Tasks with solutions for Introductory Econometrics

2. The multiple regression model

2.1 What are the standard assumptions of the multiple regression model?

Solution:

- The expected value of the disturbance term is in each period (for each statistical unit) equal to zero:$E(u_t) = 0$
- The variance of the disturbance term is constant (homoscedasticity): $Var(u_t) = E(u_t^2) = \sigma^2$
- The covariance between the disturbance terms of different periods (statistical units) is equal to zero (absence of autocorrelation): $Cov(u_t,u_s) = E(u_t \cdot u_s) = 0$ for $t \neq s$.

2.2 Derive the least squares estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ for the regression coefficients β_1 and β_2 without using the matrix calculus!

Solution:
According to the smallest square criterion, the sum of the squared disturbance variables is minimized:

$$(1) \qquad S(\beta_1,\beta_2) = \sum_{t=1}^{n} u_t^2 \rightarrow \text{Min}.$$

If the disturbance u_t is replaced by the deviation $y_t-\beta_1-\beta_2 \cdot x_t$, the least squares criterion is obtained as a function of the regression coefficients β_1 and β_2**:**

$$(2) \qquad S(\beta_1,\beta_2) = \sum_{t=1}^{n} [y_t - (\beta_1 + \beta_2 \cdot x_t)]^2 = \sum_{t=1}^{n} (y_t - \beta_1 - \beta_2 \cdot x_t)^2 \rightarrow \text{Min}.$$

Necessary condition for a minimum is that the partial derivatives disappear after β_1 as well β_2.[1]: Let us first differentiate the function according to β_1:

$$(3) \qquad \frac{\partial S(\beta_1,\beta_2)}{\partial \beta_1} = \frac{\partial \sum_{t=1}^{n} (y_t - \beta_1 - \beta_2 \cdot x_t)^2}{\partial \beta_1} = \sum_{t=1}^{n} \frac{\partial (y_t - \beta_1 - \beta_2 \cdot x_t)^2}{\partial \beta_1},$$

Here we have to apply the chain rule:

[1] The sufficient condition is not checked because it is always fulfilled for the least squares criterion (1).

- External deviation: $g = u^2$ with $u = y_t - \beta_1 - \beta_2 \cdot x_t$

  $$\rightarrow \frac{\partial g}{\partial u} = \frac{\partial u^2}{\partial u} = 2u = 2 \cdot (y_t - \beta_1 - \beta_2 \cdot x_t)$$

- Internal deviation: $u = y_t - \beta_1 - \beta_2 \cdot x_t \rightarrow \frac{\partial u}{\partial \beta_1} = \frac{\partial (y_t - \beta_1 - \beta_2 \cdot x_t)}{\partial \beta_1} = -1.$

The partial deviation (3) is obtained by multiplying the external derivative by the internal derivative:

$$(4) \qquad \frac{S(\beta_1, \beta_2)}{\partial \beta_1} = \sum_{t=1}^{n} \underbrace{2 \cdot (y_t - \beta_1 - \beta_2 \cdot x_t)}_{\text{äußere Ableitung}} \cdot \underbrace{(-1)}_{\text{innere Ableitung}} .$$

We set (4) equal to zero,

$$(5) \qquad \frac{S(\beta_1, \beta_2)}{\partial \beta_1} = \sum_{t=1}^{n} 2 \cdot (y_t - \beta_1 - \beta_2 \cdot x_t) \cdot (-1) \overset{!}{=} 0,$$

and transform the equation as follows:

$$\sum_{t=1}^{n} 2 \cdot (y_t - \beta_1 - \beta_2 \cdot x_t) \cdot (-1) = 0 \qquad | : (-2)$$

$$(6) \qquad \sum_{t=1}^{n} (y_t - \beta_1 - \beta_2 \cdot x_t) = 0$$

$$\sum_{t=1}^{n} y_t - \sum_{t=1}^{n} \beta_1 - \sum_{t=1}^{n} \beta_2 \cdot x_t = 0$$

$$\sum_{t=1}^{n} y_t - n \cdot \beta_1 - \beta_2 \cdot \sum_{t=1}^{n} x_t = 0$$

$$-n \cdot \beta_1 - \beta_2 \cdot \sum_{t=1}^{n} x_t = -\sum_{t=1}^{n} y_t .$$

From this we get the first normal equation,

$$(7) \qquad n \cdot \beta_1 + \beta_2 \cdot \sum_{t=1}^{n} x_t = \sum_{t=1}^{n} y_t ,$$

which we solve after β_1:

$$n \cdot \beta_1 = \sum_{t=1}^{n} y_t - \beta_2 \cdot \sum_{t=1}^{n} x_t \qquad | : n$$

$$(8) \qquad \beta_1 = \frac{\sum_{t=1}^{n} y_t}{n} - \beta_2 \cdot \frac{\sum_{t=1}^{n} x_t}{n}$$

If we put the least squares estimator $\hat{\beta}_2$ for β_2 in (8), we can determinate the least squares estimator $\hat{\beta}_1$ for the absolute term β_1.

