## 5.3.2 Maximum likelihood (ML) method

**<u>Excursus A</u>:**

### ● Point estimation in inductive statistics by maximum likelihood

given: sample ($x_1$, $x_2$, …, $x_n$) of n identically and independently distributed (i.i.d) sample variables $X_1$, $X_2$, …, $X_n$ from a population with an unknown parameter θ

Known type of distribution of the random variable X in the population (e.g. normal distribution)

Density function: f(x|θ) for all x

As the sample variables $X_i$ are distributed like X and independently drawn from the population, the **joint density function** $f_{X1,\ldots,Xn}$ of the n sample variables is given by the product of the **marginal density functions** f:

(A5.1) $$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n|\theta) = f(x_1|\theta)\ldots f(x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

● **Idea of the maximum likelihood method**:
Which value of the parameter θ in the population will maximise the probabilty of the given sample values to be realised?

Before drawing the sample:
Parameter θ is a fixed quantity, wheras the potential sample values $X_1.X_2,….X_n$ are variable

After the drawing of the sample:
Sample values $x_1,x_2,…,x_n$ have been realised (fixed), so that $f_{X1},…,_{Xn}$ can be viewed as a function of the unknown parameter θ

In order to stress the situation after drawing the sample, one replaces the **joint density function** $f_{X1},…,_{Xn}$ in (A5.1) by the **likelihood function** L:

(A5.2) $$L(\theta|x_1,\ldots,x_n)=\prod_{i=1}^{n} f(x_i|\theta)$$

The likelihood function of an unknown parameter θ renders the joint densitiy for each value of θ for a given sample $(x_1,x_2,…,x_n)$.

## • Maximising of the likelihood function

Maximising the likelihood function L with respect to θ is based on differential calculus. Because sums can be easier differentiated than products, usually the **log-likelihood function** (logarithm of the likelihood function) is considered:

$$\text{(A5.3)} \qquad \ln L\left(\theta \middle| x_1,\ldots,x_n\right) = \sum_{i=1}^{n} \ln f\left(x_i|\theta\right)$$

As the logarithm of a function is a monotonic transformation, the value of θ that maximises ln L as well maximises L. The necessary condition for a maximum of ln L reads

$$\text{(A5.4)} \qquad \frac{d \ln L\left(\theta \middle| x_1,\ldots,x_n\right)}{d\theta} \overset{!}{=} 0$$

The **maximum likelihood estimator** (ML estimator) $\hat{\theta}_{ML}$ for the unknown parameter θ is obtained by solving the equation (A5.4).

**Example**: **Maximum likelihood (ML) estimator of parameter p**

Probability of event A: P(A) = p

Random variable X: Number of realizations of A in n Bernoulli trials

Random variable $X_i$ in i*th* trial: 1 if A occurs, 0 else

$$X = \sum_{i=1}^{n} X_i \quad \uparrow \quad 0,1$$

Probability function of Bernoulli distribution: $f(x_i|p) = p^{x_i} \cdot (1-p)^{1-x_i}$

Joint probability function: $f_x(x|p,n) = \binom{n}{x} p^x \cdot (1-p)^{1-x}$

Likelihood function:

$$L(p|x_1, x_2, \ldots, x_n) = \binom{n}{x} \prod_{i=1}^{n} f(x_i|p,n) = f_x(x|p,n) = \binom{n}{x} p^x \cdot (1-p)^{n-x}$$

$\uparrow$

The sum x can result from $\binom{n}{x}$ permutations of n $x_i$ (0, 1) values

Necessary condition: $\frac{dL(p \mid x_1,\ldots,x_n)}{dp} \overset{!}{=} 0$

$$\frac{dL(p)}{dp} = \binom{n}{x}\left[x \cdot p^{x-1} \cdot (1-p)^{n-x} + p^{x} \cdot (n-x) \cdot (1-p)^{n-x-1} \cdot (-1)\right]$$

$$= \binom{n}{x} p^{x-1} \cdot (1-p)^{n-x-1} \cdot \left[x \cdot (1-p) - p \cdot (n-x)\right] \overset{!}{=} 0$$

$$x \cdot (1-p) - p \cdot (n-x) = 0$$

$$p \cdot (n-x) = x - x \cdot p$$

$$p \cdot (n-x+x) = x$$

$$\mathbf{p \cdot n = x}$$

$$\hat{p}_{ML} = \frac{x}{n} \quad \text{(ML estimator of p)}$$

## ● ML estimators for μ and $\sigma^2$ of the normal distribution

Suppose, the random variable X is normally distributed with the unknown parameters

(A5.5) $E(X) = \mu$ and (A5.6) $Var(X) = \sigma^2$.

