Prof. Dr. Reinhold Kosfeld
University of Kassel

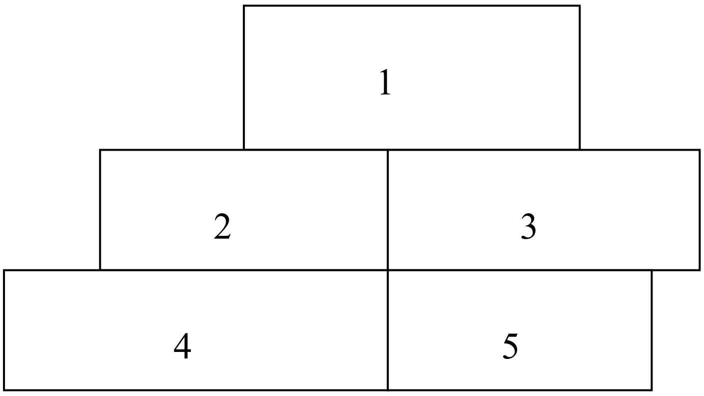
Spatial Econometrics

Assignments with Solutions

Part IV – Assignments 18 - 23

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 18 (Maximum Likelihood I)

A simple random sample ($x_1, x_2, \ldots, x_n$) is drawn from a Poission distributed population with an unkown parameter λ>0:

$$f(x|\lambda) = \frac{\lambda^x}{x!} e^{-\lambda}, \; x=0,1,2,\ldots$$

Derive the maximum likelihood (ML) estimator for λ and interpret it!

Likelihood function:

$$L(\lambda \mid x_1,\ldots,x_n) = \prod_{i=1}^{n} f(x_i \mid \lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} \cdot e^{-\lambda} = e^{-\lambda \cdot n} \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!}$$

Log likelihood funktion:

$$\ln L(\lambda \mid x_1,\ldots,x_n) = -\lambda n + \sum_{i=1}^{n} \ln \frac{\lambda^{x_i}}{x_i!} = -\lambda n + \sum_{i=1}^{n} \left(\ln \lambda^{x_i} - \ln x_i! \right)$$

$$= -\lambda n + \sum_{i=1}^{n} x_i \ln \lambda - \sum_{i=1}^{n} \ln x_i! = -\lambda n + \ln \lambda \cdot \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \ln x_i!\,.$$

Necessary condition for a maximum:

$$\frac{d \ln L(\lambda|x_1,\ldots,x_n)}{d\lambda} = -n + \frac{1}{\lambda} \cdot \sum_{i=1}^{n} x_i \overset{!}{=} 0$$

The maximum likelihood estimator $\hat{\lambda}$ for the unknown parameter λ is the value that fulfills the necessary condition:

$$-n + \frac{1}{\hat{\lambda}} \sum_{i=1}^{n} x_i = 0$$

ML estimator: $\hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}$

The ML estimator $\hat{\lambda}$ of the unknown parameter λ of a Poisson distributed population is the sample mean $\bar{x}$.

Sufficient condition for a maximum:

$$\frac{d^2 \ln L(\lambda|x_1,\ldots,x_n)}{d\lambda^2} = -\lambda^{-2} \cdot \sum_{i=1}^{n} x_i < 0$$

Assignment 19 (Maximum Likelihood II)

A simple random sample (x_1, x_2, …, x_n) is drawn from an exponentially distributed population with an unknown parameter $\lambda>0$:

$$f(x \mid \lambda) = \lambda \cdot e^{-\lambda x}, \quad x \geq 0$$

Derive the maximum likelihood (ML) estimator for λ and interpret it!

Likelihood function:

$$L(\lambda | x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i | \lambda) = \prod_{i=1}^{n} \lambda \cdot e^{-\lambda x_i} = \lambda^n \cdot \prod_{i=1}^{n} \cdot e^{-\lambda x_i}$$

Log likelihood function:

$$\ln L(\lambda | x_1, \ldots, x_n) = n \ln \lambda + \sum_{i=1}^{n} -\lambda x_i = n \ln \lambda - \lambda \sum_{i=1}^{n} x_i$$

Necessary condition for a maximum:

$$\frac{d \ln L(\lambda | x_1, \ldots, x_n)}{d\lambda} = n \cdot \frac{1}{\lambda} - \sum_{i=1}^{n} x_i \overset{!}{=} 0$$

ML estimator:

$$\hat{\lambda} = n \cdot \frac{1}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}}$$

The ML estimator $\hat{\lambda}$ of the unknown parameter λ of an exponentially distributed population is the reciprocal sample mean.