(9) $$\boxed{\hat{\beta}_1 = \frac{\sum_{t=1}^{n} y_t}{n} - \hat{\beta}_2 \cdot \frac{\sum_{t=1}^{n} x_t}{n}}$$

Accordingly, the least squares estimator is determined for β_2. For the partial derivation

(10) $$\frac{\partial S(\beta_1,\beta_2)}{\partial \beta_2} = \frac{\partial \sum_{t=1}^{n} (y_t - \beta_1 - \beta_2 \cdot x_t)^2}{\partial \beta_2} = \sum_{t=1}^{n} \frac{\partial (y_t - \beta_1 - \beta_2 \cdot x_t)^2}{\partial \beta_2},$$

we apply the chain rule again:

- External deviation: $g = u^2$ with $u = y_t - \beta_1 - \beta_2 \cdot x_t$

  $\rightarrow \frac{\partial g(u)}{\partial u} = \frac{\partial u^2}{\partial u} = 2u = 2 \cdot (y_t - \beta_1 - \beta_2 \cdot x_t)$

- Internal deviation: $u = y_t - \beta_1 - \beta_2 \cdot x_t \rightarrow \frac{\partial u(\beta_2)}{\partial \beta_2} = \frac{\partial (y_t - \beta_1 - \beta_2 \cdot x_t)}{\partial \beta_2} = -x_t$.

We set the partial derivation equal to zero

(11) $$\frac{\partial S(\beta_1,\beta_2)}{\partial \beta_2} = \sum_{t=1}^{n} 2 \cdot (y_t - \beta_1 - \beta_2 \cdot x_t) \cdot (-x_t) \overset{!}{=} 0,$$

and transform it as follows:

$$\sum_{t=1}^{n} 2 \cdot (y_t - \beta_1 - \beta_2 \cdot x_t) \cdot (-x_t) = 0 \qquad | : 2$$

(12) $$\sum_{t=1}^{n} (y_t - \beta_1 - \beta_2 \cdot x_t) \cdot (-x_t) = 0$$

$$\sum_{t=1}^{n} \left(-x_t \cdot y_t + \beta_1 \cdot x_t + \beta_2 \cdot x_t^2\right) = 0$$

$$-\sum_{t=1}^{n} x_t \cdot y_t + \sum_{t=1}^{n} \beta_1 \cdot x_t + \sum_{t=1}^{n} \beta_2 \cdot x_t^2 = 0$$

$$\sum_{t=1}^{n} \beta_1 \cdot x_t + \sum_{t=1}^{n} \beta_2 \cdot x_t^2 = \sum_{t=1}^{n} x_t \cdot y_t .$$

From this we obtain the second normal equation**:**

(13) $$\beta_1 \cdot \sum_{t=1}^{n} x_t + \beta_2 \cdot \sum_{t=1}^{n} x_t^2 = \sum_{t=1}^{n} x_t \cdot y_t .$$

We set (8) into (13),

$$\left(\frac{1}{n} \cdot \sum_{t=1}^{n} y_t - \beta_2 \cdot \frac{1}{n} \sum_{t=1}^{n} x_t\right) \cdot \sum_{t=1}^{n} x_t + \beta_2 \cdot \sum_{t=1}^{n} x_t^2 = \sum_{t=1}^{n} x_t \cdot y_t ,$$

and solve the equation to β_2:

$$\frac{1}{n} \cdot \sum_{t=1}^{n} y_t \cdot \sum_{t=1}^{n} x_t - \beta_2 \cdot \frac{1}{n} \left(\sum_{t=1}^{n} x_t\right)^2 + \beta_2 \cdot \sum_{t=1}^{n} x_t^2 = \sum_{t=1}^{n} x_t \cdot y_t$$

$$\beta_2 \cdot \left[\sum_{t=1}^{n} x_t^2 - \frac{1}{n} \left(\sum_{t=1}^{n} x_t \right)^2 \right] = \sum_{t=1}^{n} x_t \cdot y_t - \frac{1}{n} \cdot \sum_{t=1}^{n} y_t \cdot \sum_{t=1}^{n} x_t \qquad \Big| \cdot n$$

$$\beta_2 \cdot \left[n \cdot \sum_{t=1}^{n} x_t^2 - \left(\sum_{t=1}^{n} x_t \right)^2 \right] = n \cdot \sum_{t=1}^{n} x_t \cdot y_t - \sum_{t=1}^{n} y_t \cdot \sum_{t=1}^{n} x_t$$