Then the probability density function of X is given by

$$(A5.7) \quad f\left(x \mid \mu, \sigma^2\right) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

If we take a simple random sample ($x_1$, $x_2$, …, $x_n$) from the normally distributed population, the likelihood function of this sample reads

(A5.8)

$$L\left(\mu, \sigma^2 \mid x_1, \ldots, x_n\right) = \prod_{i=1}^{n} \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}} \cdot e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2} = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \cdot \prod_{i=1}^{n} \cdot e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

By taking the logs of both sides of (A5.8), the log likelihood function

$$(A5.9) \quad \ln L\left(\mu, \sigma^2 \mid x_1, \ldots, x_n\right) = \frac{n}{2} \ln\left(\frac{1}{2\pi\sigma^2}\right) + \sum_{i=1}^{n} -\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2$$

results.

By simplifying (A5.9) we obtain

$$\text{(A5.10)}\quad \ln L\left(\mu,\sigma^2 \mid x_1,\ldots,x_n\right)=\frac{n}{2}\left(\ln 1-\ln 2\pi\sigma^2\right)-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2$$

$$=-\frac{n}{2}\left(\ln 2\pi+\ln\sigma^2\right)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2$$

$$=-\frac{n}{2}\ln 2\pi-\frac{n}{2}\ln\sigma^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right)^2$$

The necessary conditions for a maximum of the log likelihood function are obtained by setting the partial derivatives of ln L with respect to μ and $\sigma^2$ equal to zero:

$$\text{(A5.11a)}\quad \frac{\partial \ln L\left(\mu,\sigma^2\right)}{\partial\mu}=\frac{1}{\sigma^2}\cdot\sum_{i=1}^{n}\left(x_i-\mu\right)\overset{!}{=}0$$

$$\text{(A5.11b)}\quad \frac{\partial \ln L\left(\mu,\sigma^2\right)}{\partial\sigma^2}=-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\cdot\sum_{i=1}^{n}\left(x_i-\mu\right)^2\overset{!}{=}0$$

ML estimators $\hat{\mu}_{ML}$ and $\hat{\sigma}^2_{ML}$ of the paramaters (A5.5) and (A5.6) are the values of μ and $\sigma^2$ which meet the conditions (A5.11a) and (A5.11b).

With $\sigma^2 \neq 0$, (A5.11a) simplifies to

$$\sum_{i=1}^{n}(x_i - \mu) = \sum_{i=1}^{n} x_i - n\mu = 0$$

what from

$$\sum_{i=1}^{n} x_i = n\mu$$

and

(A5.12) $$\hat{\mu}_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

follows. Thus, the **ML estimator for μ** is simply equal to the sample mean. Inserting $\bar{x}$ by μ in (A5.11b), we have

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 \Big/ 2\sigma^4 = n \Big/ 2\sigma^2,$$

and

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = 2\sigma^4 \, n \Big/ 2\sigma^2 = \sigma^2 \cdot n$$

From this the **ML estimator for $\sigma^2$**

(A5.13) $$\hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

follows, which is equal to the sample variance with the factor 1/n instead of 1/(n-1).

**Excursus B:**

● **Maximum likelihood estimation of β in the standard regression model**

We assume the errors u of the standard regression model

(4.1) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$

to be identically independently normally distributed (i.i.n.d.) with the expected value and variance-covariance matrix

$$E(\mathbf{u}) = \mathbf{0} \quad \text{and} \quad Cov(\mathbf{u}) = E(\mathbf{uu}') = \sigma^2 \cdot \mathbf{I},$$

respectively:

(B5.1) $\mathbf{u} \sim N(\mathbf{0}, \sigma^2 \cdot \mathbf{I})$ .

Then the density function of the disturbance variable of region i is given by

(B5.2) $$f(u_i;0,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left( -\frac{1}{2\sigma^2} u_i^2 \right)$$

In case of a normal distribution the uncorrelatedness of the errors $u_i$ entails stochastical independence. Thus, the joint density function $f_{u1,\ldots,un}$ is given by the product of the marginal density functions f:

(B5.3) $$f_{u_1,\ldots,u_n}(\mathbf{u};\mathbf{0},\sigma^2\mathbf{I}) = \prod_{t=1}^{n} f(u_i;0,\sigma^2)$$