Sufficient condition for a maximum:

$$\frac{d^2 \ln L(\lambda | x_1, \ldots, x_n)}{d\lambda^2} = -n \frac{1}{\lambda^2} < 0$$

Assignment 20 (Spatial lag model – ML Estimation I)

In 5 regions data on the unemployment rate (u) and the vacancy rate (v) are observed:

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

The row-standardized contiguity matrix reads

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

a) Estimate the OLS estimators of a regression of the spatial lag of the unemploymnt rate (SL(u)) on the vacancy rate (v) using the inverse of the product matrix **X'X** (s. assigment 14 a)) :

$$(\mathbf{X'X})^{-1} = \begin{bmatrix} 1.2 & -0.5 \\ -0.5 & 0.25 \end{bmatrix}!$$

$$\mathbf{y} = \begin{bmatrix} 6 \\ 8 \\ 8 \\ 11 \\ 12 \end{bmatrix} \text{ and } \underset{5x1}{\mathbf{Wy}} = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \end{bmatrix} \begin{bmatrix} 6 \\ 8 \\ 8 \\ 11 \\ 12 \end{bmatrix} = \begin{bmatrix} 8 \\ 8\ 1/3 \\ 8\ 2/3 \\ 10 \\ 9.5 \end{bmatrix}$$

$$\mathbf{X'Wy} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 3 & 3 & 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 8 \\ 8\ 1/3 \\ 8\ 2/3 \\ 10 \\ 9.5 \end{bmatrix} = \begin{bmatrix} 44.5 \\ 85.83333 \end{bmatrix}$$

OLS estimator $\hat{\boldsymbol{\beta}}_L = (\mathbf{X'X})^{-1}\mathbf{XWy} = \begin{bmatrix} 1.2 & -0.5 \\ -0.5 & 0.25 \end{bmatrix} \begin{bmatrix} 44.5 \\ 85.83333 \end{bmatrix} = \begin{bmatrix} 10.483 \\ -0.792 \end{bmatrix}$

b) Compute the vector of residuals, $\mathbf{e}_L$, of the regression of SL(u) on v!

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

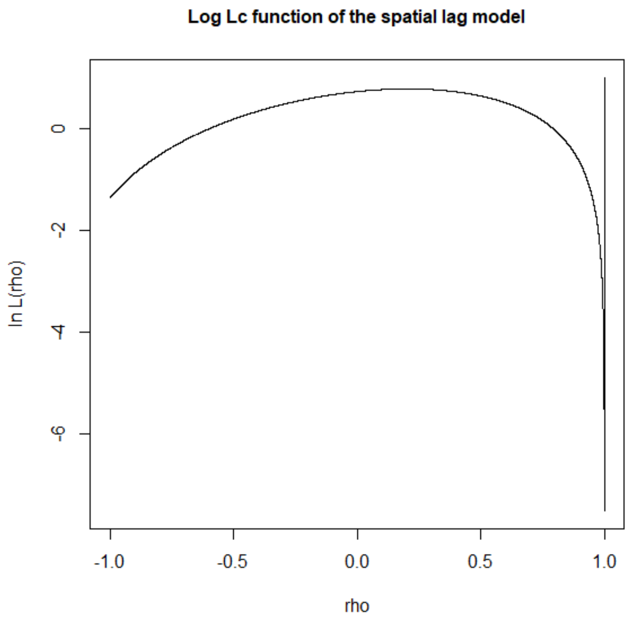
$$\hat{\mathbf{Wy}} = \mathbf{X}\hat{\boldsymbol{\beta}}_L = \begin{bmatrix} 1 & 3 \\ 1 & 3 \\ 1 & 2 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 10.483 \\ -0.792 \end{bmatrix} = \begin{bmatrix} 8.107 \\ 8.107 \\ 8.899 \\ 9.691 \\ 9.691 \end{bmatrix}$$

