University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

2. Multiple Regression Model

Prof. Dr. Miroslav Verbič

miroslav.verbic@ef.uni-lj.si

www.miroslav-verbic.si

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION of MBAs ACCREDITED

Ljubljana, October 2025

Models and Variables

MULTIPLE REGRESSION MODEL
MULTIVARIATE REGRESSION MODEL

Single equation models

dependent variable	explanatory variables
explained variable	unexplained variables
regressand	regressors
predictand	predictors
response variable	control variables

Multiple equation models

endogenous variable(s)	exogenous variables

Notation of Variables

$y_i, (y_t)$ – *value of the dependent variable at i-th (t-th) observation*

$x_{ji}, (x_{jt})$ – *value of the j-th explanatory variable at the i-th (t-th) observation*

$$\boxed{\begin{gathered} i = 1, 2, \ldots, n \quad (t = 1, 2, \ldots, T) \\ j = 1, 2, \ldots, k \end{gathered}}$$

n (T)* – *number of observations
k* – *number of parameters

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

2.1 Population Regression Model and Sample Regression Model

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

Population Regression Model

POPULATION REGRESSION MODEL
DATA GENERATING PROCESS (DGP)

Example: *PRODUCT = f (LABOUR, CAPITAL)*

$$E(Q|L,K) = f(L,K)$$

Conditional expected value of the dependent variable is a function of explanatory variables

Expected value = mathematical expectation = average value of a stochastic variable given an infinite number of measurements or realisations

Population Regression Model

Linear population regression model

(regression model = regression function = regression line = regression)

$$E(Q_i \mid L_i, K_i) = \beta_1 + \beta_2 L_i + \beta_3 K_i$$

$$Q_i = E(Q_i \mid L_i, K_i) + u_i$$

or alternatively

$$u_i = Q_i - E(Q_i \mid L_i, K_i)$$

$$Q_i = \beta_1 + \beta_2 L_i + \beta_3 K_i + u_i$$

***u_i* = disturbances (stochastic variable)**

Population Regression Model

Reasons for including u_i in the model:

- First, it is a **substitute** for all those variables that affect the dependent variable, but are *not included among the explanatory variables* (indicate some for our example!). Often this is caused by an incomplete or insufficient (economic) theory.
- Even if some variables are recognized as important explanatory variables, they often can not be included in the model specification, as they are *difficult to express numerically or obtain data* (e.g. preferences and habits, or in our case idiosyncratic knowledge).
- When there is a high probability that the total effect of omitted explanatory variables is small and insignificant, but most of all *non-systematic*, then these effects can be treated as random and captured in the model by a stochastic variable u.

Population Regression Model

- Even if one could capture all relevant explanatory variables when specifying the regression model, some random elements of the dependent variable would still be left unexplained – *pure stochastic effects or shocks*.
- Even though the classical linear regression model assumes that the values of (dependent and explanatory) *variables are measured without error*, this is often not the case in practice.
- When specifying the regression model, we should stick to the well-known rule, stating that *simpler model should be preferred* to a more complicated one, until proven that it is unsuitable due to its simplicity (Occam's razor principle).

Population Regression Model

GENERAL POPULATION REGRESSION MODEL (PRM):

$$E(y_i | x_{1i},\ldots,x_{ki}) = f(x_{1i},\ldots,x_{ki})$$

or alternatively

$$y_i = E(y_i | x_{1i},\ldots,x_{ki}) + u_i$$

Population Regression Model

Linear population regresion model:

$$E(y_i | x_{1i},\ldots,x_{ki}) = \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki}$$

$$E(y_i | x_{1i},\ldots,x_{ki}) = \beta_1 + \beta_2 x_{2i} \ldots + \beta_k x_{ki}$$

or alternatively

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + u_i$$

$$y_i = \beta_1 + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + u_i$$

β_j – ***partial regression coefficient*** **of the *j*-th explanatory variable**

u_i – ***disturbances (stochastic variable)*** **at the *i*-th observation**

Sample Regression Model

LINEAR SAMPLE REGRESSION MODEL (SRM):

$$y_i = b_1 + b_2 x_{2i} + \ldots + b_k x_{ki} + e_i$$

$$\hat{y}_i = b_1 + b_2 x_{2i} + \ldots + b_k x_{ki}$$

$$y_i = \hat{y}_i + e_i$$

Fitted values

$\hat{y}_i$ – estimator of the conditional expectation $E(y|x_{1i}, \ldots, x_{ki})$

$b_1, \ldots, b_k$ – estimators of regression coefficients $\beta_1, \ldots, \beta_k$

e_i – residuals of the sample regression model

The term estimator is used in statistical literature for sample statistics. These are formulas that tell us how to calculate the parameters for population, based on (stochastic) sample data.
A given, specific numerical value of a parameter, calculated with the estimator, is then called an estimate of a parameter.

Sample Regression Model

Linear sample regression model:

$$\hat{y}_i = b_1 x_{1i} + b_2 x_{2i} + \ldots + b_k x_{ki}$$

$$\hat{y}_i = b_1 + b_2 x_{2i} + \ldots + b_k x_{ki}$$

or alternatively

$$y_i = b_1 x_{1i} + b_2 x_{2i} + \ldots + b_k x_{ki} + e_i$$

$$y_i = b_1 + b_2 x_{2i} + \ldots + b_k x_{ki} + e_i$$

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

2.2 Estimation of the Population Regression Model and the Method of Least Squares

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

Motivation

"The only way to a position in which our science might give positive advice in a large scale to politicians and business men, leads through quantitative work. For as long as we are unable to put our arguments into figures, the voice of our science, although occasionally it may help to dispel gross errors, will never be heard by practical men. They are, by instinct, econometricians all of them, in their distrust of anything not amenable to exact proof."

Joseph A. Schumpeter: "The Common Sense of Econometrics", *Econometrica*, 1, 1933, p. 12.

Estimation of the PRM

How should the estimator minimise the residuals of the SRM:

- **should sum of the residuals be minimal?**
- **should sum of the absolute values of residuals be minimal?**
- **should sum of the squared residuals be minimal?**

Least Squares Estimator or Method (OLS, LS)

"Discovered" by the German mathematician C. F. Gauss (at age 16). First published in 1809 and finalised in 1823.

Estimation of the PRM

Carl Friederich Gauss (1777–1855)

Estimation of the PRM

$$y_i = b_1 + b_2 x_{2i} + \ldots + b_k x_{ki} + e_i = \hat{y}_i + e_i$$

Residuals of the sample regression model:

$$e_i = y_i - \hat{y}_i =$$
$$= y_i - b_1 - b_2 x_{2i} - \ldots - b_k x_{ki}$$

$$\sum \mathrm{e}_{\mathrm{i}}^2 = \sum (y_i - b_1 - b_2 x_{2i} - \ldots - b_k x_{ki})^2$$

It is obvious that this sum is a function (*S*) of estimators of regression coefficients:

$$\sum \mathrm{e}_{\mathrm{i}}^2 = S(b_1, \ldots, b_k)$$

Method of Least Squares

Linear sample regression model (SRM):

$$y_i = b_1 + b_2 x_{2i} + \ldots + b_k x_{ki} + e_i$$

$$\begin{array}{ll}
i = 1: & y_1 = b_1 + b_2 x_{21} + \ldots + b_k x_{k1} + e_1 \\
i = 2: & y_2 = b_1 + b_2 x_{22} + \ldots + b_k x_{k2} + e_2 \\
\vdots & \quad \vdots \qquad\qquad\qquad \vdots \\
i = n: & y_n = b_1 + b_2 x_{2n} + \ldots + b_k x_{kn} + e_n
\end{array}$$

Method of Least Squares

Least squares algebra and derivation of OLS:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b}$$

$$\mathbf{e}^{T}\mathbf{e} = (\mathbf{y} - \mathbf{X}\mathbf{b})^{T}(\mathbf{y} - \mathbf{X}\mathbf{b})$$

