Aufgaben mit Lösungen zur Ökonometrie I

2. Das multiple Regressionsmodell

2.1 Welche Standardannahmen liegen dem multiplen Regressionsmodell zugrunde?

Lösung:

- Die Störgröße hat in jeder Periode (für jede statistische Einheit) haben einen Erwartungswert von null: $E(u_t) = 0$
- Die Varianz der Störgröße ist konstant (Homoskedastizität): $Var(u_t) = E(u_t^2) = \sigma^2$
- Die Kovarianz zwischen den Störgrößen verschiedener Perioden (statistischer Einheiten) ist null (fehlende Autokorrelation): $Cov(u_t,u_s) = E(u_t \cdot u_s) = 0$ für $t \neq s$.

2.2 Leiten Sie die Kleinst-Quadrate-Schätzer $\hat{\beta}_1$ und $\hat{\beta}_2$ für die Regressionskoeffizienten β_1 und β_2 ohne Verwendung des Matrixkalküls her!

Lösung:
Nach dem Kleinst-Quadrate-Kriterium wird die Summe der quadrierten Störgrößen minimiert:

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

Ersetzt man hierin die Störgröße u_t durch die Abweichung $y_t-\beta_1-\beta_2 \cdot x_t$, so erhält man das Kleinst-Quadrate-Kriterium in Abhängigkeit von den Regressionskoeffizienten β_1 und β_2:

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

Notwendige Bedingung für ein Minimum ist, dass die partiellen Ableitungen nach β_1 sowie β_2 verschwinden.[1] Differenzieren wir zuerst die Funktion nach β_1:

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

wobei wir die Kettenregel anwenden müssen:

- Äußere Ableitung: $g = u^2$ mit $u = y_t - \beta_1 - \beta_2 \cdot x_t$

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

---

[1] Die hinreichende Bedingung wird nicht überprüft, da sie für das Kleinst-Quadrate-Kriterium (1) stets erfüllt ist.

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

Die partielle Ableitung (3) erhält man durch Multiplikation der äußeren mit der inneren Ableitung:

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

Wir setzen (4) gleich null,

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

und formen die Gleichung um:

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

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

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

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

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

Hieraus erhalten wir die **erste Normalgleichung**

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

die wir nach β_1 auflösen:

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

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

Setzt man den noch zu bestimmenden Kleinst-Quadrate-Schätzer $\hat{\beta}_2$ in (8) für β_2 ein, lässt sich der Kleinst-QuadrateSchätzer $\hat{\beta}_1$ für das absolute Glied β_1 berechnen:

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

Entsprechend wird der Kleinst-Quadrate-Schätzer für β_2 bestimmt. Bei der partiellen Ableitung

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

wird erneut die Kettenregel angewendet:

- Äußere Ableitung: $g = u^2$ mit $u = y_t - \beta_1 - \beta_2 \cdot x_t$

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

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

Wir setzen die partielle Ableitung gleich null,

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

und formen sie um:

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

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

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

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

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

Hieraus erhalten wir die **zweite Normalgleichung:**

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

Wir setzen (8) in (13) ein,

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

und lösen die Gleichung nach β_2 auf:

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

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

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

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

Als Ergebnis erhält man den Kleinst-Quadrate-Schätzer $\hat{\beta}_2$ für β_2.

2.3 Für die Nachfrage nach Erdgas (GASV) und die Preise für Erdgas (GASP) (≙ Energiemodell Ia) liegen die Daten für einen 7-Jahres-Zeitraum vor:

t	1	2	3	4	5	6	7
GASV	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

Schätzen Sie die Nachfragefunktion nach Erdgas mit der gewöhnlichen Methode der kleinsten Quadrate unter Verwendung der Summenformeln! Interpretieren Sie die geschätzte Nachfragefunktion!

Lösung:

Abhängige Variable (y): Nachfrage nach Erdgas (GASV)
Unabhängige Variable (x): Erdgaspreis (GASP)

Arbeitstabelle:

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-Schätzer der Regressionskoeffizienten β_1 und 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-geschätzte Nachfragefunktion für Erdgas in Abhängigkeit vom Gaspreis:

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

Interpretation:

- $\hat{\beta}_1 = 5{,}809$: Das absolute Glied ist hier allein als technische Konstante zu interpretieren, die die Lage der Regressionsgerade im zweidimensionalen Variablenraum festlegt. Formal gibt sie die Nachfrage nach Erdgas bei einem Gaspreis von 0 an. Da jedoch ein Gaspreis von 0 nicht im Stützbereich liegt, ist diese Interpretation jedoch nicht zulässig.
- $\hat{\beta}_2 = -0{,}368$: Erhöht sich der Erdgaspreis um eine Einheit, dann sinkt der Erdgasverbrauch durchschnittlich um 0,368 Einheiten.