Vector of residuals $\mathbf{e}_L$:

$$\mathbf{e}_L = \mathbf{Wy} - \hat{\mathbf{Wy}} = \begin{bmatrix} 8 \\ 8.333. \\ 8.667 \\ 10 \\ 9.5 \end{bmatrix} - \begin{bmatrix} 8.107 \\ 8.107 \\ 8.899 \\ 9.691 \\ 9.691 \end{bmatrix} = \begin{bmatrix} -0.107 \\ 0.226 \\ -0.232 \\ 0.309 \\ -0.191 \end{bmatrix}$$

c) Determine the ML estimate for the autoregressive parameter ρ by grid search with the aid of the R file LCLAGPLOT using the residual vector (s. assignment 14 b)

$$\mathbf{e}_O = \begin{bmatrix} -0.75 \\ 1.25 \\ -1 \\ -0.25 \\ 0.75 \end{bmatrix}!$$



ML estimator for ρ: $\hat{\rho}_{ML} = 0.2042$ with $L_c^*(\hat{\rho}_{ML} = 0.2042) = 0.7715$

d) Determine the ML estimator $\hat{\boldsymbol{\beta}}_{ML}$ of the vector of regression coefficients $\boldsymbol{\beta}$ using the OLS estimator (s. assignment 14 a):

$$\hat{\boldsymbol{\beta}}_O = \begin{bmatrix} 13.5 \\ -2.25 \end{bmatrix}!$$

ML estimator for $\boldsymbol{\beta}$:

$$\hat{\beta}_{ML} = \hat{\beta}_O - \hat{\rho}_{ML} \cdot \hat{\beta}_L = \begin{bmatrix} 13.5 \\ -2.25 \end{bmatrix} - 0.2042 \cdot \begin{bmatrix} 10.483 \\ -0.792 \end{bmatrix} = \begin{bmatrix} 11.359 \\ -2.088 \end{bmatrix}$$

e) Compute the standard error of regression (SER) of ML estimation!

Vector of ML residuals:

$$e_{ML} = e_O - \hat{\rho}_{ML} \cdot e_L = \begin{bmatrix} -0.75 \\ 1.25 \\ -1 \\ -0.25 \\ 0.75 \end{bmatrix} - 0.2042 \cdot \begin{bmatrix} -0.107 \\ 0.226 \\ -0.232 \\ 0.309 \\ -0.191 \end{bmatrix} = \begin{bmatrix} -0.728 \\ 1.204 \\ -0.953 \\ -0.313 \\ 0.789 \end{bmatrix}$$

ML estimator of the error variance σ^2:

$$\hat{\sigma}^2_{ML} = \frac{1}{n}(e_O - \hat{\rho}_{ML} \cdot e_L)'(e_O - \hat{\rho}_{ML} \cdot e_L)$$

$$= \frac{1}{5} \cdot [-0.728 \quad 1.204 \quad -0.953 \quad -0.313 \quad 0.789] \begin{bmatrix} -0.728 \\ 1.204 \\ -0..953 \\ -0.313 \\ 0.789 \end{bmatrix}$$

$$= \frac{1}{5} \cdot 3.608 = 0.722$$

Standard error of regression (SER):

$$\text{SER} = \hat{\sigma} = \sqrt{0.722} = 0.850$$

Assignment 21 (Spatial lag model – ML estimation II)

ML estimation of the spatial lag model for the Beveridge curve brought out the following results (see “General information on a system of regions and data”):

```
REGRESSION
SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD
ESTIMATION
Data set: region5uv
Spatial Weight: region5_ex.GAL
Dependent Variable: U              Number of Observations: 5
Mean dependent var: 9              Number of Variables: 3
S.D. dependent var: 2.19089        Degrees of Freedom: 2
Lag coeff. (Rho):   ?

R-squared: 0.849709                Log likelihood: -6.32298
Sq. Correlation: -                 Akaike info criterion: 18.646
Sigma-square: 0.721396             Bayes criterion: 17.4743
S.E of regression: 0.84935
```

Variable	Coefficient	Std.Error	?-value	Probability
W_U	?	0.3879349	0.5273139	0.5979756
CONSTANT	11.35549	3.991297	2.845064	0.0044404
V	-2.088054	0.4804627	-4.345923	0.0000139

a) Specify the regression equation of the spatial lag model and interpret it!

