Klausur zu Ökonometrie (Master)

Technische Universität Dortmund | Fakultät Wirtschaftswissenschaften | 1. August 2025

Bitte tragen sie Ihre Daten sorgfältig und leserlich ein:

Matrikelnummer ☐☐☐☐☐☐ Nachname ____________

Studiengang ____________ Vorname ____________

Bearbeitungshinweise:

Diese Klausur besteht aus fünf Aufgaben, von welchen **vier Aufgaben ihrer Wahl** zu bearbeiten sind.

Bearbeiten sie alle Aufgaben, so werden nur die ersten vier bewertet.

Alle Antworten sind zu begründen.

Bitte verwenden sie einen Kugelschreiber oder nicht zu starken Filzstift.

Für jede der fünf Aufgaben sind maximal je 18 Punkte zu erreichen.

Die Bearbeitungszeit beträgt 90 Minuten.

Erlaubte Hilfsmittel:

- Taschenrechner (nicht programmierbar)
- Ein DIN-A4 Blatt mit handschriftlichen Notizen (Vorder- und Rückseite)

Viel Erfolg!

Vom Prüfer auszufüllen:

Punkte Aufgabe 1 ☐ / 18

Punkte Aufgabe 2 ☐ / 18

Punkte Aufgabe 3 ☐ / 18

Punkte Aufgabe 4 ☐ / 18

Punkte Aufgabe 5 ☐ / 18

Gesamtpunkte ☐ / 72 **Note:**

Aufgabe 1

Betrachten sie das Regressionsmodell

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_K x_{iK} + u_i,\ i = 1, \ldots, n$$

a) Schreiben sie obige Regressionsgleichung in Matrixnotation auf und erklären sie die Bedeutung aller Variablen aus dieser Gleichung. Benennen sie ebenfalls die Ordnungen (# Zeilen, # Spalten) aller Matrizen und Vektoren.

b) Benennen sie die Annahmen **MLR 1** bis **MLR 6** an das Regressionsmodell.

Betrachten sie einen linearen Schätzer $\check{\boldsymbol{\beta}} = \mathbf{A}\mathbf{y}$ für $\boldsymbol{\beta}$, wobei $\mathbf{A}$ eine feste Matrix und $\mathbf{y}$ der Vektor der Beobachtungen y_1 bis y_n sei.

c) Welche Ordnung hat $\mathbf{A}$ und welche Bedingung muss $\mathbf{A}$ erfüllen, damit $\check{\boldsymbol{\beta}}$ erwartungstreu für $\boldsymbol{\beta}$ ist?

d) Welche der Annahmen **MLR 1** bis **MLR 6** müssen erfüllt sein, damit $\check{\boldsymbol{\beta}}$ erwartungstreu für $\boldsymbol{\beta}$ ist? Welche Annahmen werden hierfür nicht benötigt?

e) Zeigen sie, dass unter ihren Antworten zu c) und d) $\check{\boldsymbol{\beta}}$ tatsächlich erwartungstreu für $\boldsymbol{\beta}$ ist.

f) Geben sie ein Beispiel für einen linearen unverzerrten Schätzer für $\boldsymbol{\beta}$, welcher sich auf Grundlage der Daten aus der Stichprobe berechnen lässt und benennen sie die Matrix $\mathbf{A}$.

Aufgabe 2

Es bezeichne $\Sigma(\mathbf{v})$ die Varianz-Kovarianz-Matrix eines Zufallsvektors $\mathbf{v}$ der Länge $n \geq 2$.

a) Wie lautet die Definition von $\Sigma(\mathbf{v})$?

b) Welche Eigenschaften muss die Matrix $\Sigma(\mathbf{v})$ erfüllen, damit sie positiv definit symmetrisch ist?

c) Nehmen sie an, der Zufallsvektor $\mathbf{v}$ habe die Länge 2 und es gelte $v_2 = 1 + 2v_1$ und $Var(v_1) = 1$. Geben sie $\Sigma(\mathbf{v})$ an. Ist $\Sigma(\mathbf{v})$ positiv definit symmetrisch?

Für die Störterme $\mathbf{u} \in \mathbb{R}^n$ gelte

$$\Sigma(\mathbf{u}|\mathbf{X}) = \sigma^2 \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \text{ mit } \sigma^2 > 0.$$

d) Ist die Varianz-Kovarianz-Matrix $\Sigma(\mathbf{u}|\mathbf{X})$ positiv definit symmetrisch?

e) Wie lautet die auf $\mathbf{X}$ bedingte Varianz-Kovarianz-Matrix von $\mathbf{y}$ unter den Annahmen **MLR 1** und **MLR 5**?

f) Wie lautet die auf $\mathbf{X}$ bedingte Varianz-Kovarianz-Matrix von $\check{\boldsymbol{\beta}}$ aus Aufgabe 1?