Method of Least Squares

$$\frac{\partial \mathbf{e}^{T}\mathbf{e}}{\partial \mathbf{b}} = -2\mathbf{X}^{T}\mathbf{y} + 2\mathbf{X}^{T}\mathbf{X}\mathbf{b} = 0$$

$$\mathbf{X}^{T}\mathbf{y} = \mathbf{X}^{T}\mathbf{X}\mathbf{b}$$

Least squares estimator of regression coefficients:

$$\boxed{\mathbf{b} = (\mathbf{X}^{T}\mathbf{X})^{-1}\mathbf{X}^{T}\mathbf{y}}$$

Properties of Least Squares Estimator

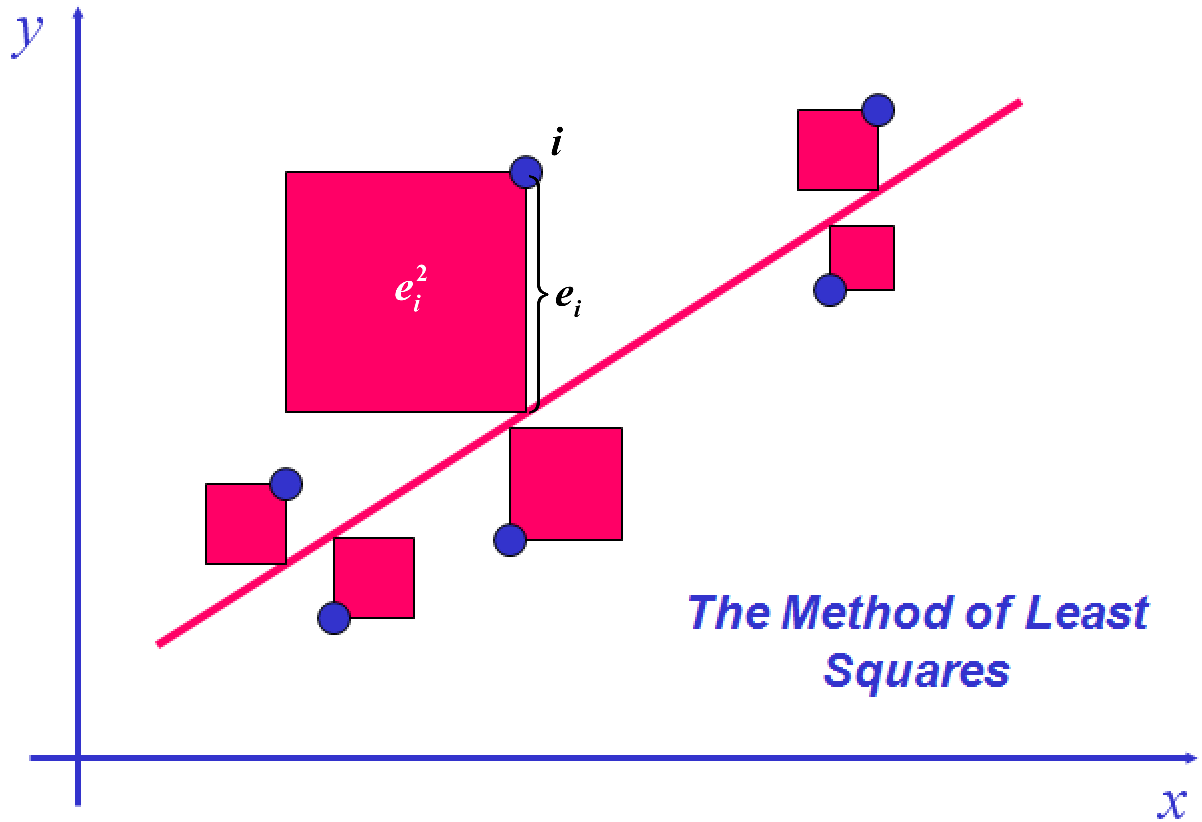
1. $\mathbf{X}^{\mathrm{T}}\mathbf{e} = \mathbf{0}$

2. $\hat{\mathbf{y}}^{\mathrm{T}}\mathbf{e} = \mathbf{0}$

3. $\overline{\hat{y}} = \overline{y}$

4. $\sum_i e_i = 0$

Graphic Illustration of OLS



University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

2.3 Assumptions of the Method of Least Squares

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION OF MBAS AMBA ACCREDITED

Assumption 1

Linearity of regression model:

$$y_i = \beta_1 + \beta_2 x_i + u_i$$

Assumption 2

Fixed (non-stochastic) values of explanatory variables at sampling (drawing samples from a distribution):

CONDITIONAL REGRESSION ANALYSIS

Assumption 3

Zero conditional mathematical expectation of u_i:

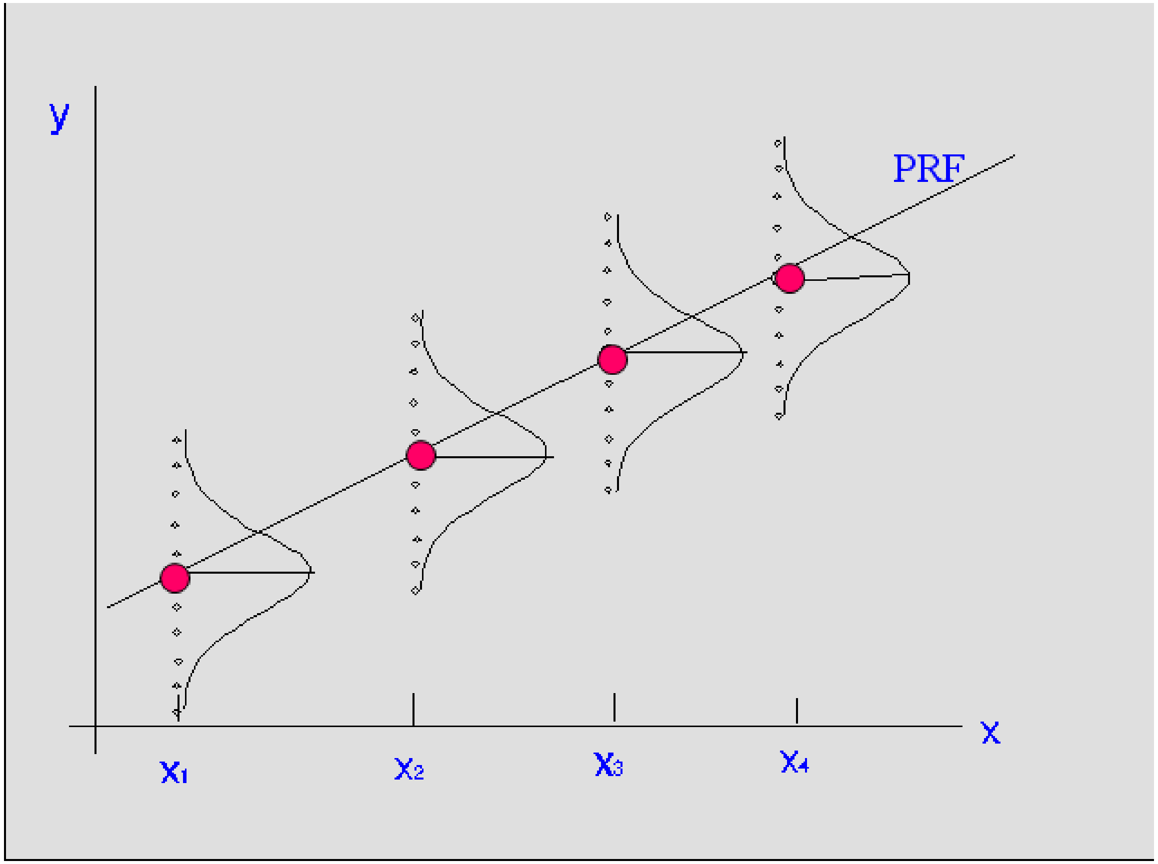
$$E(u_i \mid x_{2i}, \dots, x_{ki}) = 0 \quad \text{or in short } E(u_i) = 0;$$

which also means:

$$E(y_i \mid x_{2i}, \dots, x_{ki}) = \beta_1 + \beta_2 x_{2i} + \dots + \beta_k x_{ki}.$$

Assumption 3

Distribution of stochastic variables y and u:



Assumption 4

Homoscedasticity:

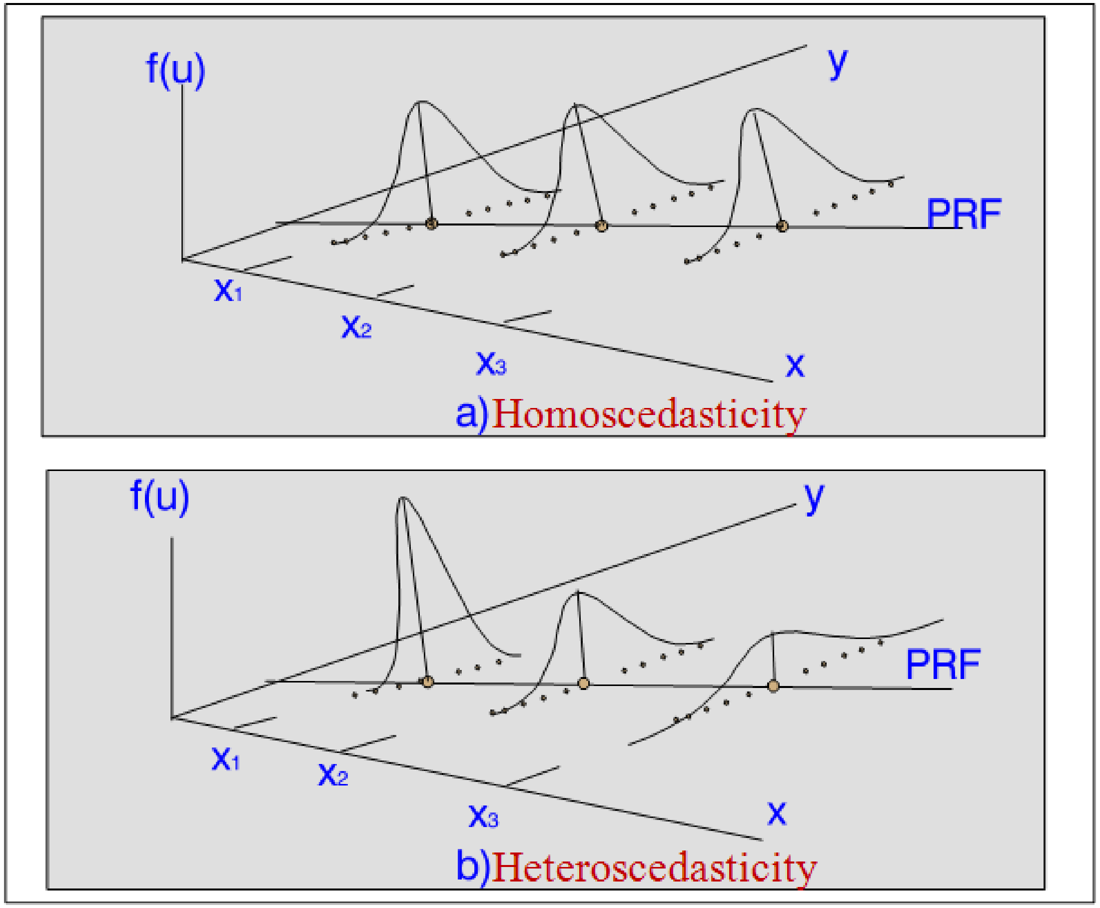
$$Var(u_i|x_i) = E\left[\left(u_i - E(u_i|x_i)\right)^2 \middle| x_i\right] = E\left[u_i^2 \middle| x_i\right] = E(u_i^2) = \sigma^2$$

or alternatively:

$$Var(u_i) = E(u_i^2) = \sigma^2$$

Assumption 4

Variability of stochastic variable *u*:



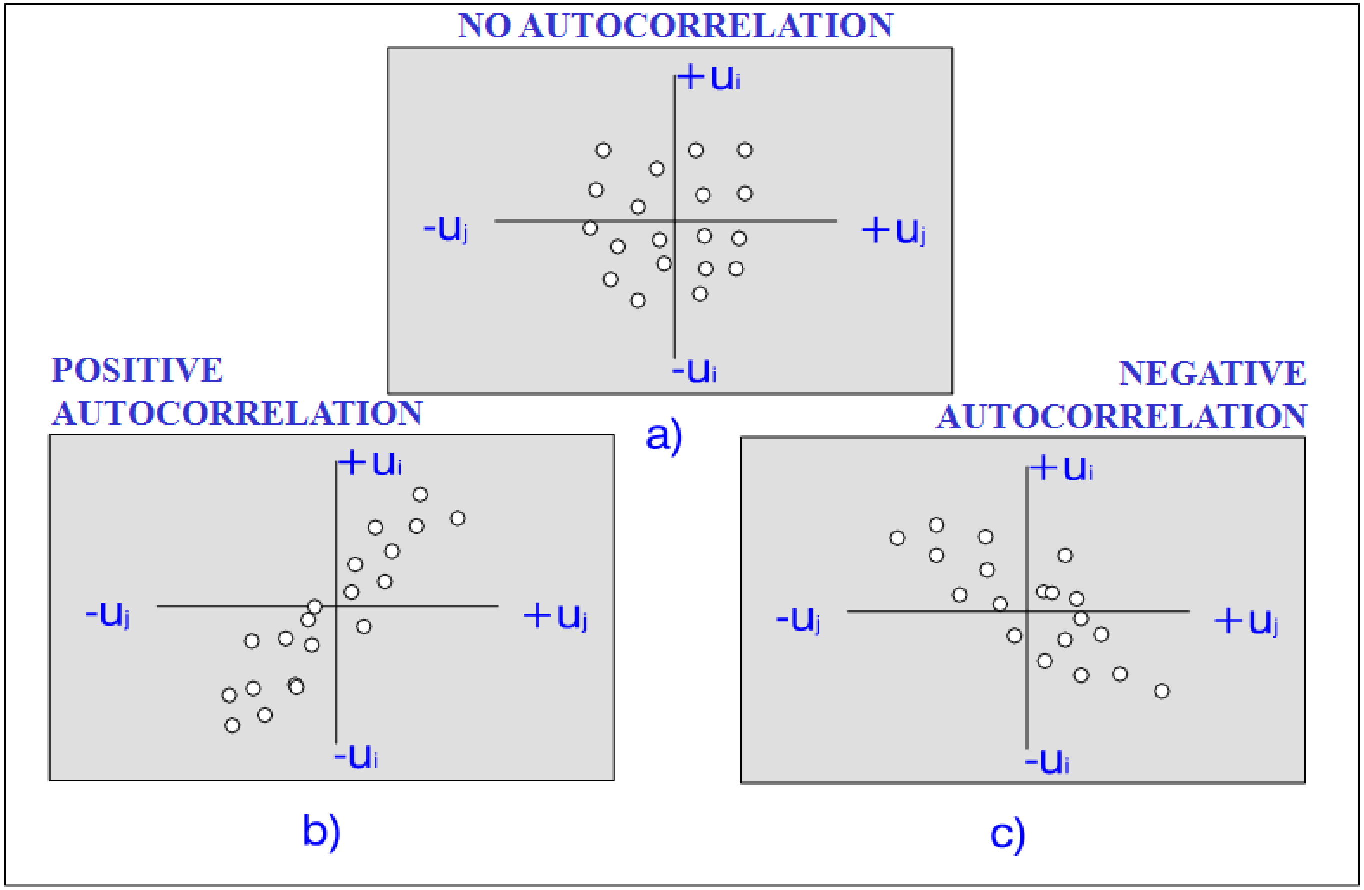
Assumption 5

No autocorrelation:

$$Cov(u_i, u_j \mid x_i, x_j) = 0; \quad i \neq j$$

Assumption 5

Scatter plots of (lagged) stochastic variable u:



Assumption 6

No correlation between explanatory variables and the stochastic variable *u*:

$$Cov(x_2, u) = Cov(x_3, u) = \ldots = Cov(x_k, u) = 0$$

Assumption 7

Number of observations should be greater than the number of estimated parameters (explanatory variables):

$$n > k$$

Assumption 8

Variability in values of explanatory variables:

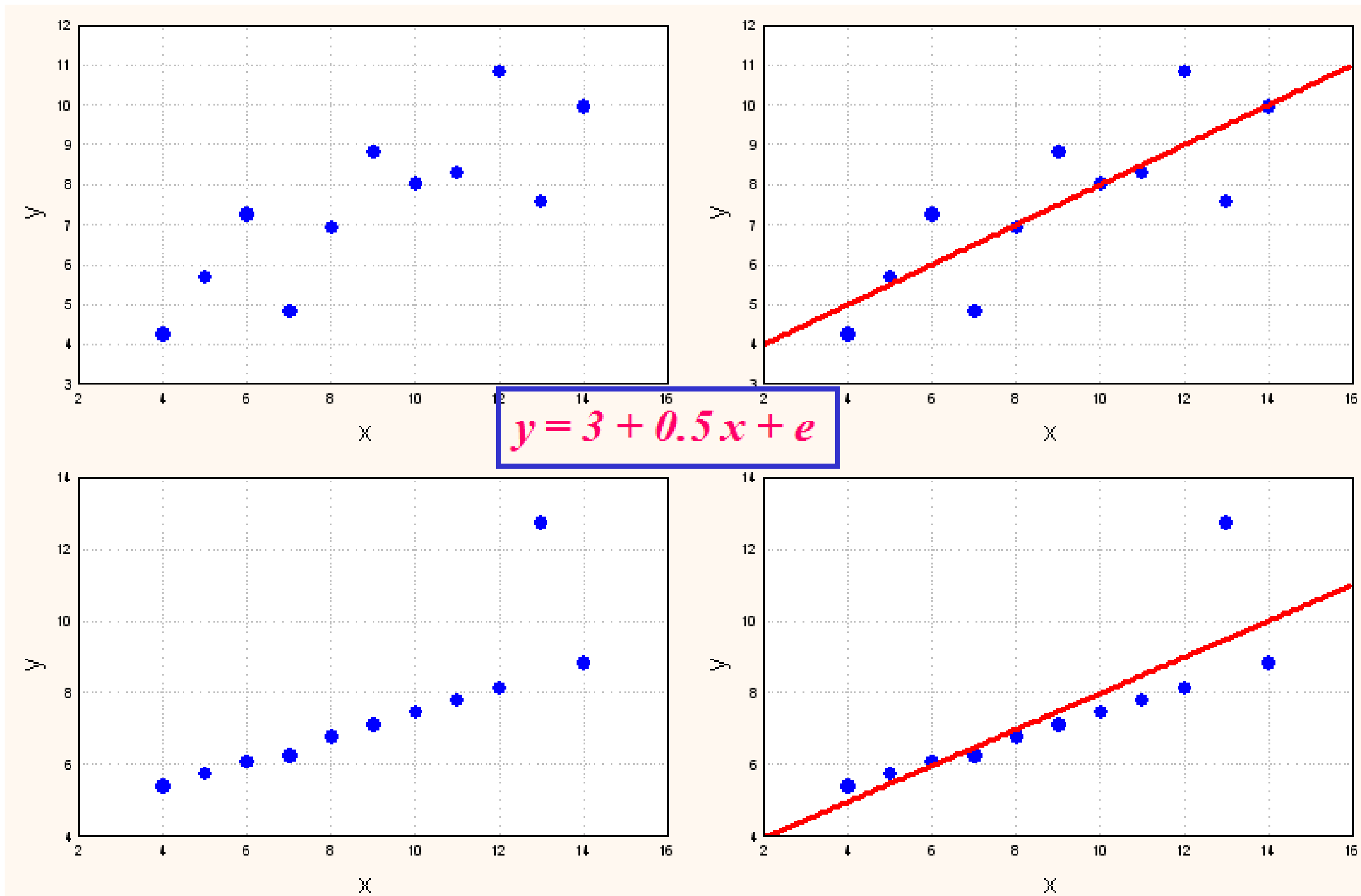
var(*X*) is a positive finite value

Assumption 9

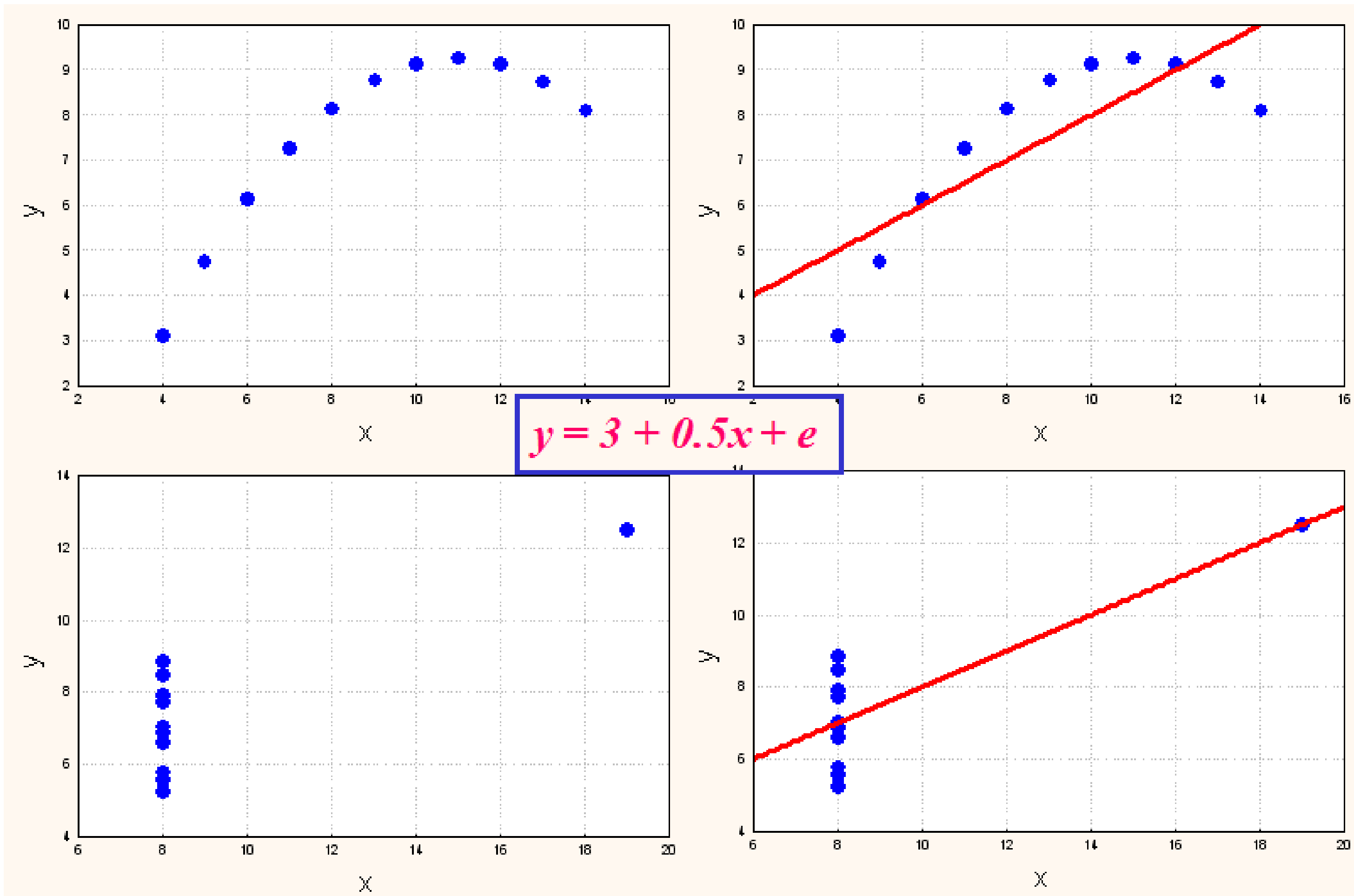
Regression model is correctly specified:

- **All relevant explanatory variables are included**
- **The right functional form of the model is chosen**

Assumption 9



Assumption 9



Assumption 10

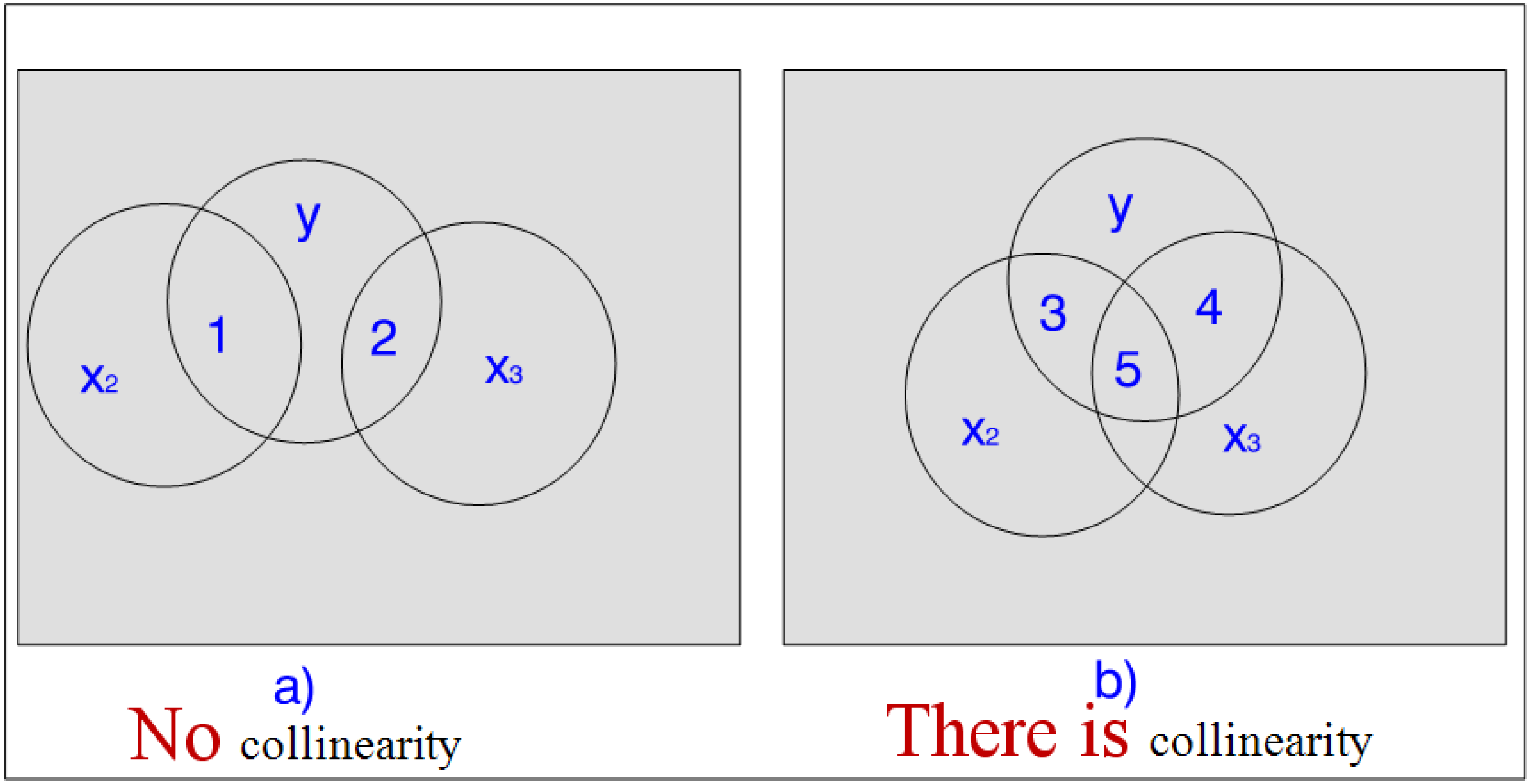
Absence of perfect multicollinearity:

There exists no perfect *linear* relationship among the explanatory variables of the form:

$$\lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_2 + \ldots + \lambda_k \mathbf{x}_k = \mathbf{0}$$

Assumption 10

Analogy with Venn diagrams:



Operationalisation of the PRM

Stochastic variable *u* is normally distributed:

$$u_i \sim N(0, \sigma_u^2)$$

Consequently, the dependent variable *y* is also a normally distributed stochastic variable:

$$y_i \sim N(\beta_1 x_{1i} + \ldots + \beta_k x_{ki}, \sigma_u^2)$$

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

2.4 Sample Properties of the Method of Least Squares

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION OF MBAS ACCREDITED

Motivation

- The estimates of regression coefficients, obtained by the least squares estimator, are *stochastic variables*. How to establish their mean, variance, covariances, and probability distribution?

- The method of least squares is only one of possible estimators of regression coefficients. How *successful is the chosen estimator* compared to other possible stimators?

- What is the *reliability of estimates* of the regression coefficients? How good are the estimates, obtained based on one sample only?

Sample Properties of OLS

PROPERTY A: LINEARITY

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Estimates of regression coefficients based on sample data are linear combinations of sample values of the dependent variable *y*.

PROPERTY B: CONSISTENCY

$$\mathbf{b} \xrightarrow{n \to \infty} \beta$$

Estimates of regression coefficients tend to the true values of parameters, when the sample size increases.

Sample Properties of OLS

PROPERTY C: UNBIASEDNESS

$$E(\mathbf{b}) = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u})$$
$$= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{0} = \beta$$

Estimates of regression coefficients are unbiased estimates of their population values.

Sample Properties of OLS

PROPERTY D: EFFICIENCY

$$Var-Cov(\mathbf{b})=\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

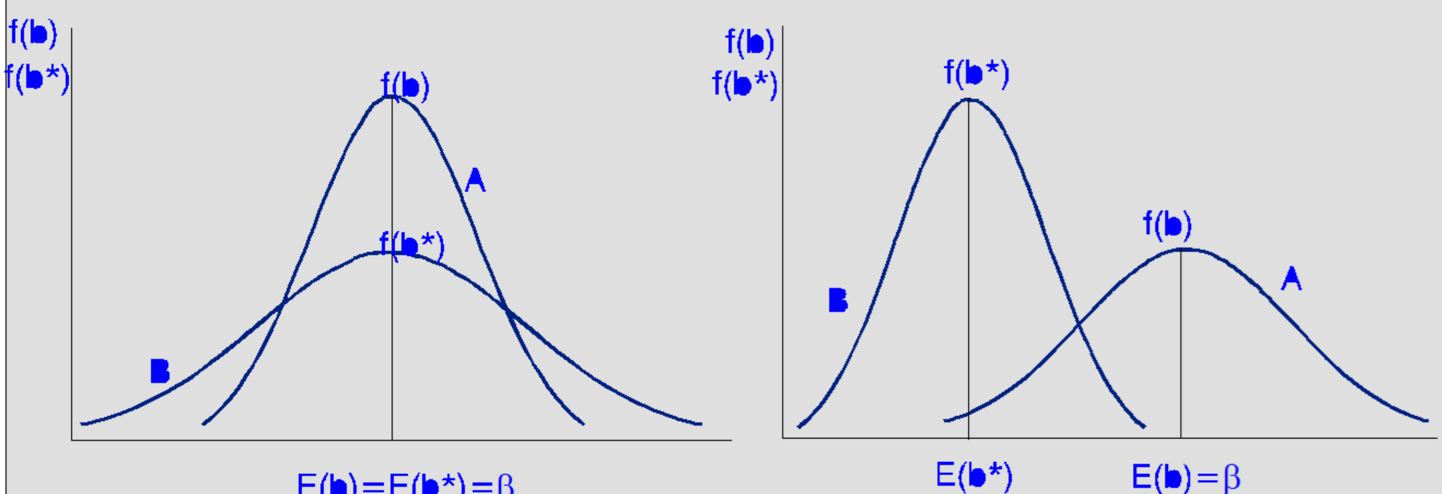
Estimates of regression coefficients are efficient.

$$Var-Cov(\mathbf{b})=\begin{bmatrix} Var(b_1) & Cov(b_1,b_2) & \cdots & Cov(b_1,b_k) \\ Cov(b_2,b_1) & Var(b_2) & \cdots & Cov(b_2,b_k) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(b_k,b_1) & Cov(b_k,b_2) & \cdots & Var(b_k) \end{bmatrix}$$

Sample Properties of OLS

Distribution of estimates of regression coefficients based on a large number of repeated samples is normal:

$$\mathbf{b} \sim N\left[\boldsymbol{\beta}, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}\right]$$



a) Both estimators are unbiased, but A is more efficient.

b) Estimator B is more efficient, but biased. Estimator A is unbiased, but less efficient.

Sample Properties of OLS

Variance of the regression coefficient estimates:

$$Var(b_2) = \frac{\sigma^2}{nVar(x_2)} \cdot \frac{1}{(1 - r_{x_2 x_3}^2)}$$

Variance decreases, when:

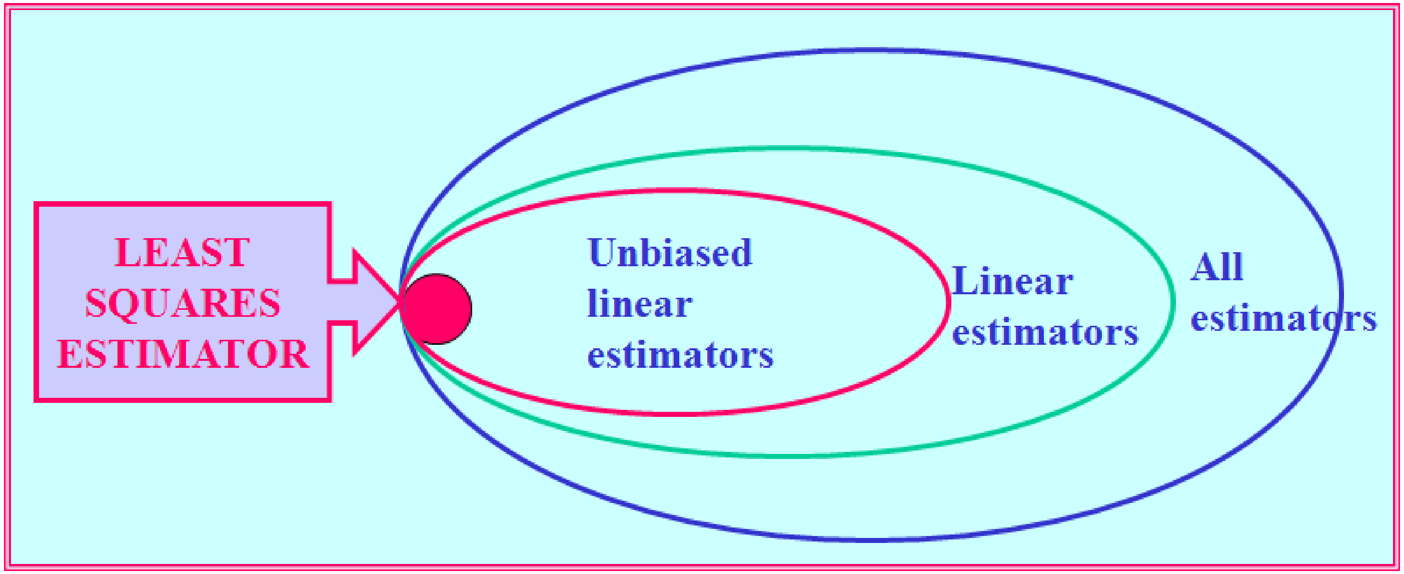
1. **the sample size (n) increases, i.e. when the calculation of a regression coefficient estimate is based on more observations;**
2. **the variability (variance) of an explanatory variable x increases;**
3. **the variability (variance) of the stochastic variable u decreases;**
4. **the linear dependence among the explanatory variables decreases.**

Gauss–Markov Theorem

The method of least squares is

Best **L**inear **U**nbiased **E**stimator.