According to (5.24) the spatial lag model with a dummy variable for the intercept and one substantial x-variable can be written in the form

$$y_i = \beta_1 + \beta_2 \cdot x_i + \rho \cdot \sum_{j=1}^{n} w_{ij} \cdot y_j + \varepsilon_i .$$

By replacing the errors ε_i by the ML residuals e_i and the ML estimates $\hat{\beta}_1 = 11.355$ for β_1, $\hat{\beta}_2 = -2.088$ for β_2 and $\hat{\rho} = z_{\hat{\rho}} \cdot \hat{\sigma}_{\hat{\rho}} = 0.5273139 \cdot 0.3879349 = 0.205$ we obtain the numerically specified regression equation

$$\text{(1)} \qquad y_i = 11.335 - 2.088 \cdot x_i + 0.205 \cdot \sum_{j=1}^{n} w_{ij} \cdot y_j + e_i$$

for the Beveridge curve.

Equation (1) explains the regional unemployment rate (u) by own regions’s vacancy rate (v) and the spatially lagged unemployment rate (Lu). The latter variable captures unemployment spillovers in surrounding regions on a region’s unemployment. It is measured by the average unemployment rate in neighbouring regions.

ML estimation has brought out a significant negative relationship between the unemployment rate and the vacancy rate. A rise in the vacancy rate by 1 percentage point on average comes along with a drop of the unemployment rate by 2.1 percentage points. When the unemployment rate in contiguous regions rises by 1 percentage point, unemployment rate of the *ith* region is expected to increase by 0.2 percentage points. In our small regional system, however, the effect of the spatially lagged variable Lu cannot be proved to be significant.

b) Explain the structure of the log-likelihood function of the spatial lag model!

For ease of exposition, take the error variance σ^2 is given. The log-likelihood function of the spatial lag model,

$$\ln L\left(\boldsymbol{\beta},\rho,\sigma^2 \middle| \mathbf{y},\mathbf{X}\right) = C + \underbrace{\ln\left|\mathbf{I}-\rho\mathbf{W}\right|}_{=\text{log of Jacobian}} - \frac{1}{2\sigma^2}\underbrace{(\mathbf{y}-\rho\mathbf{W}\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\rho\mathbf{W}\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}_{=\text{sum of squared erros}},$$

then consists – apart from a constant C - of

- the sum of squared residuals $(\mathbf{y}-\rho\mathbf{W}\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\rho\mathbf{W}\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$ (times a const.),
- the log of the Jacobian determinant $\ln\left|\mathbf{I}-\rho\mathbf{W}\right|$.

Ordinary least-squares (OLS) ignores the Jacobian that accounts for spatial interaction in estimating the regression coefficients. This is only admissible when the autoregressive parameter goes to zero. In this case, maximising the log-likelihood function is equivalent to minimising the sum of squared residuals. Thus, for $\rho \neq 0$, the maximum likelihood (ML) method provides a correction of distortions resulting from applying the “pure” least-squares principle. According to this correction, the ML estimates for $\boldsymbol{\beta}$ and ρ are ensured to be unbiased, consistent and efficient (= most precise) in large samples.

c) Is the autoregressive parameter of the spatial lag model significant (α=0.05)?

● Test of significance for ρ (two-sided test)

H_0: $\rho = 0$ and H_1: $\rho \neq 0$

ML estimator for ρ: $\hat{\rho}_{ML} = 0.205$ (see part a))

Test statistic: $z_\rho = 0.527$

Critical value (α=0.05, two-sided test): $z_{0.975} = 1.96$

Testing decision: $|z_\rho| = 0.527 < z_{0.975} = 1.96$ => Accept H_0

d) Specify the variance-covariance matrix of the errors in equation (5.5b)/(5.6b)!

Hint: $(\mathbf{I}-\rho\mathbf{W}')^{-1} = [(\mathbf{I}-\rho\mathbf{W})^{-1}]'$

(5.5b) $\mathbf{y} = (\mathbf{I}-\rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}$

(5.6b) $\mathbf{y} = (\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I}-\rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}$