(14)
$$\boxed{\hat{\beta}_2 = \frac{n \cdot \sum_{t=1}^{n} x_t \cdot y_t - \sum_{t=1}^{n} y_t \cdot \sum_{t=1}^{n} x_t}{n \cdot \sum_{t=1}^{n} x_t^2 - \left(\sum_{t=1}^{n} x_t \right)^2}}.$$

The result is the least squares estimator $\hat{\beta}_2$ for β_2.

2.3 For the demand of natural gas (GASD) and natural gas price (GASP) data are available for a 7-year period (energy model Ia):

t	1	2	3	4	5	6	7
GASD	1,0	0,8	0,7	0,9	1,2	1,1	1,4
GASP	12,9	13,8	13,7	13,2	12,2	12,9	12,5

Estimate the demand function for natural gas with the ordinary least squares method by the help of the sum formulas! Interpret the estimated demand function!

Solution:

Dependent variable (y): demand of natural gas (GASD)

Independent variable (x): price of natural gas (GASP)

Working table:

t	x_t	y_t	x_t^2	$x_t \cdot y_t$
1	12,9	1,0	166,41	12,90
2	13,8	0,8	190,44	11,04
3	13,7	0,7	187,69	9,59
4	13,2	0,9	174,24	11,88
5	12,2	1,2	148,84	14,64
6	12,9	1,1	166,41	14,19
7	12,5	1,4	156,25	17,50
Σ	91,2	7,1	1190,28	91,74

OLS-estimater of the regression coefficients β_1 and β_2:

$$\hat{\beta}_2 = \frac{n \cdot \sum x_t y_t - \sum x_t \sum y_t}{n \cdot \sum x_t^2 - (\sum x_t)^2} = \frac{7 \cdot 91{,}74 - 7{,}1 \cdot 91{,}2}{7 \cdot 1190{,}28 - 91{,}2^2} = \frac{-5{,}34}{14{,}52} = -0{,}368$$

$$\hat{\beta}_1 = \frac{\sum y_t}{n} - \hat{\beta}_2 \cdot \frac{\sum x_t}{n} = \frac{7{,}1}{7} - (-0{,}368) \cdot \frac{91{,}2}{7} = 1{,}014 + \underbrace{0{,}368 \cdot 13{,}029}_{=4{,}795} = 5{,}809$$

OLS-estimated demand function for natural gas as a function of gas price:

$$\underbrace{\hat{GASD}_t}_{\hat{y}_t} = 5{,}809 - 0{,}368 \cdot \underbrace{GASP_t}_{x_t}$$

Interpretation:

- $\hat{\beta}_1 = 5{,}809$: The absolute term can interpreted as a technical constant that defines the position of the regression line in the two-dimensional variable space. Formally, the absolute term indicates the demand for natural gas at a gas price of 0.
- $\hat{\beta}_2 = -0{,}368$: If the natural gas price increases by one unit, then the consumption of natuarl gas will decrease by 0.368 units on average.

2.4

For the demand of natural gas (GASD) and disposable income (DINC) data are available for a 7-year period (energy model Ib):

t	1	2	3	4	5	6	7
GASD	1,0	0,8	0,7	0,9	1,2	1,1	1,4
DINC	12	12	9	11	13	13	15

Estimate the demand for natural gas depending on the disposable income with the OLS method using matrix calculus! Interpret the estimated demand function!