Using (B5.2) we obtain

$$f(\mathbf{u};\mathbf{0},\sigma^2\mathbf{I}) = \frac{1}{(2\pi)^{1/2}(\sigma^2)^{1/2}}\exp\left(-\frac{u_1^2}{2\sigma^2}\right) \cdot \ \ldots \ \cdot \frac{1}{(2\pi)^{1/2}(\sigma^2)^{1/2}}\exp\left(-\frac{u_n^2}{2\sigma^2}\right)$$

$$= \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}}\exp\left(-\frac{1}{2\sigma^2}(u_1^2+u_2^2+\ldots+u_n^2\right) = \frac{1}{(2\sigma)^{n/2}(\sigma^2)^{n/2}}\exp\left(-\frac{1}{2\sigma^2}\sum u_i^2\right)$$

or because of $\sum u_t^2$ = **u'u**

(B5.4) $$f(\mathbf{u};\mathbf{0},\sigma^2\mathbf{I}) = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}}\exp\left(-\frac{1}{2\sigma^2}\mathbf{u'u}\right)$$

The joint density function for **y**, $f_{y1,\ldots,yn}$, is given by

$$\text{(B5.5)} \qquad f_{y_1,\ldots,y_n}(\mathbf{y}) = f_{u_1,\ldots,u_n}(\mathbf{u}) \cdot \left|\frac{\partial \mathbf{u}}{\partial \mathbf{y}}\right|$$

where $\left|\partial \mathbf{u} / \partial \mathbf{y}\right|$ is absolute value of the Jacobian (Jacobian determinant) J formed by the partial derivatives of the elements of **u** with respect to the elements of **y**:

$$\text{(B5.6)} \qquad J = \left|\frac{\partial \mathbf{u}}{\partial \mathbf{y}}\right| = \begin{vmatrix} \partial u_1/\partial y_1 & \partial u_1/\partial y_2 & \cdots & \partial u_1/\partial y_n \\ \partial u_2/\partial y_1 & \partial u_2/\partial y_2 & \cdots & \partial u_2/\partial y_n \\ \vdots & \vdots & \ddots & \vdots \\ \partial u_n/\partial y_1 & \partial u_n/\partial y_2 & \cdots & \partial u_n/\partial y_n \end{vmatrix}$$

As the relation between **u** and **y** reads [from (4.1)]

$$\text{(B5.7)} \qquad \mathbf{u} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$$

the partial derivatives $\partial u_i/\partial y_j$ are equal to 1 for i=j and 0 for i≠j. Thus, in the standard regression model, the Jacobian J is equal to 1, so that

$$\text{(B5.8)} \quad f_{y_1,\ldots,y_n}(\mathbf{y}) = f_{u_1,\ldots,u_n}(\mathbf{u}) = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right)$$

follows on account of (B5.5).

Given the sample data **y** and **X**, the likelihood function L depends on **β** and $\sigma^2$:

$$\text{(B5.9)}\quad L(\boldsymbol{\beta},\sigma^2 \mid \mathbf{y},\mathbf{X}) = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right)$$

ML estimators $\hat{\boldsymbol{\beta}}_{ML}$ and $\hat{\sigma}^2_{ML}$ are given by the parameter values **β** and $\sigma^2$ that maximise the likelihood function:

$$\text{(B5.10)}\quad L\left(\hat{\boldsymbol{\beta}}_{ML},\hat{\sigma}^2_{ML}\right) = \underset{\boldsymbol{\beta},\sigma^2}{\text{Max}}\, L\left(\boldsymbol{\beta},\sigma^2 \mid \mathbf{y},\mathbf{X}\right)$$

For ease of calculations, we maximise the log likelihood function

$$\text{(B5.11)}\quad \ln L\left(\boldsymbol{\beta},\sigma^2 \mid \mathbf{y},\mathbf{X}\right) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$$

instead of L. As **β** does only appear in the last term of (B5.11), maximising the log likelihood function L with respect to **β** is equal to minimising the function

$$\text{(B5.12)}\quad Q(\boldsymbol{\beta}) = \mathbf{u}'\mathbf{u} = (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$$

The function Q renders exactly the least-squares criterion. Thus, in the standard regression model, the ML estimator for **β** coincides with the OLS estimator:

$$\text{(B5.13)}\quad \hat{\boldsymbol{\beta}}_{ML} = \hat{\boldsymbol{\beta}}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

According to the Gauss-Markov theorem, $\hat{\boldsymbol{\beta}}_{ML}(=\hat{\boldsymbol{\beta}}_{OLS})$ is the best unbiased linear estimator. In case of the normal distribution $\hat{\boldsymbol{\beta}}_{ML}$ has the lowest variance of any unbiased estimator (efficient estimator). Moreover, $\hat{\boldsymbol{\beta}}_{ML}$ is consistent and normally distributed.