Variance-covariance matrix of the errors $\boldsymbol{\upsilon} = (\mathbf{I}-\rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}$:

$$\begin{aligned}
\mathrm{Cov}(\boldsymbol{\upsilon}) &= \mathrm{E}(\boldsymbol{\upsilon}\boldsymbol{\upsilon}') = \mathrm{E}\{[(\mathbf{I}-\rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}][(\mathbf{I}-\rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}]'\} \\
&= \mathrm{E}\{(\mathbf{I}-\rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'[(\mathbf{I}-\rho\mathbf{W})^{-1}]'\} \\
&= (\mathbf{I}-\rho\mathbf{W})^{-1}\sigma^2\mathbf{I}[(\mathbf{I}-\rho\mathbf{W})^{-1}]' \\
&= \sigma^2(\mathbf{I}-\rho\mathbf{W})^{-1}(\mathbf{I}-\rho\mathbf{W}')^{-1}
\end{aligned}$$

Estimated variance-covariance matrix of the errors $\boldsymbol{\upsilon} = (\mathbf{I}-\rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}$:

$$\widehat{\mathrm{Cov}}(\boldsymbol{\upsilon}) = \hat{\sigma}^2(\mathbf{I}-\hat{\rho}\mathbf{W})^{-1}(\mathbf{I}-\hat{\rho}\mathbf{W}')^{-1}$$

with $\hat{\sigma}^2 = 0.721396$, $\hat{\rho} = z_{\hat{\rho}} \cdot \hat{\sigma}_{\hat{\rho}} = 0.5273139 \cdot 0.3879349 = 0.204563$

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

$\mathbf{I} - \hat{\rho}\cdot\mathbf{W} = \mathbf{I} - 0.204563\cdot\mathbf{W}$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} - 0.204563\cdot\begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \end{bmatrix}$$

$$
=\begin{bmatrix}1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\\0&0&0&1&0\\0&0&0&0&1\end{bmatrix}-\begin{bmatrix}0&0.102282&0.102282&0&0\\0.068188&0&0.068188&0.068188&0\\0.068188&0.068188&0&0&0.068188\\0&0.102282&0&0&0.102282\\0&0&0.102282&0.102282&0\end{bmatrix}
$$

$$
=\begin{bmatrix}1&-0.102282&-0.102282&0&0\\-0.068188&1&-0.068188&-0.068188&0\\-0.068188&-0.068188&1&0&-0.068188\\0&-0.102282&0&1&-0.102282\\0&0&-0.102282&-0.102282&1\end{bmatrix}
$$

$$
(\mathbf{I}-\hat{\rho}\cdot\mathbf{W})^{-1}=(\mathbf{I}-0.204563\cdot\mathbf{W})^{-1}
$$

$$
=\begin{bmatrix}1.0153&0.1124&0.1124&0.0085&0.0085\\0.0749&1.0203&0.0785&0.0709&0.0126\\0.0749&0.0785&1.0203&0.0126&0.0709\\0.0085&0.1063&0.0189&1.0180&0.1054\\0.0085&0.0189&0.1063&0.1054&1.0180\end{bmatrix}
$$

$$
(\mathbf{I}-\hat{\rho}\cdot\mathbf{W}')^{-1}=(\mathbf{I}-0.204563\cdot\mathbf{W}')^{-1}=[(\mathbf{I}-0.204563\cdot\mathbf{W})^{-1}]'
$$

$$
=\begin{bmatrix}1.0153&0.0749&0.0749&0.0085&0.0085\\0.1124&1.0203&0.0785&0.1063&0.0189\\0.1124&0.0785&1.0203&0.0189&0.1063\\0.0085&0.0709&0.0126&1.0180&0.1054\\0.0085&0.0126&0.0709&0.1054&1.0180\end{bmatrix}
$$

$$
\widehat{\mathrm{Cov}}(\boldsymbol{\upsilon})=\hat{\sigma}^2(\mathbf{I}-\hat{\rho}\mathbf{W})^{-1}(\mathbf{I}-\hat{\rho}\mathbf{W}')^{-1}
$$

$$
=0.721396\cdot\begin{bmatrix}1.0563&0.2003&0.2003&0.0323&0.0323\\0.2003&1.0579&0.1676&0.1840&0.0486\\0.2003&0.1676&1.0579&0.0486&0.1840\\0.0323&0.1840&0.0486&1.0592&0.2187\\0.0323&0.0486&0.1840&0.2187&1.0592\end{bmatrix}
$$

$$
=\begin{bmatrix}0.7620&0.1445&0.1445&0.0233&0.0233\\0.1445&0.7632&0.1209&0.1328&0.0350\\0.1445&0.1209&0.7632&0.0350&0.1328\\0.0233&0.1328&0.0350&0.7641&0.1578\\0.0233&0.0350&0.1328&0.1578&0.7641\end{bmatrix}[
$$