Solution:

Dependent variable (y): Erdgasverbrauch (GASV)

Independent variable (x): disposable income (DIINC)

- vector of the dependent variable **y** and of the observation matrix **X**:

$$\underset{7\times1}{\mathbf{y}} = \begin{pmatrix} 1{,}0 \\ 0{,}8 \\ 0{,}7 \\ 0{,}9 \\ 1{,}2 \\ 1{,}1 \\ 1{,}4 \end{pmatrix} \qquad \underset{7\times2}{\mathbf{X}} = \begin{pmatrix} 1 & 12 \\ 1 & 12 \\ 1 & 9 \\ 1 & 11 \\ 1 & 13 \\ 1 & 13 \\ 1 & 15 \end{pmatrix}$$

- product matrix **X**'**X**:

$$\mathbf{X'X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 12 & 12 & 9 & 11 & 13 & 13 & 15 \end{pmatrix} \begin{pmatrix} 1 & 12 \\ 1 & 12 \\ 1 & 9 \\ 1 & 11 \\ 1 & 13 \\ 1 & 13 \\ 1 & 15 \end{pmatrix} = \begin{pmatrix} n = 7 & \sum x_t = 85 \\ \sum x_t = 85 & \sum x_t^2 = 1053 \end{pmatrix}$$

- inverse of the product matrix $(\mathbf{X'X})^{-1}$:

Adjoint matrix:

Rule for an symmetric 2x2-matrix: $\mathbf{A} = \begin{pmatrix} a & b \\ b & d \end{pmatrix} \rightarrow \mathbf{A}^{adj} = \begin{pmatrix} d & -b \\ -b & a \end{pmatrix}$

$$(\mathbf{X'X})^{adj} = \begin{pmatrix} 7 & 85 \\ 85 & 1053 \end{pmatrix}^{adj} = \begin{pmatrix} 1053 & -85 \\ -85 & 7 \end{pmatrix}$$

determinant:

$$|\mathbf{X'X}| = \begin{vmatrix} 7 & 85 \\ 85 & 1053 \end{vmatrix} = 7 \cdot 1053 - 85^2 = 146$$

inverse:

$$(\mathbf{X'X})^{-1} = \frac{1}{|\mathbf{X'X}|} \cdot (\mathbf{X'X})^{\text{adj}} = \frac{1}{146} \cdot \begin{pmatrix} 1053 & -85 \\ -85 & 7 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1053}{146} & \frac{-85}{146} \\ \frac{-85}{146} & \frac{7}{146} \end{pmatrix} = \begin{pmatrix} 7{,}212329 & -0{,}582192 \\ -0{,}582192 & 0{,}047945 \end{pmatrix}$$

- product of the of the transpose vector and the vector $\mathbf{X'y}$:

$$\mathbf{X'y} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 12 & 12 & 9 & 11 & 13 & 13 & 15 \end{pmatrix} \begin{pmatrix} 1{,}0 \\ 0{,}8 \\ 0{,}7 \\ 0{,}9 \\ 1{,}2 \\ 1{,}1 \\ 1{,}4 \end{pmatrix} = \begin{pmatrix} \sum y_t = 7{,}1 \\ \sum x_t \cdot y_t = 88{,}7 \end{pmatrix}$$

- OLS-estimator of $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'y} = \begin{pmatrix} 7{,}212329 & -0{,}582192 \\ -0{,}582192 & 0{,}047945 \end{pmatrix} \begin{pmatrix} 7{,}1 \\ 88{,}7 \end{pmatrix} = \begin{pmatrix} \hat{\beta}_1 = -0{,}433 \\ \hat{\beta}_2 = 0{,}119 \end{pmatrix}$$

Via OLS estimated demande function of natrual gas as a function of the disposable income:

$$\underbrace{\hat{GASD}_t}_{\hat{y}_t} = -0{,}433 + 0{,}119 \cdot \underbrace{DIINC_t}_{x_t}$$

Interpretation:

- $\hat{\beta}_1 = -0{,}433$: The absolute term can interpreted as a technical constant that defines the position of the regression line in the two-dimensional variable space. Formally, the absolute term indicates the demand for natural gas at a gas price of 0.
- $\hat{\beta}_2 = 0{,}119$: If the natural gas price increases by one unit, then the consumption of natuarl gas will decrease by 0.119 units on average.

2.5 The demand for natural gas can explained by the natural gas price (GASP) and disposable income (DIINC) (energy model II). Determine the OLS estimator of the linear demand function for natural gas (for data, see exercise 2.3 and 2.4)!