Aufgabe 3

Das Regressionsmodell

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

erfülle die Annahmen **MLR 1, 2, 3, 4**. Für die Varianz-Kovarianz-Matrix der Störterme gelte

$$\mathbb{\Sigma}\left(\mathbf{u}|\mathbf{X}\right) = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix} ,$$

wobei $\sigma_i^2 > 0$ für alle $i = 1, \dots, n$ und $\sigma_t^2 \neq \sigma_s^2$ für mindestens ein Paar $t, s = 1, \dots, n$, $t \neq s$.

a) Liegt hier eine Verletzung von Annahme **MLR 5** vor? Falls ja, wie wird diese Verletzung bezeichnet?

b) Ist der OLS-Schätzer $\hat{\boldsymbol{\beta}}$ erwartungstreu für $\boldsymbol{\beta}$? Könnte anhand $\sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}$ als Schätzer für $\mathbb{\Sigma}\,(\hat{\boldsymbol{\beta}})$ die Nullhypothese $H_0: \ \beta_1 = 0$ valide getestet werden?

c) Nehmen sie an, die σ_i^2 seien für alle $i = 1, \dots, n$ bekannt. Wie berechnet sich in diesem Fall der beste lineare unverzerrte Schätzer $\tilde{\boldsymbol{\beta}}$ für $\boldsymbol{\beta}$? Könnte anhand von $\tilde{\boldsymbol{\beta}}$ und $\mathbb{\Sigma}\,(\tilde{\boldsymbol{\beta}})$ die Nullhypothese $H_0: \ \beta_1 = 0$ valide getestet werden?

d) Nehmen sie nun abweichend zu c) an, die σ_i^2 seien für kein $i = 1, \dots, n$ bekannt. Es sei ebenfalls unklar, ob **MLR 5** verletzt wäre. Nennen sie zwei statistische Tests, welche die Nullhypothese homoskedastischer Störterme zurückweisen können.

e) Beschreiben sie einen statistischen Test, welcher die Nullhypothese homoskedastischer Störterme zurückweisen kann.

f) Nehmen sie an, ein geeigneter statistischer Test würde die Nullhypothese homoskedastischer Störtermne zurückweisen. Wie würden sie vorgehen um die Nullhypothese $H_0 : \beta_1 = 0$ valide zu testen? Antworten sie in wenigen Sätzen.

Aufgabe 4

Auf Grundlage des Datensatzes `smoke.xls` werde der Zusammenhang zwischen dem durchschnittlichen Konsum von Zigaretten pro Tag (`cigs`) und Schuljahren (`educ`), Packungspreis (`cigpric`), Alter und Alter^2 (`age`, `agesq`) und Jahreseinkommen (`income`, in $) untersucht. Der Output der linearen Regressionsanalyse sei wie folgt:

Modell 1: KQ, benutze die Beobachtungen 1–807
Abhängige Variable: cigs

	Koeffizient	Std. Fehler	t-Quotient	p-Wert
const	1,87774	6,87287	0,2732	0,7848
educ	−0,504037	0,168659	−2,988	0,0029
cigpric	−0,0345002	0,100216	−0,3443	0,7307
age	0,796047	0,159838	4,980	0,0000
agesq	−0,00927067	0,00174150	−5,323	0,0000
income	4,13093e–05	5,68945e–05	0,7261	0,4680

Mittel abhängige Var.	8,686493	Stdabw. abhängige Var.	13,72152
Summe quad. Residuen	145026,3	Stdfehler Regression	13,45573
R^2	0,044331	Korrigiertes R^2	0,038365
$F(5, 801)$	7,431220	P-Wert(F)	7,94e–07

Gehen sie bei der Beantwortung der folgenden Fragen davon aus, dass die Modellannahmen **MLR 1** bis **6** erfüllt seien.

a) Welche der folgenden Aussagen ist richtig?

A) Abgesehen von der Konstante ist der Regressor `cigpric` ist am signifikantesten, da dessen p-Wert am größten ist.

B) Annahme **MLR 3** ist verletzt, da in der Regressormatrix sowohl Alter (age), als auch Alter^2 (`agesq`) enthalten ist.