Interpretation:

While the errors ε_i in the representation (5.5a)/(5.6a) of the spatial lag model follow the standard assumptions, the errors in the representation (5.5b)/(5.6b) (here denoted by υ_i) have a slight tendency to heteroscedasticity (slight differences between the diagonal elements of $\widehat{Cov}(\boldsymbol{\upsilon})$) and are spatially auto-correlated (non-zero off-diagonal elements of $\widehat{Cov}(\boldsymbol{\upsilon})$). Both effects can be ascribed to the spatially lagged endogenous variable **Wy**, as they vanish as soon as this variable enters the regression model.

Assignment 22 (Spatial error model – ML estimation)

ML estimation of the spatial error model for the Beveridge curve brought out the following results (see "General information on a system of regions and data"):

```
REGRESSION
SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD
ESTIMATION
Data set: region5uv
Spatial Weight: region5_ex.GAL
Dependent Variable:U             Number of Observations: 5
Mean dependent var: 9.000000     Number of Variables: 2
S.D. dependent var: 2.190890     Degree of Freedom: 3
Lag coeff. (Lambda): -0.774416

R-squared:    0.937623           R-squared (BUSE): -
Sq. Correlation: -               Log likelihood: -4.803067
Sigma-square: 0.299409           Akaike info criterion:13.6061
S.E of regression: 0.547183      Bayes criterion: 12.825011
```

Variable	Coefficient	Std.Error	?-value	Probability
CONSTANT	13.95179	0.4773741	29.22611	0.0000000
V	-2.461901	0.2252333	-10.93045	0.0000000
LAMBDA	-0.7744158	0.2980455	?	?

a) Specify the spatial error process and interpret it!

According to (5.35), in ML estimation of the spatial error model, a spatially autoregressive error process is adopted:

$$\boldsymbol{\varepsilon} = \boldsymbol{\lambda}\mathbf{W}\boldsymbol{\varepsilon} + \mathbf{v} \quad \text{or} \quad \varepsilon_i = \lambda \cdot \sum_{j=1}^{n} w_{ij} \cdot \varepsilon_j$$

By replacing the ML estimate $\hat{\lambda}$ = -0.774 for the autoregressive parameter λ, we obtain the numerically specified spatial error process

(1a) $\boldsymbol{\varepsilon} = -0.774 \cdot \mathbf{W}\boldsymbol{\varepsilon} + \mathbf{v}$

or

(1b) $\varepsilon_i = -0.774 \cdot \sum_{j=1}^{n} w_{ij} \cdot \varepsilon_j$

for the Beveridge curve.

Equation (1) states negative dependencies between disturbances in the *ith* region and its neighbouring regions. This means that exogenous shocks in a region will negatively affect unemployment (u) in the surrounding regions. This can be explained when the errors ε_i capture the effects of omitted regional attributes. Suppose, the region i is an eligible area to financial support from EU funds but its neighbouring regions not. In this case, firms situated in the surrounding regions may relocate due to cost benefits. Thus, the positive exogenous shock (= regional promotion) in region i may entail an increase in unemployment in the surrounding area thereby reflecting negative error dependence.

b) Give the Jacobian term of the log-likelihood function and interpret its role in estimation!

The Jacobian term of the log-likelihood function of the spatial error model is given by the log of the expression

$$J = \left|\frac{\partial \mathbf{v}}{\partial \mathbf{y}}\right| = \left|\mathbf{I} - \lambda \mathbf{W}\right|.$$

It measures the change in the log-likelihood function resulting from spatially autocorrelated errors compared to a regression model with uncorrelated errors. Due to the entrance of ln $\left|\mathbf{I} - \lambda \mathbf{W}\right|$ in the log-likelihood function, the ML method corrects the pure least-squares approach to estimating the regression coefficients in the spatial error model for spatial effects. In this way, efficient (= most precise) parameter estimates and unbiased standard errors of the regression coefficients are ensured in large samples.

c) Is the autoregressive parameter of the spatial error model significant (α=0.01)?