Note: The inverse of the product matrix X'X is given

$$\mathbf{X'X} = \begin{pmatrix} 7 & 91,2 & 85 \\ 91,2 & 1190,28 & 1102,7 \\ 85 & 1102,7 & 1053 \end{pmatrix}$$

Solution:
Dependent variable (y): natural gas demand (GASD)

Independent variables:
x_1: natural gas price (GASP)
x_2: disposable income (DIINC)

- Vector of dependent variable **y** and observation matrix **X:**

$$\underset{7\times1}{\mathbf{y}} = \begin{pmatrix} 1,0 \\ 0,8 \\ 0,7 \\ 0,9 \\ 1,2 \\ 1,1 \\ 1,4 \end{pmatrix}, \quad \underset{7\times3}{\mathbf{X}} = \begin{pmatrix} 1 & 12,9 & 12 \\ 1 & 13,8 & 12 \\ 1 & 13,7 & 9 \\ 1 & 13,2 & 11 \\ 1 & 12,2 & 13 \\ 1 & 12,9 & 13 \\ 1 & 12,5 & 15 \end{pmatrix}$$

- Vector **X'y**

$$\mathbf{X'y} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 12,9 & 13,8 & 13,7 & 13,2 & 12,2 & 12,9 & 12,5 \\ 12 & 12 & 9 & 11 & 13 & 13 & 15 \end{pmatrix} \begin{pmatrix} 1,0 \\ 0,8 \\ 0,7 \\ 0,9 \\ 1,2 \\ 1,1 \\ 1,4 \end{pmatrix} = \begin{pmatrix} \sum y_t = 7,1 \\ \sum x_{2t} y_t = 91,74 \\ \sum x_{3t} y_t = 88,7 \end{pmatrix}$$

The inverse of the product matrix X'X is above given:

$$(\mathbf{X'X})^{-1} = \begin{pmatrix} 255,7066 & -15,7459 & -4,1520 \\ -15,7459 & 0,997745 & 0,2262 \\ -4,1520 & 0,2262 & 0,099228 \end{pmatrix}$$

- OLS-Schätzer für **β**:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'y} = (\mathbf{X'X})^{-1} = \begin{pmatrix} 255,7066 & -15,7459 & -4,1520 \\ -15,7459 & 0,997745 & 0,2262 \\ -4,1520 & 0,2262 & 0,099228 \end{pmatrix} \begin{pmatrix} 7,1 \\ 91,74 \\ 88,7 \end{pmatrix} = \begin{pmatrix} 2,706 \\ -0,199 \\ 0,074 \end{pmatrix}$$

2.6 What is a BLUE estimator?

Solution:

- A BLUE estimator is an unbiased and efficient linear estimator.
- An estimator is unbiasd if his expected value matches the unknown parameter of the population. An unbiased estimator is undistorted.
- An unbiased estimator is efficient if it has the least variance of all unbiased estimators.

A BLUE estimator has the least variance of any linear, unbiased estimators.

2.7 2.7 Explain the consistency property verbally! What value would you attach to this estimation property in applied econometrics depending on the type of data?

Solution:

An estimation approach is consistent when the estimated parameter coincides with the true parameter for a sample size that grows across all borders. Consistency is ensured when two properties are met:

- Asyptotic unbiasedness: In a sample size that is going to infinity, the expected value of the estimator coincides with the true parameter.

- The variance of the estimator goes to zero for a sample size that goes to infinity.

Consistency is a minimum requirement. If econometric models do not meet expectations and efficiency, the estimation function should at least be consistent.

However, the consistency characteristic is only noticeable in large samples. While this requirement is often met for cross-sectional data, e.g. from company and consumer surveys, it often cannot be met for time series data. As a rule, annual data do not contain sufficiently large samples. In the case of time series data, consistency is therefore more relevant for monthly and quarterly data.

2.8 Calculate the determination coefficient for the energy model Ib (see task 2.4) using the sum formula and interpret it!

Solution:

sum formula of the coefficient of determination: $R^2 = \frac{\sum \hat{y}_t^2 - n\bar{y}^2}{\sum y_t^2 - n\bar{y}^2}$