C) Die Summe der quadratischen Abweichungen von `cigs` von dessen Durchschnitt beträgt 145026,3.

D) Etwa 0,4% der Variation von `cigs` wird durch die Variation der Regressoren erklärt.

b) Welche der folgenden Aussagen ist richtig?

A) Die Nullhypothese, dass der Parameter des Regressors age gleich null ist kann nur zum Siginifikanzniveau $\alpha = 15,98$ zurückgewiesen werden, da der Standardfehler dessen Schätzers den Wert 0,1598 hat.

B) Die Parameter der Regressoren sind gemeinsam signifikant von null verschieden.

C) Der Schätzer $\hat{\sigma}^2$ für die Varianz der Störterme, σ^2 hat den Wert 13,45573.

D) Im Mittel beträgt das Jahreseinkommen (`income`) 41309$

Aufgabenteil c) von Aufgabe 4 auf Seite 6!

Noch Aufgabe 4

Es werde eine zweite Regression durchgeführt:

Modell 2: KQ, benutze die Beobachtungen 1–807
Abhängige Variable: cigs

	Koeffizient	Std. Fehler	t-Quotient	p-Wert
const	−0,331768	3,50945	−0,09454	0,9247
educ	−0,471677	0,161790	−2,915	0,0037
age	0,823768	0,154691	5,325	0,0000
agesq	−0,00958225	0,00168340	−5,692	0,0000

Mittel abhängige Var.	8,686493	Stdabw. abhängige Var.	13,72152
Summe quad. Residuen	145140,0	Stdfehler Regression	13,44422
R^2	0,043582	Korrigiertes R^2	0,040009
$F(3, 803)$	12,19696	P-Wert(F)	8,24e–08

c) Unterscheidet sich der Erklärungsgehalt von Modell 1 und Modell 2?
Begründen sie ihre Aussage mit Bezugnahme auf konkrete Angaben in den beiden Outputs.

Beachten sie nun einen statistischen Test, welcher geeignet wäre zu prüfen, ob die Parameter von `cigpric` und `income` in Modell 1 beide gleich null sind.

d) Nennen sie eine geeignete Teststatistik, welche im obigen Test Anwendung findet. Wie lautet der konkrete Wert dieser Teststatistik hier?

e) Wie ist diese Teststatistik und den Modellannahmen **MLR 1** bis **6** verteilt? Wie lautet der kritische Wert für diese Teststatistik für $\alpha = 1\%$?

f) Beschreiben sie den statistischen Test ausführlich und treffen sie eine begründete Testentscheidung.

Aufgabe 5

Sie untersuchen den Einfluss von der monatlichen Temperatur c_t auf den monatlichen Pegel des Rheines rp_t. Dafür schätzen sie das folgende Modell:

$$rp_t = \beta_0 + \beta_1 c_t + u_t$$

a) Sie haben Grund für den Verdacht, dass der Störterm u_t autokorreliert sein könnte. Nennen sie eine MLR Bedinung, welche dadurch verletzt werden könnte und die möglichen Folgen für die OLS-Schätzung.

b) Wofür steht das Akronym AR(1)? Welche Bedingung entspricht einem *stationären* AR(1)-Prozess? Nehmen sie im Folgenden an, dass u_t ein stationärer AR(1)-Prozess sei.

c) Sie schätzen folgendes Modell unter Nutzung der OLS Residuen:

$$u_t = \rho u_{t-1} + e_t$$

und erhalten eine OLS-Schätzung von $\hat{\rho} = .54$. Nennen sie eine geeignete Teststatistik (Name, Formel und Hypothesen) für dieses Modell.

d) Bestimmen sie den *ungefähren* Wert für ihre gewählte Teststatistik.

e) Ihre Teststatistik liefert Evidenz für autokorrelierte Störterme. Nennen sie die FGLS Transformation, welche unter diesen Umständen sinnvoll wäre.

f) Welche Eigenschaften haben ihre Störterme nach der Transformation?