● Test of significance for λ (two-sided test)

H_0: $\lambda = 0$ and H_1: $\lambda \neq 0$

ML estimator for λ: $\hat{\lambda}_{ML} = -0.774$

Test statistic: $z_\lambda = \dfrac{\hat{\lambda}}{\hat{\sigma}_\lambda} = \dfrac{-0.774}{0.298} = -2.597$

Critical value (α=0.05, two-sided test): $z_{0.995} = 2.5758$

Testing decision: $| z_\lambda = -2.597| > z_{0.995} = 2.5758$ => Reject H_0

d) Specify the variance-covariance matrix of the errors!

Hint: $(\mathbf{I}-\lambda\mathbf{W}')^{-1} = [(\mathbf{I}-\lambda\mathbf{W})^{-1}]'$

Variance-covariance matrix of the errors $\boldsymbol{\varepsilon} = (\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{v}$:

(5.40) $\mathrm{Cov}(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2(\mathbf{I}-\lambda\mathbf{W})^{-1}[(\mathbf{I}-\lambda\mathbf{W}')^{-1}]$

Estimated variance-covariance matrix of the errors $\boldsymbol{\varepsilon} = (\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{v}$:

$$\hat{\mathrm{Cov}}(\boldsymbol{\varepsilon}) = \hat{\sigma}^2(\mathbf{I}-\hat{\lambda}\mathbf{W})^{-1}(\mathbf{I}-\hat{\lambda}\mathbf{W}')^{-1}$$

with $\hat{\sigma}^2 = 0.299409$ and $\hat{\lambda} = -0.774416$

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

As the variance-covariance matrix of the errors ε_i in the spatial error model is of the same structure as that of the errors υ_i in the spatial lag model, the calculation of the matrix product $(\mathbf{I}+0.774416\cdot\mathbf{W})^{-1}[(\mathbf{I}+0.774416\cdot\mathbf{W}')^{-1}]$ can be tracked from assignment 21.

$$\hat{\mathrm{Cov}}(\boldsymbol{\varepsilon}) = 0.299409\cdot(\mathbf{I}+0.774416\cdot\mathbf{W})^{-1}[(\mathbf{I}+0.774416\cdot\mathbf{W}')^{-1}]$$

$$= 0.299409\cdot\begin{bmatrix} 1.7655 & -0.7267 & -0.7267 & 0.2555 & 0.2555 \\ -0.7267 & 2.3581 & -1.1572 & -1.9401 & 1.4689 \\ -0.7267 & -1.1572 & 2.3581 & 1.4689 & -1.9401 \\ 0.2555 & -1.9401 & 1.4689 & 3.1284 & -2.4392 \\ 0.2555 & 1.4689 & -1.9401 & -2.4392 & 3.1284 \end{bmatrix}$$

$$= \begin{bmatrix} 0.5286 & -0.2176 & -0.2176 & 0.0765 & 0.0765 \\ -0.2176 & 0.7060 & -0.3465 & -0.5809 & 0.4398 \\ -0.2176 & -0.3465 & 0.7060 & 0.4398 & -0.5809 \\ 0.0765 & -0.5809 & 0.4398 & 0.9367 & -0.7303 \\ 0.0765 & 0.4398 & -0.5809 & -0.7303 & 0.9367 \end{bmatrix}$$

Interpretation:

The non-zero off-diagonal elements of $\hat{\text{Cov}}(\boldsymbol{\varepsilon})$ indicate spatially autocorrelated errors in the regression equation of the Beveridge curve. The diagonal elements of $\hat{\text{Cov}}(\boldsymbol{\varepsilon})$ differ stronger from one another than those of $\hat{\text{Cov}}(\boldsymbol{\upsilon})$ in the spatial lag model. Heteroscedasticity is therefore stronger marked in the spatial error model. Both effects are accounted for by a spatially autocorrelated error process in ML estimation of the regression coefficients.

Assignment 23 (Model comparison)

Estimation of the Beveridge curve (see "General information on a system of regions and data") has brought out the following summary statistics with four different regression models:

	Standard regression model	Spatial cross-regressive model	Spatial lag model	Spatial error model
R^2	0.844	0.988	0.850	0.938
Adj. R^2	0.792	0.976	0.699	0.875
AIC	16.761	5.899	18.646	13.606
BIC	15.970	4.727	17.474	12.825

a) Explain the summary statistics for model selection!