In order to obtain a ML estimator for the error variance $\sigma^2$, we set the partial derivative of the log likelihood function (B5.11) with respect to $\sigma^2$ equal to 0:

(B5.14) $$\frac{\partial \ln L\left(\beta,\sigma^2 \mid \mathbf{y},\mathbf{X}\right)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\cdot(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\cdot(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) = 0$$

From (B5.14) we obtain

$$n = \frac{1}{\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$$

and with $\hat{\boldsymbol{\beta}}_{ML}$ for $\boldsymbol{\beta}$

(B5.15) $$\hat{\sigma}^2_{ML} = \frac{1}{n}\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}_{ML}\right)'\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}_{ML}\right)$$

On account of $\hat{\boldsymbol{\beta}}_{ML} = \hat{\boldsymbol{\beta}}_{OLS}$ the components of $\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}_{ML}$ are identical with the OLS residuals $\hat{u}_i$. Thus, the ML estimator for the error variance is given by

(B5.16) $$\hat{\sigma}^2_{ML} = \frac{1}{n}\hat{\mathbf{u}}'\hat{\mathbf{u}} = \frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^2$$

$\hat{\sigma}^2_{ML}$ is asymptotically unbiased. An unbiased estimator for $\sigma^2$ is obtained by replacing the sample size n by the degrees of freedom n-k:

(B5.17) $$\hat{\sigma}^2 = \frac{1}{n-k}\hat{\mathbf{u}}'\hat{\mathbf{u}} = \frac{1}{n-k}\sum_{t=1}^{n}\hat{u}_t^2$$

In general, ML estimators must not possess desirable small but large sample properties. Under usually fulfilled regularity conditions they are asymptotically unbiased, consistent and asymptotically efficient.

### 5.3.3 Maximum likelihood estimation in the spatial lag model

The spatial lag model discussed in section 5.3.1 is given in augmented form by the equation (5.6),

(5.6a) $$\mathbf{y} = \mathbf{X\beta} + \rho\mathbf{W}\mathbf{y} + \boldsymbol{\varepsilon} \ ,$$

or equation (5.6b):

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

The errors $\varepsilon_i$ are assumed to be identically and independently normally distributed (i.i.n.d.):

(5.27) $$\boldsymbol{\varepsilon} \sim \mathrm{N}(0, \sigma^2\mathbf{I}) \ .$$

By solving (5.6a) for $\boldsymbol{\varepsilon}$,

(5.28) $$\boldsymbol{\varepsilon} = (\mathbf{I} - \rho\mathbf{W}) \cdot \mathbf{y} - \mathbf{X\beta} \ ,$$

we obtain the Jacobian

(5.29) $$\mathrm{J} = \left| \frac{\partial\boldsymbol{\varepsilon}}{\partial\mathbf{y}} \right| = |\mathbf{I} - \rho\mathbf{W}|$$

Thus, the **log likelihood function** for **y** of the **spatial lag model** is obtained by adding the term ln |**I**-ρ**W**| to the log likelihood function of the standard regression model (B5.11) and replacing **u** by **ε** according to (5.28):

(5.30a) $$\ln L\left(\boldsymbol{\beta},\rho,\sigma^2 \middle| \mathbf{y},\mathbf{X}\right) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln\left|\mathbf{I}-\rho\mathbf{W}\right| - \frac{1}{2\sigma^2}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} \quad .$$

Replacing the random vector **ε** by (5.28) with

$$\boldsymbol{\varepsilon} = (\mathbf{I}-\rho\mathbf{W})\cdot\mathbf{y} - \mathbf{X}\boldsymbol{\beta} = \mathbf{y} - \rho\mathbf{W}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}$$

the log likelihood function reads

(5.30b) $$\ln L\left(\boldsymbol{\beta},\rho,\sigma^2 \middle| \mathbf{y},\mathbf{X}\right) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln\left|\mathbf{I}-\rho\mathbf{W}\right| - \frac{1}{2\sigma^2}(\mathbf{y}-\rho\mathbf{W}\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\rho\mathbf{W}\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \quad .$$

Maximising the log likelihood function, ln L ,is equal to minimising the sum of squared errors, **ε**'**ε**, corrected by the log of the Jacobian, ln $|\mathbf{I}-\rho\mathbf{W}|$. On account of this correction the ML estimates will differ from the OLS estimates. They coincide for $\rho \to 0$ where the spatial lag model approaches the standard regression model.