Kritische Werte der t-Verteilung

		Signifikanzniveau				
	einseitig:	10%	5%	2,5%	1%	0,5%
	zweiseitig:	20%	10%	5%	2%	1%
Freiheits-	1	3,078	6,314	12,706	31,821	63,657
grade	2	1,886	2,920	4,303	6,965	9,925
	3	1,638	2,353	3,182	4,541	5,841
	4	1,533	2,132	2,776	3,747	4,604
	5	1,476	2,015	2,571	3,365	4,032
	6	1,440	1,943	2,447	3,143	3,707
	7	1,415	1,895	2,365	2,998	3,499
	8	1,397	1,860	2,306	2,896	3,355
	9	1,383	1,833	2,262	2,821	3,250
	10	1,372	1,812	2,228	2,764	3,169
	11	1,363	1,796	2,201	2,718	3,106
	12	1,356	1,782	2,179	2,681	3,055
	13	1,350	1,771	2,160	2,650	3,012
	14	1,345	1,761	2,145	2,624	2,977
	15	1,341	1,753	2,131	2,602	2,947
	16	1,337	1,746	2,120	2,583	2,921
	17	1,333	1,740	2,110	2,567	2,898
	18	1,330	1,734	2,101	2,552	2,878
	19	1,328	1,729	2,093	2,539	2,861
	20	1,325	1,725	2,086	2,528	2,845
	21	1,323	1,721	2,080	2,518	2,831
	22	1,321	1,717	2,074	2,508	2,819
	23	1,319	1,714	2,069	2,500	2,807
	24	1,318	1,711	2,064	2,492	2,797
	25	1,316	1,708	2,060	2,485	2,787
	26	1,315	1,706	2,056	2,479	2,779
	27	1,314	1,703	2,052	2,473	2,771
	28	1,313	1,701	2,048	2,467	2,763
	29	1,311	1,699	2,045	2,462	2,756
	30	1,310	1,697	2,042	2,457	2,750
	40	1,303	1,684	2,021	2,423	2,704
	60	1,296	1,671	2,000	2,390	2,660
	90	1,291	1,662	1,987	2,368	2,632
	120	1,289	1,658	1,980	2,358	2,617
	∞	1,282	1,645	1,960	2,326	2,576

Kritische Werte der F-Verteilung zum Signifikanzniveau von 1%

		Anzahl der Restriktionen									
		1	2	3	4	5	6	7	8	9	10
$n-k-1$	10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85
	11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54
	12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30
	13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10
	14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94
	15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80
	16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69
	17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59
	18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51
	19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43
	20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37
	21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31
	22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26
	23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21
	24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17
	25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13
	26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09
	27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06
	28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03
	29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00
	30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98
	40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80
	60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63
	90	6,93	4,85	4,01	3,54	3,23	3,01	2,84	2,72	2,61	2,52
	120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47
	∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32

Kritische Werte der F-Verteilung zum Signifikanzniveau von 5%

		Anzahl der Restriktionen									
		1	2	3	4	5	6	7	8	9	10
$n-k-1$	10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
	11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
	12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
	13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
	14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
	15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
	16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
	17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
	18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
	19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
	20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
	21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32
	22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30
	23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27
	24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25
	25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24
	26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22
	27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20
	28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19
	29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18
	30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16
	40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08
	60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99
	90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,99	1,94
	120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91
	∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83

Kritische Werte der χ^2-Verteilung

		Signifikanzniveau		
		10%	5%	1%
Freiheits-	1	2,71	3,84	6,63
Grade	2	4,61	5,99	9,21
	3	6,25	7,81	11,34
	4	7,78	9,49	13,28
	5	9,24	11,07	15,09
	6	10,64	12,59	16,81
	7	12,02	14,07	18,48
	8	13,36	15,51	20,09
	9	14,68	16,92	21,67
	10	15,99	18,31	23,21
	11	17,28	19,68	24,72
	12	18,55	21,03	26,22
	13	19,81	22,36	27,69
	14	21,06	23,68	29,14
	15	22,31	25,00	30,58
	16	23,54	26,30	32,00
	17	24,77	27,59	33,41
	18	25,99	28,87	34,81
	19	27,20	30,14	36,19
	20	28,41	31,41	37,57
	21	29,62	32,67	38,93
	22	30,81	33,92	40,29
	23	32,01	35,17	41,64
	24	33,20	36,42	42,98
	25	34,38	37,65	44,31
	26	35,56	38,89	45,64
	27	36,74	40,11	46,96
	28	37,92	41,34	48,28
	29	39,09	42,56	49,59
	30	40,26	43,77	50,89

Lösung für Aufgabe 1

a) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, wobei

- $\mathbf{y} \in \mathbb{R}^n$ Regressand
- $\mathbf{X} \in \mathbb{R}^{n \times K+1}$ Regressormatrix
- $\boldsymbol{\beta} \in \mathbb{R}^{K+1}$ Parametervektor
- $\mathbf{u} \in \mathbb{R}^n$ Störterme

b)