The **coefficient of determination** (R^2) serves as a measure of goodness of fit in regression analysis. It relates the variance explained the regression (= variance of the fitted values) ($s_{\hat{y}}^2$) to the variance of the dependent variable (s_y^2):

$$R^2 = \frac{s_{\hat{y}}^2}{s_y^2}.$$

As the explained variance $s_{\hat{y}}^2$ increases when additional variables enter the regression equation regardless of whether their contribution is substantive or nor, R^2 is not a suitable measure for model selection in case of different numbers of explanatory variables. It is only a measure of goodness of fit in a single regression model.

To account for increases of R^2 simply arising from random effects, an **adjusted coefficient of determination** $\overline{R}^2$ is defined. It modifies the coefficient of determination by replacing the sample size n by the numbers of degrees of freedom in $s_{\hat{y}}^2$ (=n-k) and in s_y^2 (=n-1):

$$\overline{R}^2 = 1 - \frac{n-1}{n-k}(1 - R^2).$$

The isolated effect of the factor (n-1)/(n-k) is a decrease of $\overline{R}^2$ when the number of exploratory variables (k) is increased. $\overline{R}^2$ will, however, only increase if the additional explanation by a new variable overcompensates this effect. In this case the explanation of an additional variable is viewed to be substantive. In contrast to R^2, the adjusted coefficient of determination can be used for model selection.

Like the adjusted coefficient of determination, the **information criteria** AIC and BIC aim at differentiating purely random from substantive contributions of an additional explanatory variable. They combine the log likelihood function (ln L) with a "penalty function". As ln L crucially depends on the sum of squared residuals (SSR), it increases when additional variables enter the regression equation (following the decrease of SSR). Thus, SSR and hence ln L are not suitable measures for model selection in case of different numbers of explanatory variables (k). The information criteria AIC and BIC aim to eliminate purely random increases of ln L by adding a function of k. This function is called a "penalty function" as it penalizes extensive model building using as much as possible explanatory variables. The information criteria AIC and BIC differ in the strength by which models with a larger number of explanatory variables are penalized.

The Akaike information criterion (AIC) is given by

$$AIC = -2 \cdot \ln(L) + 2k$$

and the Schwartz information criterion (BIC) by

$$BIC = -2 \cdot \ln(L) + k \cdot \ln(n).$$

The lower SSR and the larger ln(L), the lower AIC and BIC will be. Thus, according to the information criteria, regression models with the lowest AIC and BIC values are preferable. For usual sample ($n > 8$) sizes additional variables are stronger penalized with the Schwartz information criterion (BIC) than with the Akaike information criterion (AIC).

For our small regional system ($n=5$), however, additional variables are stronger penalized with AIC than with SC:

Penalty function for AIC: $2 \cdot k$

Penalty function for BIC: $k \cdot \ln(n) = k \cdot \ln(5) = 1.609 \cdot k$.

The information criteria AIC and BIC are available with both least-squares (LS) and maximum likelihood (ML) estimation.

b) Which regression model would you select for modelling the Beveridge curve? Substantiate your selection!

The **coefficient of determination** indicates the share of the variance of the unemployment rate that is explained by a regression model for the Beveridge curve. As an increase of R^2 is expected when passing from the standard regression model to a spatial model with additional explanatory variables, the coefficient of determination cannot be used for model selection.

Based on the **adjusted coefficient of determination**, the spatial lag model is inferior to the standard regression model as $\overline{R}^2$ drops from 0.792 to 0.699. By contrast, both the spatial error model and the spatial cross-regressive model markedly outperform the standard regression model. While the increase amounts to 8.3 percentage points (from 0.792 to 0,875) in the former case and 18.4 percentage points in the latter case (from 0.792 to 0.976), the $\overline{R}^2$ criterion clearly favours the SLX model.

For a well-founded assessment of the performance of models, **information criteria** hould be used. In our case, the model ranking of both information criteria AIC and BIC are in line with that of the adjusted coefficient of determintion. The spatial cross-regressive model clearly performs best, as AIC and BIC show the lowest values. The AIC and BIC values of the second-best model, the spatial error model, increase considerably. Obviously, regional interaction cannot be explained by spatially lagged unemployment at all as the standard regression model outperforms the spatial lag model.