BLUE



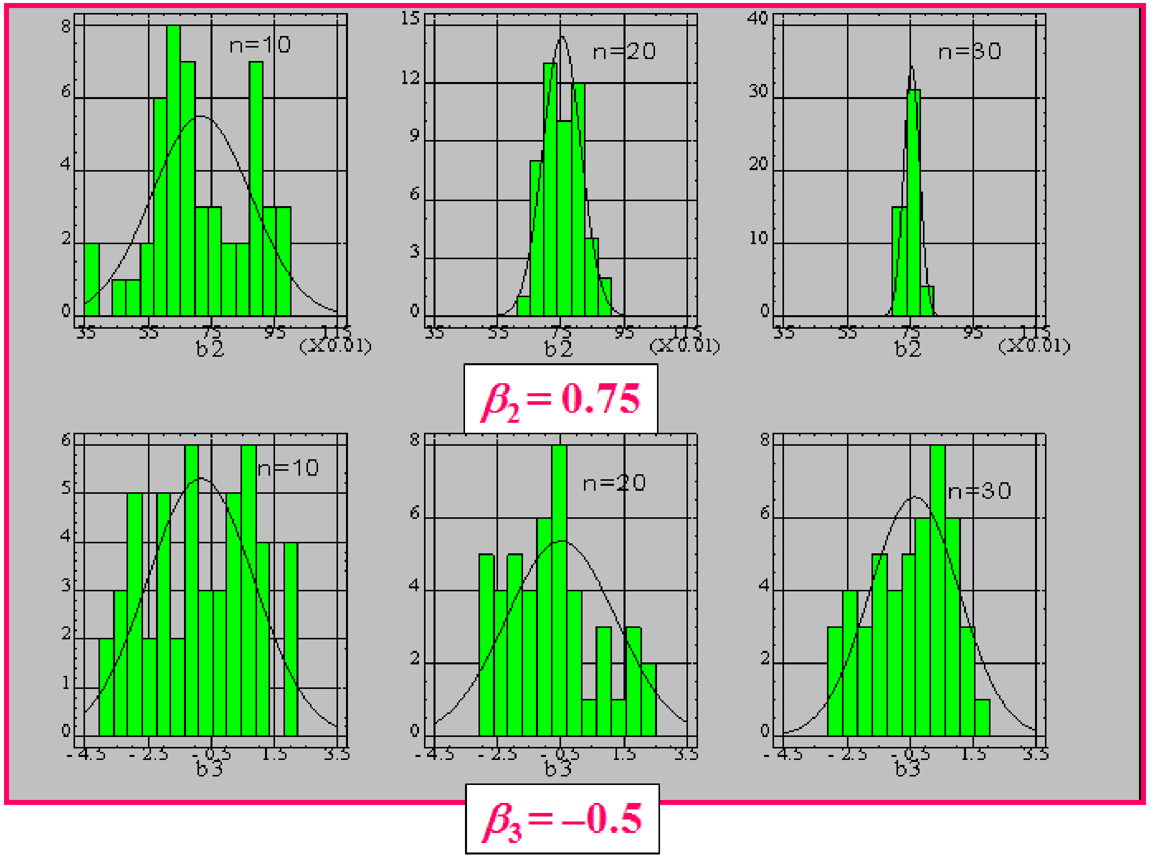
Monte Carlo Experiments

Monte Carlo experiments
(Los Alamos – Manhattan)

- We define a specific form for the PRM;
- We set the values of regression coefficients, values of explanatory values, and calculate the values of the dependent variable;
- We prepare the "mended" values of the dependent variable for a large number of samples, such that at any given sample we add a random value (drawing from a distribution) of the stochastic variable u;
- For each of the prepared samples, we calculate the estimates of regression coefficients by applying OLS;
- We prepare the distributions of estimates of regression coefficients and we analyze them.

Monte Carlo Experiments

Distributions of regression coefficient estimates
(50 sample models; n = 10, n = 20, n = 30)



Variance estimator of disturbances *u*

Variance estimator of disturbances u, σ_u^2:

$$s_e^2 = \frac{\mathbf{e}^{\mathrm{T}}\mathbf{e}}{n-k}$$

$\boxed{s_e^2 \text{ is an unbiased estimator of } \sigma^2}$

Estimator of the var–cov matrix of **b**

Estimator of the var–cov matrix of regression coefficient estimates, Var – Cov(**b**)**:**

$$\text{var}-\text{cov}(\mathbf{b}) = s_e^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$$

Distribution of regression coef. estimates

$$\mathbf{b} \sim N\left[\boldsymbol{\beta}, \sigma^2(\mathbf{X}^\mathrm{T}\mathbf{X})^{-1}\right]$$

$$b_j \sim N(\beta_j, Var(b_j))$$

$$z = \frac{b_j - \beta_j}{\sqrt{Var(b_j)}} \sim N(0,1)$$

Distribution of regression coef. estimates

Calculation of the *t*–statistic based on sample data

$$\boxed{t_j = \frac{b_j - \beta_j}{\sqrt{\mathrm{var}(b_j)}} = \frac{b_j - \beta_j}{se(b_j)}}$$

$$se(b_j) = \sqrt{\mathrm{var}(b_j)} = \sqrt{s_e^2 \left[(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} \right]_{jj}}$$

Distribution of regression coef. estimates

William Sealy Gosset (1876–1937)

Distribution of regression coef. estimates

Student's *t*–distribution:

$$t_j = \frac{b_j - \beta_j}{se(b_j)} \sim t_{(n-k)}$$

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

2.5 Measures of Fit of a Regression Model

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION OF MBAS ACCREDITED

Standard Error of Regression

Standard error of regression

$$s_e = \sqrt{s_e^2} = \sqrt{\frac{\mathbf{e}^{\mathrm{T}}\mathbf{e}}{n-k}}$$

Coefficient of variation

$$KV = \frac{s_e}{\bar{y}} \quad \text{or} \quad KV\% = \frac{s_e}{\bar{y}} 100$$

Analysis of Variance (ANOVA)

Decomposition of sum of squares (*SS*):

$$\boldsymbol{TSS = ESS + RSS}$$

$$TSS = \sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2 = \sum_{i=1}^{n} y_i^2 - n\overline{y}^2 = \mathbf{y}'\mathbf{y} - n\overline{y}^2$$

$$ESS = \sum_{i=1}^{n}\left(\hat{y}_i - \overline{y}\right)^2 = \sum_{i=1}^{n} \hat{y}_i^2 - n\overline{y}^2 = \hat{\mathbf{y}}'\hat{\mathbf{y}} - n\overline{y}^2$$

$$ESS = \mathbf{b}'\mathbf{X}'\mathbf{y} - n\overline{y}^2 = \mathbf{y}'\mathbf{X}\mathbf{b} - n\overline{y}^2$$

$$RSS = \sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 = \sum_{i=1}^{n} e_i^2 = \mathbf{e}'\mathbf{e} = \mathbf{y}'\mathbf{y} - \hat{\mathbf{y}}'\hat{\mathbf{y}}$$

Determination Coefficient

Determination coefficient of multiple regression
Multiple determination coefficient

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

$$R^2 = \frac{\hat{\mathbf{y}}'\hat{\mathbf{y}} - n\bar{y}^2}{\mathbf{y}'\mathbf{y} - n\bar{y}^2} = \frac{\mathbf{b}'\mathbf{X}'\mathbf{y} - n\bar{y}^2}{\mathbf{y}'\mathbf{y} - n\bar{y}^2}$$

$$\boxed{R_*^2 = \frac{\hat{\mathbf{y}}'\hat{\mathbf{y}}}{\mathbf{y}'\mathbf{y}} = \frac{\mathbf{b}'\mathbf{X}'\mathbf{y}}{\mathbf{y}'\mathbf{y}}}$$

$$R^2 = (r_{y\hat{y}})^2 = r_{y\hat{y}}^2$$

Determination Coefficient

Adjusted determination coefficient (H. Theil, 1971)

TSS = ESS + RSS

TSS = (TSS - RSS) + RSS

Degrees of freedom:

$$(n-1) = ((n-1)-(n-k)) + (n-k)$$

$$(n-1) = (k-1) + (n-k)$$

$$\bar{R}^2 = 1 - \frac{\frac{\text{RSS}}{(n-k)}}{\frac{\text{TSS}}{(n-1)}} = 1 - \frac{\text{RSS}}{\text{TSS}}\frac{(n-1)}{(n-k)}$$

$$\boxed{\bar{R}^2 = 1-(1-R^2)\frac{(n-1)}{(n-k)}}$$

Determination Coefficient

Weaknesses of determination coefficient and adjusted determination coefficient:

- A high value of determination coefficient does not yet mean that the right explanatory variables are included in the model;
- Values of determination coefficient are not comparable among models with differently defined dependant variable;
- Values of determination coefficient are in general higher for time series models than for cross-section based models;
- A low value of determination coefficient does not mean that the model does not include relevant explanatory variables.