- **MLR 1** Regressionsmodell linear in $\boldsymbol{\beta}$
- **MLR 2** $\mathbf{y}, \mathbf{X}$ Zufallsstichprobe
- **MLR 3** $rk(X) = K + 1$ voller Spaltenrang von $\mathbf{X}$
- **MLR 4** $\mathbb{E}[\mathbf{u}|\mathbf{X}] = \mathbb{E}[\mathbf{u}] = \mathbf{0}$ strikte Exogenität
- **MLR 5** $\Sigma(\mathbf{u}) = \sigma^2\mathbf{I}$ Homoskedastizität, serielle Unkorreliertheit
- **MLR 6** $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ normalverteilte Störterme

c) $\mathbf{A} \in \mathbb{R}^{K+1 \times n}$, $\mathbf{AX} = \mathbf{I}$ ($\mathbf{A}$ Pseudoinverse von $\mathbf{X}$)

d) Es müssen **MLR 1,3,4** erfüllt sein, die Annahmen **MLR 2,5,6** müssen nicht erfüllt sein.

e) $\mathbb{E}[\mathbf{Ay}|\mathbf{X}] = \mathbf{A}\mathbb{E}[\underbrace{\mathbf{y}}_{\substack{\mathbf{X}\boldsymbol{\beta}+\mathbf{u} \\ (MLR\ 1)}}|\mathbf{X}] = \mathbf{AX}\boldsymbol{\beta} + \mathbf{A}\underbrace{\mathbb{E}[\mathbf{u}|\mathbf{X}]}_{\substack{=0 \\ (MLR\ 4)}} = \boldsymbol{\beta}\ \forall \in \mathbb{R}^{K+1} \Leftrightarrow \underset{(MLR\ 3)}{rk(\mathbf{X}) = K+1} \;\&\; \mathbf{AX} = \mathbf{I}$

f) $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$

Lösung zu Aufgabe 2

a)

$$\Sigma(\mathbf{v}) = \mathbb{E}\left[(\mathbf{v} - \mathbb{E}[\mathbf{v}]) (\mathbf{v} - \mathbb{E}[\mathbf{v}])'\right] = \begin{pmatrix} Var(v_1) & Cov(v_1, v_2) & \dots & Cov(v_1, v_n) \\ Cov(v_2, v_1) & Var(v_2) & \dots & Cov(v_2, v_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(v_n, v_1) & Cov(v_n, v_2) & \dots & Var(v_n) \end{pmatrix}$$

b) Eine Matrix $\mathbf{B} \in \mathbb{R}^{n \times n}$ heißt **positiv definit**, falls $\mathbf{x}'\mathbf{B}\mathbf{x} > 0$ für alle $\mathbf{x} \in \mathbb{R}^n$. $\mathbf{B}$ heißt **symmetrisch**, falls $\mathbf{B}' = \mathbf{B}$.

c) Mit $Cov(v_1, v_2) = Cov(v_1, 1 + 2v_1) = 2Var(v_1)$ und $Var(v_2) = Var(1 + 2v_1) = 4Var(v_1)$ gilt für $\Sigma(\mathbf{v}) = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$:

$$(x_1, x_2) \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1^2 + 4x_1x_2 + 4x_2^2 = (x_1 + 2x_2)^2 = 0 \Leftrightarrow x_1 = -2x_2$$

Zum Beispiel für den Vektor $(x_1, x_2) = (-2, 1)$ erfüllt die Varianz-Kovarianz-Matrix also nicht die Bedingung positiv definit.

d) Für einen beliebigen Vektor $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} \neq \mathbf{0}$ gilt

$$\mathbf{x}'\Sigma(\mathbf{u}|\mathbf{X})\mathbf{x} = \sigma^2 \sum_{i=1}^{n} x_i^2$$

Die Summe der Quadrate ist positiv, falls mindestens ein Eintrag von $\mathbf{x}$ von null verschieden ist. Also ist $\Sigma(\mathbf{u}|\mathbf{X})$ positiv definit. Da abseits der Diagonalen alle Einträge identisch sind, ist die Matrix auch symmetrisch.