[ Asymptotic properties like consistency, asymptotic efficiency and asymptotic normality of the ML estimates will only hold if regularity conditions are satisfied. Particularly the Jacobian J is required to be positive: $J = \left|\mathbf{I}-\rho\mathbf{W}\right| > 0$ ]

The ML estimator for β results from the first partial derivative of the log likelihood function ln L with respect to β:

(5.31a) $$\hat{\boldsymbol{\beta}}_{ML} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I} - \rho\mathbf{W})\mathbf{y}.$$

The ML estimator for β depends on the unknown autoregressive parameter ρ. It can be decomposed in the OLS estimator $\hat{\boldsymbol{\beta}}_O$ of a regression of **y** on **X** and the OLS estimator $\hat{\boldsymbol{\beta}}_L$ of a regression of **Wy** on **X**:

(5.31b) $$\hat{\boldsymbol{\beta}}_{ML} = \underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}}_{=\hat{\beta}_O} - \rho\underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}}_{=\hat{\beta}_L} = \hat{\boldsymbol{\beta}}_O - \rho\cdot\hat{\boldsymbol{\beta}}_L \quad .$$

The vector of ML residuals $e_{ML}$ also depends on ρ and can be similarly decomposed:

(5.32)
$$\begin{aligned}\mathbf{e}_{ML} &= \mathbf{y} - \rho\mathbf{W}\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{ML} \\ &= \mathbf{y} - \rho\mathbf{W}\mathbf{y} - \mathbf{X}(\hat{\boldsymbol{\beta}}_O - \rho\cdot\hat{\boldsymbol{\beta}}_L) \\ &= \underbrace{\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_O}_{=e_O} - \rho\underbrace{(\mathbf{W}\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_L)}_{=e_L} = \mathbf{e}_O - \rho\cdot\mathbf{e}_L \quad .\end{aligned}$$

Thus, according to the first order condition, the error variance $\sigma^2$ can estimated by

(5.33) $$\hat{\sigma}^2_{ML} = \frac{1}{n}(\mathbf{e}'_{ML}\cdot\mathbf{e}_{ML}) = \frac{1}{n}(\mathbf{e}_O - \rho\cdot\mathbf{e}_L)'(\mathbf{e}_O - \rho\cdot\mathbf{e}_L) \quad .$$

Substituting β and $\sigma^2$ by their estimates (5.31) and (5.33) into the likelihood function(5.30) yields the concentrated loglikelihood function that is nonlinear in ρ:

(5.34) $$\ln L_c(\rho) = C - \frac{n}{2}\cdot\ln[(\mathbf{e}_O - \rho\mathbf{e}_L)'(\mathbf{e}_O - \rho\mathbf{e}_L)] + \ln|\mathbf{I} - \rho\mathbf{W}| \quad .$$

- **Estimation procedure**

Preliminary values for $\beta$ and $\sigma^2$ are estimated for the log likelihood function only depending on the autoregressive parameter $\rho$:

**1st step**: Regression of **y** only on **X**

OLS estimator: $\hat{\boldsymbol{\beta}}_O = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$

Residual vector: $\hat{\mathbf{e}}_O = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_O$

**2nd step**: Regression of **Wy** only on **X**

OLS estimator: $\hat{\boldsymbol{\beta}}_L = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}$

Residual vector: $\hat{\mathbf{e}}_L = \mathbf{W}\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_L$

**3rd step**: Maximisation of the concentrated likelihood function $L_C$ given $\hat{\mathbf{e}}_O$ and $\hat{\mathbf{e}}_L$ to obtain an ML estimator $\hat{\rho}_{ML}$ for the autoregressive parameter $\rho$:

$$\max_{\rho} \quad \ln L_c(\rho) = C - \frac{n}{2} \cdot \ln\left[\frac{1}{n}(\mathbf{e}_O - \rho\mathbf{e}_L)'(\mathbf{e}_O - \rho\mathbf{e}_L)\right] + \ln\left|\mathbf{I} - \rho\mathbf{W}\right|$$

**4rth step**: ML estimates for $\beta$ and $\sigma^2$

$$\hat{\boldsymbol{\beta}}_{ML} = \hat{\boldsymbol{\beta}}_O - \hat{\rho}_{ML} \cdot \hat{\boldsymbol{\beta}}_L \quad \text{and} \quad \hat{\sigma}^2_{ML} = \frac{1}{n}(\mathbf{e}_O - \hat{\rho}_{ML} \cdot \mathbf{e}_L)'(\mathbf{e}_O - \hat{\rho}_{ML} \cdot \mathbf{e}_L)$$

**Example:**

We use the data on output growth (X) and productivity growth (Y) of the 5-region example in order to illustrate maximum likelihood (ML) estimation of the mixed regressive, spatial autoregressive model:

| Region | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Output growth (X) | 0.6 | 1.0 | 1.6 | 2.6 | 2.2 |
| Productivity growth (Y) | 0.4 | 0.6 | 0.9 | 1.1 | 1.2 |

The extended spatial lag model presumes that regional productivity growth is determined by own region‘s output growth and productivity growth in neighbouring regions

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

with $x_{i1}=1$ for all i and $x_{i2} = x_i$. The endogenous spatial lag may capture regional productivity spilllovers suggested by endogenous growth theory.