Testing the Model as a Whole

Testing for the statistical significance of a regression model as a whole:

R. A. Fisher established in 1922 that the ratio between explained and residual variance, taking into account the respective degrees of freedom, is distributed according to a specific distribution, named after him the ***F*–distribution**.

Testing the Model as a Whole

Ronald Aylmer Fisher (1890–1962)

Testing the Model as a Whole

$$F = \frac{\frac{ESS}{k-1}}{\frac{RSS}{n-k}} \sim F_{(k-1,n-k)}$$

$$F = \frac{\frac{ESS}{TSS(k-1)}}{\frac{RSS}{TSS(n-k)}} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \sim F_{(k-1,n-k)}$$

Testing the Model as a Whole

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

$$F > F_c$$

$$R^2 > R^2_c$$

$$R_c^2 = \frac{F_c(k-1)}{F_c(k-1)+(n-k)}$$

Likelihood Function and Information Criteria

Log–likelihood value of the model:

$$\ln L = -\frac{n}{2}\left[\ln(2\pi) + \ln\left(\frac{\text{RSS}}{n}\right) + 1\right]$$

Akaike information criterion (AIC):

$$AIC = -2 \ln L + 2k$$

Schwarz criterion (SC) or Bayesian information criterion (BIC):

$$SC = -2 \ln L + k \ln(n)$$

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

2.6 Interpretation of Regression Coefficients

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

Representation of Estimation Results

Representation of results of an estimated regression model

$$y = \beta_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$$

$$\begin{array}{ccccccc}
\hat{y} & = & b_1 & + & b_2 x_2 & + \quad \cdots & + b_k x_k \\
 & & (se(b_1)) & & (se(b_2)) & & (se(b_k)) \\
 & & (t_1) & & (t_2) & \cdots & (t_k) \\
 & & (p_1) & & (p_2) & \cdots & (p_k) \\
\end{array}$$

$$n = \ldots. \qquad R^2 = \ldots \qquad \bar{R}^2 = \ldots\ldots$$

$$s_e = \ldots.. \qquad (F; DW; h)$$

Computer Printout of the Results

Computer printout of econometric estimation of a linear production function (Stata)

```
. regress q l k

      Source |       SS       df       MS              Number of obs =      81
-------------+------------------------------           F(  2,    78) =   52.90
       Model |  6.9350e+12     2  3.4675e+12           Prob > F      =  0.0000
    Residual |  5.1130e+12    78  6.5551e+10           R-squared     =  0.5756
-------------+------------------------------           Adj R-squared =  0.5647
       Total |  1.2048e+13    80  1.5060e+11           Root MSE      = 2.6e+05

------------------------------------------------------------------------------
           q |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           l |   9687.383   3640.852     2.66   0.009     2439.003    16935.76
           k |    2.27941   .7553228     3.02   0.003      .775678    3.783142
       _cons |  -11875.29   34865.13    -0.34   0.734    -81286.43    57535.85
------------------------------------------------------------------------------
```

Computer Printout of the Results

Computer printout of econometric estimation of a linear production function (R)

```
> mod_lin = lm(q ~ l + k, data = production)
> summary(mod_lin)

Call:
lm(formula = q ~ l + k, data = production)

Residuals:
Value added in 1000 monetary units
    Min      1Q  Median      3Q     Max
-928125  -30862   -3095    7945 1310726

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.188e+04  3.487e+04  -0.341  0.73432
l            9.687e+03  3.641e+03   2.661  0.00946 **
k            2.279e+00  7.553e-01   3.018  0.00344 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 256000 on 78 degrees of freedom
Multiple R-squared:  0.5756,   Adjusted R-squared:  0.5647
F-statistic:  52.9 on 2 and 78 DF,  p-value: 3.038e-15
```

Computer Printout of the Results

```
> summary.aov(mod_lin)
            Df    Sum Sq   Mean Sq F value   Pr(>F)    
l            1 6.338e+12 6.338e+12  96.688 2.64e-15 ***
k            1 5.970e+11 5.970e+11   9.107  0.00344 ** 
Residuals   78 5.113e+12 6.555e+10                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> sum(anova(mod_lin)[-3,2])
[1] 6.935018e+12

> sum(anova(mod_lin)[,2])
[1] 1.204801e+13

> nobs(mod_lin)
[1] 81

> confint(mod_lin, level=0.95)
                   2.5 %       97.5 %
(Intercept) -81286.433705 57535.852836
l             2439.003091 16935.763850
k                0.775678     3.783142
```

Computer Printout of the Results

Computer printout of econometric estimation of a linearized Cobb-Douglas production function (Stata)

```
. regress lq ll lk

      Source |       SS       df       MS              Number of obs =      81
-------------+------------------------------           F(  2,    78) =  190.75
       Model |  178.261263     2  89.1306313           Prob > F      =  0.0000
    Residual |    36.44752    78  .467275898           R-squared     =  0.8302
-------------+------------------------------           Adj R-squared =  0.8259
       Total |  214.708783    80  2.68385978           Root MSE      =  .68358

------------------------------------------------------------------------------
          lq |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ll |   .9645479   .1199229     8.04   0.000     .7257997    1.203296
          lk |   .1885438   .0673358     2.80   0.006     .0544886     .322599
       _cons |   7.546026   .4617465    16.34   0.000      6.62676    8.465293
------------------------------------------------------------------------------
```

Computer Printout of the Results

Computer printout of econometric estimation of a linearized Cobb-Douglas production function (R)

```
> mod_log = lm(lq ~ ll + lk, data = production)
> summary(mod_log)

Call:
lm(formula = lq ~ ll + lk, data = production)

Residuals:
Value added in 1000 monetary units
     Min       1Q   Median       3Q      Max
-1.22501 -0.46545 -0.08825  0.42991  1.78679

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.54603    0.46175  16.342  < 2e-16 ***
ll           0.96455    0.11992   8.043 7.77e-12 ***
lk           0.18854    0.06734   2.800  0.00644 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6836 on 78 degrees of freedom
Multiple R-squared:  0.8302,    Adjusted R-squared:  0.8259
F-statistic: 190.7 on 2 and 78 DF,  p-value: < 2.2e-16
```

Computer Printout of the Results

```
> summary.aov(mod_log)
            Df Sum Sq Mean Sq F value  Pr(>F)    
ll           1 174.60  174.60  373.65 < 2e-16 ***
lk           1   3.66    3.66    7.84 0.00644 ** 
Residuals   78  36.45    0.47                    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> sum(anova(mod_log)[-3,2])
[1] 178.2613

> sum(anova(mod_log)[,2])
[1] 214.7088

> nobs(mod_log)
[1] 81

> confint(mod_log, level=0.95)
                2.5 %    97.5 %
(Intercept) 6.6267597 8.4652928
ll          0.7257997 1.2032961
lk          0.0544886 0.3225991
```

Definition of a Regression Coefficient

$$E(y|x_2,\ldots,x_k) = \beta_1 + \beta_2 x_2 \ldots + \beta_k x_k$$

$$\beta_j = \frac{\partial E(y|x_2,\ldots,x_k)}{\partial x_j} \quad ; \quad j = 2,3,\ldots,k$$