2.4 Für die Nachfrage nach Erdgas (GASV) und das verfügbare Einkommen (VEINK) (≙ Energiemodell Ib) liegen die Daten für einen 7-Jahres-Zeitraum vor:

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

Schätzen Sie die Nachfrage nach Erdgas in Abhängigkeit vom verfügbaren Einkommen mit der OLS-Methode unter Verwendung der Matrizenrechnung! Interpretieren Sie die geschätzte Nachfragefunktion!

Lösung:

Abhängige Variable (y): Erdgasverbrauch (GASV)

Unabhängige Variable (x): Verfügbares Einkommen (VEINK)

- Vektor der abhängigen Variablen **y** und Beobachtungsmatrix **X**:

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

- Produktmatrix **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 Produktmatrix $(\mathbf{X'X})^{-1}$:

Adjungierte Matrix:

Regel für symmetrische 2x2-Matrix : $\mathbf{A} = \begin{pmatrix} a & b \\ b & d \end{pmatrix} \rightarrow \mathbf{A}^{adj} = \begin{pmatrix} d & -b \\ -b & a \end{pmatrix}$

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

Determinante:

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

Inverse:

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

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

- Vektor $\mathbf{X'y}$:

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

- OLS-Schätzer für $\boldsymbol{\beta}$:

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

OLS-geschätzte Nachfragefunktion für Erdgas in Abhängigkeit vom verfügbaren Einkommen:

$$\underbrace{\widehat{\text{GASV}}_t}_{\hat{y}_t} = -0{,}433 + 0{,}119 \cdot \underbrace{\text{VEINK}_t}_{x_t}$$

Interpretation:

- $\hat{\beta}_1 = -0{,}433$: Das absolute Glied ist hier allein als technische Konstante zu interpretieren, die die Lage der Regressionsgerade im zweidimensionalen Variablenraum festlegt. Da ein Einkommen von 0 nicht im Stützbereich liegt, lässt sich die Konstante ökonomisch nicht interpretieren.
- $\hat{\beta}_2 = 0{,}119$: Erhöht sich das verfügbare Einkommen um eine Einheit, dann steigt der Erdgasverbrauch durchschnittlich um 0,119 Einheiten.

2.5 Die Nachfrage nach Erdgas soll durch den Erdgaspreis (GASP) und das verfügbare Einkommen (VEINK) mit einer lineare Nachfragefunktion erklärt werden (≙Energiemodell II). Bestimmen Sie die OLS-Schätzer der Nachfragefunktion nach Erdgas (Daten s. Aufg. 2.3 und 2.4)!

<u>Hinweis:</u> Bilden Sie hierzu die Inverse der Produktmatrix **X'X** mit dem Programm R:

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

Lösung:

Abhängige Variable (y): Erdgasverbrauch (GASV)

Unabhängige Variablen:
x_1: Ergaspreis (GASP)
x_2: Verfügbares Einkommen (VEINK)

- Vektor der abhängigen Variablen **y** und Beobachtungsmatrix **X**:

$$\underset{7\times 1}{\mathbf{y}} = \begin{pmatrix} 1{,}0 \\ 0{,}8 \\ 0{,}7 \\ 0{,}9 \\ 1{,}2 \\ 1{,}1 \\ 1{,}4 \end{pmatrix}, \quad \underset{7\times 3}{\mathbf{X}} = \begin{pmatrix} 1 & 12{,}9 & 12 \\ 1 & 13{,}8 & 12 \\ 1 & 13{,}7 & 9 \\ 1 & 13{,}2 & 11 \\ 1 & 12{,}2 & 13 \\ 1 & 12{,}9 & 13 \\ 1 & 12{,}5 & 15 \end{pmatrix}$$

- Vektor **X'y**

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

Die mit dem Programm R berechnete Inverse der Produktmatrix **X'X** lautet:

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

- 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 Was versteht man unter einem BLUE-Schätzer?

Lösung:

- Ein BLUE-Schätzer ist ein erwartungstreuer (unbiased) und effizienter (best) linearer (linear) Schätzer (estimator).
- Ein Schätzer ist erwartungstreu, wenn sein Erwartungswert mit dem unbekannten Parameter der Grundgesamtheit übereinstimmt. Ein erwartungstreuer Schätzer ist unverzerrt.
- Ein erwartungstreuer Schätzer ist effizient, wenn er die geringste Varianz von allen erwartungstreuen Schätzern aufweist.
- Ein BLUE-Schätzer weist die geringste Varianz unter allen linearen, erwartungstreuen Schätzfunktionen auf.

2.7 Erläutern Sie verbal die Konsistenzeigenschaft! Welchen Stellenwert würden Sie dieser Schätzeigenschaft in der angewandten Ökonometrie in Abhängigkeit von der Datenart zubilligen?