OLS-estimated demand function for natural gas as a function of disposable income:

$$\underbrace{\hat{GASD}_t}_{\hat{y}_t} = -0{,}433 + 0{,}119 \cdot \underbrace{DIINC_t}_{x_t}$$

regression values ($\hat{y}_t$):

t=1: $\hat{y}_1 = -0{,}433 + 0{,}119 \cdot (x_1 = 12) = 0{,}995$

t=2: $\hat{y}_2 = -0{,}433 + 0{,}119 \cdot (x_2 = 12) = 0{,}995$

t=3: $\hat{y}_3 = -0{,}433 + 0{,}119 \cdot (x_3 = 9) = 0{,}638$

t=4: $\hat{y}_4 = -0{,}433 + 0{,}119 \cdot (x_4 = 11) = 0{,}876$

t=5: $\hat{y}_5 = -0{,}433 + 0{,}119 \cdot (x_5 = 13) = 1{,}114$

t=6: $\hat{y}_6 = -0{,}433 + 0{,}119 \cdot (x_6 = 13) = 1{,}114$

t=7: $\hat{y}_7 = -0{,}433 + 0{,}119 \cdot (x_1 = 15) = 1{,}352$

working table:

t	y_t	y_t^2	$\hat{y}_t$	$\hat{y}_t^2$
1	1,0	1,00	0,995	0,9900
2	0,8	0,64	0,995	0,9900
3	0,7	0,49	0,638	0,4070
4	0,9	0,81	0,876	0,7674
5	1,2	1,44	1,114	1,2410
6	1,1	1,21	1,114	1,2410
7	1,4	1,96	1,352	1,8279
Σ	7,1	7,55	≈7,1	7,4643

Arithmetic mean of the dependent variable y: $\bar{y} = \frac{1}{n}\sum y_t = \frac{1}{7} \cdot 7{,}1 = 1{,}014$

coefficient of determination:

$$R^2 = \frac{\sum \hat{y}_t^2 - n \cdot \bar{y}^2}{\sum y_t^2 - n \cdot \bar{y}^2} = \frac{7{,}4643 - 7 \cdot 1{,}014^2}{7{,}55 - 7 \cdot 1{,}014^2} = \frac{0{,}2669}{0{,}3526} = 0{,}757$$

Interpretation:

- The disposable income explains 75.7% of the gas consumption variance.
- The squared correlation between the y-values and the regression values (and thus also the x-values) is 0.757.

2.9 Calculate the coefficient of determination for the energy model II (see exercise 2.5) by the of the OLS estimator and interpret it!

Solution:

Calculation formula: $R^2 = \frac{\hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{y} - n\bar{y}^2}{\mathbf{y}' \mathbf{y} - n\bar{y}^2}$

$$\hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{y} = \begin{pmatrix} 2{,}706 & -0{,}199 & 0{,}074 \end{pmatrix} \begin{pmatrix} 7{,}1 \\ 91{,}74 \\ 88{,}7 \end{pmatrix} = 7{,}5201$$

$$\mathbf{y}'\mathbf{y} = \begin{pmatrix}1{,}0 & 0{,}8 & 0{,}7 & 0{,}9 & 1{,}2 & 1{,}1 & 1{,}4\end{pmatrix}\begin{pmatrix}1{,}0\\0{,}8\\0{,}7\\0{,}9\\1{,}2\\1{,}1\\1{,}4\end{pmatrix} = 7{,}55$$

$$R^2 = \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{y}^2}{\mathbf{y}'\mathbf{y} - n\bar{y}^2} = \frac{7{,}5201 - 7 \cdot 1{,}014^2}{7{,}55 - 7 \cdot 1{,}014^2} = \frac{7{,}5201 - 7{,}1974}{7{,}55 - 7{,}1974} = \frac{0{,}3227}{0{,}3526} = 0{,}915$$

Interpretation:

- The natural gas price and disposable income explaining 91.5% of the gas consumption variance.
- The squared correlation between the y-values and the regression values is 0.915.

2.10 Explain the two different ways of interpreting the coefficient of determination!

Solution:

- The coefficient of determination indicates the proportion of the variance of the dependent variable which is explained by the regression approach.
- The determination coefficient is the squared multiple correlation coefficient $R^2 = r^2_{y\hat{y}}$, that quantifies the relationship between the dependent variable y and the regression values $\hat{y}_t$.

2.11 What role could the adjusted coefficient of determination play in model selection?

Solution:

The usual coefficient of determination can only increase by the addition of further exogenous variables or remain the same in borderline cases. The corrected coefficient of determination therefore performs a correction by replacing the number of observations with the number of degrees of freedom. The more exogenous variables the regression model contains, the greater the correction factor. The possible applications are derived from this:

- The adjusted coefficient of determination should be used to check whether it makes sense to include other exogenous variables. Then the model with the larger corrected coefficient of determination should be selected.
- The adjusted coefficient of determination can be used when completely different regression models are to be compared. Here, too, the regression model with the largest corrected explanatory contribution (≙ larger corrected determination coefficient) should be used.