Multiple regression coefficients

Partial regression coefficients or partial slope coefficients

$$\beta_1 = E(y|x_2 = 0,\ldots,x_k = 0)$$

Definition of a Regression Coefficient

Regression coefficients represent

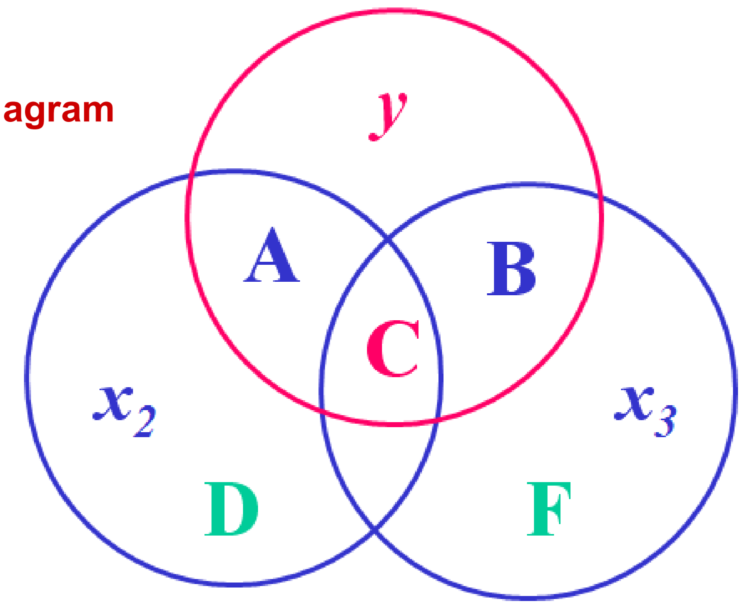
pure or direct or net effects

$$\hat{y}_i = b_1 + b_2 x_{2i} + b_3 x_{3i}$$

Venn diagram

$$\hat{x}_{2i} = c_1 + c_2 x_{3i}$$

$$e_{2i} = x_{2i} - \hat{x}_{2i}$$



$$\hat{x}_{3i} = f_1 + f_2 x_{2i}$$

$$e_{3i} = x_{3i} - \hat{x}_{3i}$$

$$\hat{y}_i = d_1 + d_2 e_{2i} + d_3 x_{3i}$$

$$\hat{y}_i = g_1 + g_2 x_{2i} + g_3 e_{3i}$$

Decomposition of the Total Effect

Effects of explanatory variable x_2

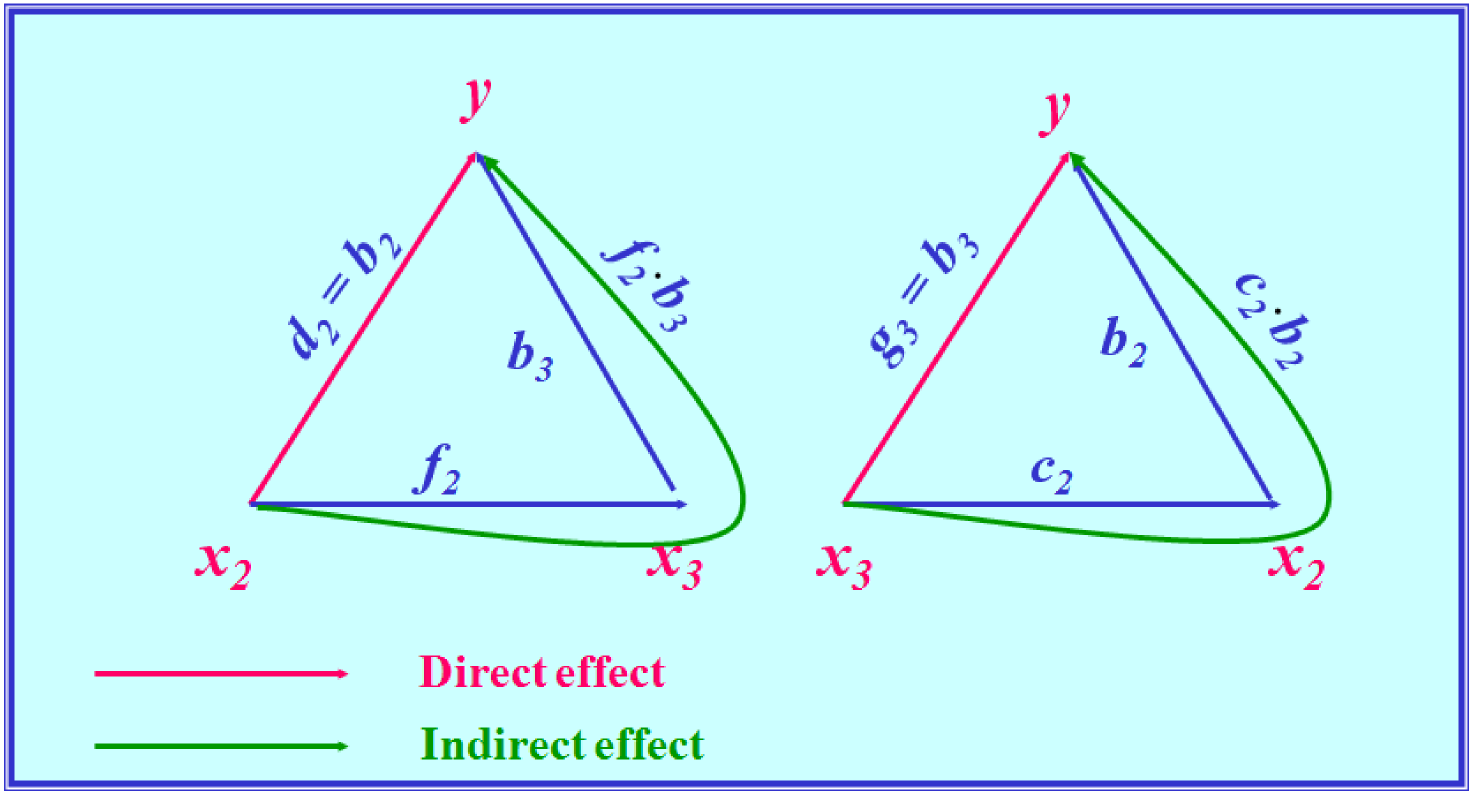
Direct effect	:	$b_2 = d_2$
Indirect effect (effect of x_2 on x_3 and through the latter on y)	:	$f_2 \cdot b_3$
Total effect	:	$g_2 = d_2 + f_2 \cdot b_3$

Effects of explanatory variable x_3

Direct effect	:	$b_3 = g_3$
Indirect effect (efect of x_3 on x_2 and through the latter on y)	:	$c_2 \cdot b_2$
Total effect	:	$d_3 = g_3 + c_2 \cdot b_2$

Decomposition of the Total Effect

Graphic illustration of direct and indirect effects of explanatory variables on the dependent variable



Frisch–Waugh (–Lovell) Theorem

Frisch–Waugh theorem
(Econometrica, 1933)

$$y_i = b_1 + b_2 x_{2i} + \underline{b_3} x_{3i} + \underline{e_i}$$

$$y_i = c_1 + c_2 x_{2i} + e_{y_i} \quad \Rightarrow \quad \tilde{y}_i = e_{y_i} = y_i - \hat{y}_i$$

$$x_{3i} = d_1 + d_2 x_{2i} + e_{x_{3i}} \quad \Rightarrow \quad \tilde{x}_{3i} = e_{x_{3i}} = x_{3i} - \hat{x}_{3i}$$

$$\tilde{y}_i = g_1 + \underline{b_3} \tilde{x}_{3i} + \underline{e_i}$$ FW(L) regression

Frisch–Waugh (–Lovell) Theorem

AN EXAMPLE OF ITS USE:

Method of individual trend

First step

We purge the “redundant” component from the time series

Second step

We estimate the regression model that is cleaned off of the effects of “redundant” component

Frisch–Waugh (–Lovell) Theorem

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \ldots + \beta_k x_{kt} + u_t$$

First step

$$y_t = c_1 + c_2 t + c_3 t^2 + \ldots + e_{y_t} \quad ; \quad e_{y_t} = y_t - \hat{y}_t$$

$$x_{2t} = c_{21} + c_{22} t + c_{23} t^2 + \ldots + e_{x_{2t}} \quad ; \quad e_{x_{2t}} = x_{2t} - \hat{x}_{2t}$$

$$\ldots$$

$$x_{kt} = c_{k1} + c_{k2} t + c_{k3} t^2 + \ldots + e_{x_{kt}} \quad ; \quad e_{x_{kt}} = x_{kt} - \hat{x}_{kt}$$

Second step

$$e_{y_t} = b_1 + b_2 e_{x_{2t}} + b_3 e_{x_{3t}} + \ldots + b_k e_{x_{kt}} + e_t$$

Frisch–Waugh (–Lovell) Theorem

AN ALTERNATIVE APPROACH:

Method of partial time-series regression

$$y_t = d_1 + b_2 x_{2t} + b_3 x_{3t} + \ldots + b_k x_{kt} + \underline{d_2 t + d_3 t^2 + \ldots} + e_t$$

University of Ljubljana

SEB SCHOOL OF ECONOMICS AND BUSINESS

2. Multiple Regression Model

Prof. Dr. Miroslav Verbič

miroslav.verbic@ef.uni-lj.si

www.miroslav-verbic.si

EFMD EQUIS ACCREDITED

AACSB ACCREDITED

ASSOCIATION AMBA ACCREDITED

Ljubljana, October 2025