Lösung:

Ein Schätzansatz ist konsistent, wenn der geschätzte Parameter bei einem über alle Grenzen wachsenden Stichprobenumfang mit dem wahren Parameter zusammenfällt. Die Konsistenz ist gesichert, wenn zwei Eigenschaften erfüllt sind:

- Asyptotische Erwartungstreue: Bei einem gegen unendlich strebenen Stichprobenumfang fällt der Erwartungswert des Schätzers mit dem wahren Parameter zusammen.
- Die Varianz des Schätzers geht bei einem gegen unendlich strebenden Stichprobenumfang gegen null.

Die Konsistenz ist eine Minimalforderung. Wenn in ökonometrischen Modellen nicht die Erwartungstreue und Effizienz erfüllt ist, so sollte die Schätzfunktion doch zumindest konsistent sein.

Allerdings macht sich die Konsistenzeigenschaft nur bei großen Stichproben bemerkbar. Während diese Voraussetzung bei Querschnittsdaten z.B. aus Unternehmens- und Verbraucherbefragungen häufig erfüllt ist, lässt sie sich bei Zeitreihendaten oft nicht realisieren. Bei Jahresdaten liegen in der Regel nicht ausreichend große Stichproben vor. Die Konsistenz ist im Falle von Zeitreihendaten deshalb eher bei Monats- und Quartalsdaten relevant.

2.8 Berechnen Sie den Determinationskoeffizienten für das Energiemodell Ib (s. Aufgabe 2.4) unter Verwendung der Summenformel und interpretieren Sie ihn!

Lösung:

Berechnungsformel: $R^2 = \frac{\sum \hat{y}_t^2 - n\bar{y}^2}{\sum y_t^2 - n\bar{y}^2}$

OLS-geschätzte Nachfragefunktion für Erdgas in Abhängigkeit vom verfügbaren Einkommen:

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

Regressionswerte ($\hat{y}_t$):

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

Arbeitstabelle:

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

Arithmetisches Mittel der abhängigen Variablen y: $\bar{y} = \frac{1}{n}\sum y_t = \frac{1}{7} \cdot 7{,}1 = 1{,}014$

Determinationskoeffizient:

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

Interpretation:

- Das verfügbare Einkommen erklärt 75,7% der Varianz des Gasverbrauchs.
- Die quadrierte Korrelation zwischen den y-Werten und den Regressionswerten $\hat{y}$ und damit auch den x-Werten beträgt 0,757.

2.9 Berechnen Sie den Determinationskoeffizienten für das Energiemodell II (s. Aufgabe 2.5) unter Verwendung des OLS-Schätzers $\hat{\boldsymbol{\beta}}$ und interpretieren Sie ihn!

Lösung:

Berechnungsformel: $R^2 = \dfrac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{y}^2}{\mathbf{y}'\mathbf{y} - n\bar{y}^2}$

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

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

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

Interpretation:

- Durch den Erdgaspreis und das verfügbare Einkommen wird 91,5% der Varianz des Gasverbrauchs erklärt.
- Die quadrierte Korrelation zwischen den y-Werten und den Regressionswerten $\hat{y}$ beträgt 0,915.

2.10 Erläutern Sie die beiden unterschiedlichen Interpretationsmöglichkeiten des Determinationskoeffizienten!

Lösung:

- Der Determinationskoeffizient gibt den Anteil der Varianz der abhängigen Variablen an, der durch den Regressionsansatz erklärt wird.
- Der Determinationskoeffizient ist der quadrierte multiple Korrelationskoeffizient $R^2 = r^2_{y\hat{y}}$, der den Zusammenhang zwischen der abhängigen Variablen y und den Regressionswerten $\hat{y}_t$ quantifiziert.

2.11 Welche Rolle könnte der korrigierte Determinationskoeffizient bei der Modellauswahl einnehmen?

Lösung:

Der gewöhnliche Determinationskoeffizient kann durch die Aufnahme weiterer exogener Variabler nur steigen oder im Grenzfall gleich bleiben. Der korrigierte Determinationskoeffizient führt deswegen eine Korrektur durch, indem die Anzahl der Beobachtungen durch die Zahl der Freiheitsgrade ersetzt wird. Der Korrekturfaktor ist umso größer, je mehr exogene Variablen das Regressionsmodell enthält. Hieraus leiten sich die Einsatzmöglichkeiten ab:

- Der korrigierte Determinationskoeffizient ist dann einzusetzen, wenn geprüft werden soll, ob die Einbeziehung weiterer exogener Variabler sinnvoll ist. Danach wäre das Modell mit dem größeren korrigierten Bestimmtheitsmaß auszuwählen.
- Der korrigerte Determinationskoeffizient kann verwendet werden, wenn vollkommen unterschiedliche Regressionsmodelle miteinander verglichen werden sollen. Auch hier wäre das Regressionsmodell mit dem größten korrigierten Erklärungsbeitrag ($\hat{=}$ größer korrigierter Determinationskoeffizient) heranzuziehen.