## ML estimation in the extended spatial lag model

**1st step**: Regression of **y** only on **X** (see section 4.1)

OLS estimator: $\hat{\boldsymbol{\beta}}_O = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix} 1.1412 & -0.5882 \\ -0.5882 & 0.3676 \end{bmatrix}\begin{bmatrix} 4.20 \\ 7.78 \end{bmatrix} = \begin{bmatrix} 0.2165 \\ 0.3897 \end{bmatrix}$

Residual vector: $\mathbf{e}_O = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_O = \begin{bmatrix} 0.4 \\ 0.6 \\ 0.9 \\ 1.1 \\ 1.2 \end{bmatrix} - \begin{bmatrix} 0.4503 \\ 0.6062 \\ 0.8400 \\ 1.2297 \\ 1.0738 \end{bmatrix} = \begin{bmatrix} -0.0503 \\ -0.0062 \\ 0.0600 \\ -0.1297 \\ 0.1262 \end{bmatrix}$

**2nd step**: Regression of **Wy** only on **X**

$$\underset{5x1}{\mathbf{Wy}} = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \\ 1/3 & 1/3 & 0 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 0.4 \\ 0.6 \\ 0.9 \\ 1.1 \\ 1.2 \end{bmatrix} = \begin{bmatrix} 0.75 \\ 0.8 \\ 0.7 \\ 0.9 \\ 1.1 \end{bmatrix} \quad \text{and} \quad \underset{5x2}{\mathbf{X}} = \begin{bmatrix} 1 & 0.6 \\ 1 & 1.0 \\ 1 & 1.6 \\ 1 & 2.6 \\ 1 & 2.2 \end{bmatrix}$$

$$\mathbf{X'Wy} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0.6 & 1.0 & 1.6 & 2.6 & 2.2 \end{bmatrix} \begin{bmatrix} 0.75 \\ 0.8 \\ 0.7 \\ 0.9 \\ 1.1 \end{bmatrix} = \begin{bmatrix} 4.25 \\ 7.13 \end{bmatrix}$$

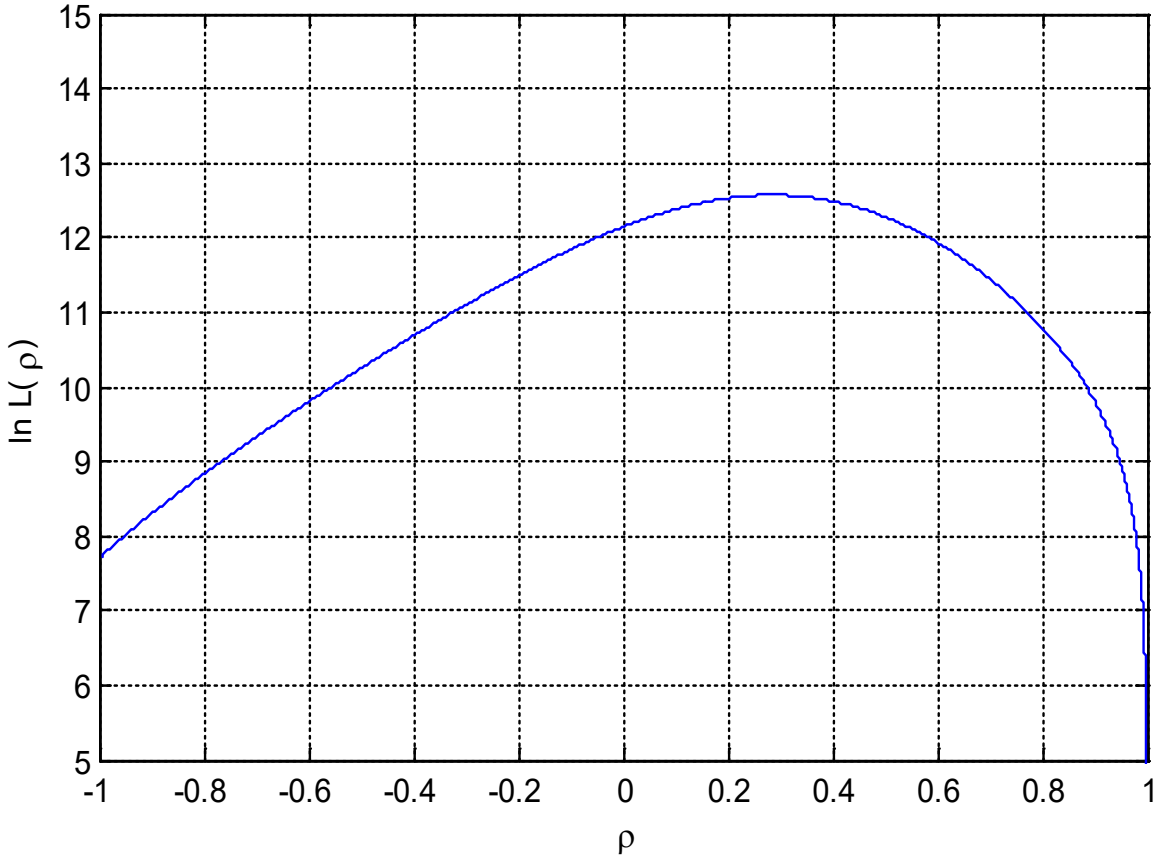
OLS estimator: $\hat{\boldsymbol{\beta}}_{\mathrm{L}} = (\mathbf{X'X})^{-1}\mathbf{X'Wy} = \begin{bmatrix} 1.1412 & -0.5882 \\ -0.5882 & 0.3676 \end{bmatrix} \begin{bmatrix} 4.25 \\ 7.13 \end{bmatrix} = \begin{bmatrix} 0.6562 \\ 0.1211 \end{bmatrix}$

$$\mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{L}} = \begin{bmatrix} 1 & 0.6 \\ 1 & 1.0 \\ 1 & 1.6 \\ 1 & 2.6 \\ 1 & 2.2 \end{bmatrix} \begin{bmatrix} 0.6562 \\ 0.1211 \end{bmatrix} = \begin{bmatrix} 0.7289 \\ 0.7773 \\ 0.8500 \\ 0.9711 \\ 0.9226 \end{bmatrix}$$

Residual vector: $\mathbf{e}_{\mathrm{L}} = \mathbf{Wy} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{L}} = \begin{bmatrix} 0.75 \\ 0.8 \\ 0.7 \\ 0.9 \\ 1.1 \end{bmatrix} - \begin{bmatrix} 0.7289 \\ 0.7773 \\ 0.8500 \\ 0.9711 \\ 0.9226 \end{bmatrix} = \begin{bmatrix} 0.0211 \\ 0.0227 \\ -0.1500 \\ -0.0711 \\ 0.1774 \end{bmatrix}$

**3rd step**: Maximisation of the concentrated likelihood function $L_C$ given $\hat{\mathbf{e}}_O$ and $\hat{\mathbf{e}}_L$ to obtain an ML estimator $\hat{\rho}_{ML}$ for the autoregressive parameter $\rho$

Figure: Log $L_c$ function of the extended spatial lag model




ML estimator for $\rho$:

$$\hat{\rho}_{ML} = 0.2872$$

Log Lc function at $\hat{\rho}_{ML} = 0.2872$:

$\ln L_c^* = 12.5729$

(neglecting the constant C)

## MATLAB code for plotting the ln LC function of the extended spatial lag model

```
% Program LCLAGPLOT
%
% Plot of the concentrated log likelihood function
% (ln Lc) of the extended spatial lag model
%
% Input:
% Residual vectors: eO and eL
% Spatial weights matrix: W
% Number of regions: n
% Output:
% Coordiate system
% x-axis: Autoregressive parameter rho
% y-axis: Concentrated log likelihood function (ln Lc)
% Autoregressive parameter: rho
% Loglikelihood function at rho: Lcmax
%-----------------------------------------------------------------
%  Plot of Concentrated log likelihood function (ln Lc)
%-----------------------------------------------------------------
n = 5;
eO = [-0.0503; -0.0062; 0.0600; -0.1297; 0.1262];
eL = [0.0211; 0.0227; -0.1500; -0.0711; 0.1774];
eOt = eO';
eLt = eL';
I5 = eye(5);
W = [0 0.5 0.5 0 0;1/3 0 1/3 1/3 0;1/3 1/3 0 1/3 0;0 1/3 1/3 0 1/3;0 0 0 1
0];
```

```
x=-1:0.0001:1;
Lc=x;
rho = -0.9999;
for j=1:20000
    if j<10000
        d=-(10000-j)/10000;
    else
        d = (j-1-10000)/10000;
    end
    Lc(j) = - n/2*log((eOt-d*eLt)*(eO-d*eL)/n) + log(det(I5-d*W));
    if j==1
        Lcmax = Lc(1);
    end
    if Lc(j)>Lcmax
        rho = d;
        Lcmax = Lc(j);
    end
end
plot(x,Lc);
grid on;
xlabel(' \rho');
ylabel('ln L( \rho)');
'rho = ', rho, 'ln Lc(rho) = ', Lcmax
```

# R code for plotting the ln LC function of the extended spatial lag model

```
# Program LCLAGPLOT
#
# Plot of the concentrated log likelihood function
# (ln Lc) of the spatial lag model (SAR model)
#
# Input:
# Residual vectors: eO and eL
# Spatial weights matrix: W
# Number of regions: n
# Output:
# Coordiate system
# x-axis: Autoregressive parameter rho
# y-axis: Concentrated log likelihood function (ln Lc)
# Autoregressive parameter: rho
# Loglikelihood function at rho: Lcmax
# ------------------------------------------------------------------
# Plot of Concentrated log likelihood function (ln Lc)
# ------------------------------------------------------------------
n = 5;
eO = c(-0.0503, -0.0062, 0.0600, -0.1297, 0.1262)
eL = c(0.0211, 0.0227, -0.1500, -0.0711, 0.1774)
I5 = diag(5)
W = matrix(c(0, 0.5, 0.5, 0, 0, 1/3, 0, 1/3, 1/3, 0, 1/3, 1/3, 0, 1/3, 0, 0, 1/3, 1/3, 0, 1/3, 0, 0, 0, 1,
0), nrow=5, ncol=5, byrow=TRUE)
```

```
x = seq(-1, 1, by=0.0001)
Lc = x
rho = -0.9999
for (j in 1:20000) {
    if (j<10000) {
     d =-(10000-j)/10000
  } else {
     d = (j-1-10000)/10000
  }
  Lc[j] = -n/2*log((eO-d*eL)%*%(eO-d*eL)/n) + log(det(I5-d*W))
    if (j==1) {
    Lcmax = Lc[1]
  }
    if (Lc[j]>Lcmax) {
    rho = d;
    Lcmax = Lc[j]
  }
  }
plot(x, Lc, main="Log Lc function of the spatial lag model", cex.main=1.0, xlab="rho",
ylab="ln L(rho)", type="l")
```

**4rth step**: ML estimates for $\beta$ and $\sigma^2$

ML estimator for $\beta$:

$$\hat{\boldsymbol{\beta}}_{ML} = \hat{\boldsymbol{\beta}}_{O} - \hat{\rho}_{ML} \cdot \hat{\boldsymbol{\beta}}_{L} = \begin{bmatrix} 0.2165 \\ 0.3897 \end{bmatrix} - 0.2872 \cdot \begin{bmatrix} 0.6562 \\ 0.1211 \end{bmatrix} = \begin{bmatrix} 0.0280 \\ 0.3549 \end{bmatrix}$$

ML estimator of the error variance:

$$\mathbf{e}_{ML} = \mathbf{e}_{O} - \hat{\rho}_{ML} \cdot \mathbf{e}_{L} = \begin{bmatrix} -0.0503 \\ -0.0062 \\ 0.0600 \\ -0.1297 \\ 0.1262 \end{bmatrix} - 0.2872 \cdot \begin{bmatrix} 0.0211 \\ 0.0227 \\ -0.1500 \\ -0.0711 \\ 0.1774 \end{bmatrix} = \begin{bmatrix} -0.0564 \\ -0.0127 \\ 0.1031 \\ -0.1093 \\ 0.0753 \end{bmatrix}$$

$$\hat{\sigma}^2_{ML} = \frac{1}{n}(\mathbf{e}_{O} - \hat{\rho}_{ML} \cdot \mathbf{e}_{L})'(\mathbf{e}_{O} - \hat{\rho}_{ML} \cdot \mathbf{e}_{L})$$

$$= \frac{1}{5} \cdot \begin{bmatrix} -0.0564 & -0.0127 & 0.1031 & -0.1093 & 0.0753 \end{bmatrix} \begin{bmatrix} -0.0564 \\ -0.0127 \\ 0.1031 \\ -0.1093 \\ 0.0753 \end{bmatrix}$$

$$= \frac{1}{5} \cdot 0.0316 = 0.0063$$