e) Für $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ (MLR 1) gilt:

$$\Sigma(\mathbf{y}|\mathbf{X}) = \Sigma(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}|\mathbf{X}) = \Sigma(\mathbf{u}|\mathbf{X}) = \sigma^2\mathbf{I} \ (MLR\ 5)$$

f) Für $\check{\boldsymbol{\beta}} = \mathbf{A}\mathbf{y}$ gilt

$$\Sigma(\check{\boldsymbol{\beta}}|\mathbf{X}) = \Sigma(\mathbf{A}\mathbf{y}|\mathbf{X}) = \mathbf{A}\Sigma(\mathbf{y}|\mathbf{X})\mathbf{A}' = \sigma^2\mathbf{A}\mathbf{A}'$$

Lösung zu Aufgabe 3

a) Annahme **MLR 5** fordert $\Sigma(\mathbf{u}) = \sigma^2\mathbf{I}$ bzw. $\sigma_i^2 = \sigma^2 \; \forall \; i = 1, \ldots, n$. Also ist **MLR 5** verletzt. Diese Verletzung bezeichnet man als Heteroskedastizität.

b) $\hat{\boldsymbol{\beta}}$ ist erwartungstreu, da für diese Eigenschaft lediglich die Annahmen **MLR 1,3,4** benötigt werden. Da $\Sigma(\hat{\boldsymbol{\beta}}|\mathbf{X})$ wegen der heteroskedastischen Störterme nicht $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ entspricht, würden die Standardabweichungen von $\hat{\beta}_j$ falsch geschätzt. Damit wäre der t-Test nicht valide.

c) Wenn σ_i^2 für alle i bekannt ist, kann die Matrix

$$\mathbf{P} = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & \ldots & 0 \\ 0 & \frac{1}{\sigma_2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \frac{1}{\sigma_n} \end{pmatrix}$$

berechnet werden. Der GLS-Schätzer $\tilde{\boldsymbol{\beta}}$ berechnet sich dann wie der OLS-Schätzer, wobei anstelle der Daten $\mathbf{y}, \mathbf{X}$ die transformierten Daten $\tilde{\mathbf{y}}, \tilde{\mathbf{X}} = \mathbf{Py}, \mathbf{PX}$ benutzt werden:

$$\tilde{\boldsymbol{\beta}} = \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{y}}$$

Dies entspricht dem Ausdruck

$$\tilde{\boldsymbol{\beta}} = \left(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y} \; ,$$

wobei $\boldsymbol{\Omega} = \Sigma(\mathbf{u})$.

d) Breusch-Pagan-Test, White-Test, alternativer White-Test

e) Breusch-Pagan-Test:
Schätze mit OLS das Modell $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ und speichere die quadrierten Residuen $\hat{u}_i^2$.
Schätze mit OLS das Modell

$$\hat{\mathbf{u}}^2 = \mathbf{X}\boldsymbol{\delta} + \boldsymbol{\epsilon}$$

und führe einen globalen F-Test auf $\delta_1 = \ldots = \delta_K = 0$ aus. Wird diese Hypothese verworfen, liegt Heteroskedastizität vor.

f) Statistische Tests mit heteroskedastizitäts-robusten Standardfehlern sind valide. Alternativ kann auch ein konkreter funktionialer Zusammenhang von σ_i^2 und $\mathbf{X}$ angenommen werden (wie z.B. multiplikative Heteroskedastizität) und die Parameter dieser Funktion geschätzt werden. Mit den Schätzern für die Parameter können die prognositizierten quadrierten Residuen berechnet werden. Diese dienen dann als Schätzer für die unbekannten Varianzen σ_i^2 und werden benutzt um die Daten geeignet zu transformieren. Der OLS-Schätzer unter Verwendung der transformierten Daten ist dann ein FGLS-Schätzer.

Lösung zu Aufgabe 4

a) Fehler in der Aufgabenstellung: Alle vier Antworten sind falsch. Alle Personen bekommen 3 Punkte für a).

A) Der p-Wert gibt das kleinste α an zu welchem ein Parameter signifikant von null verschieden ist. Damit wäre der Parameter zu `cigpric` zu den gängigen αs (1%, 5%, 10%) nicht signifikant von null verschieden (`educ`,`age`,`agesq` aber schon). → A) falsch

B) x und x^2 sind nicht linear abhängig. MLR 3 ist daher nicht wegen `age` und `agesq` verletzt. → [B)] falsch

C) Die Summe der quadrierten Residuen beträgt 145026,3. Die in der Aussage gemeinte Summe beträgt $(n-1)$(Stdabw. abhängige Var)$^2 = 806 * 13,72152^2 = 151753,77$. Test $1 - R^2 = 0,955669 = 145026,3/151753,77 = SSR/SST$ → [C)] falsch

D) falsch, wegen $R^2 = 0,044331 \approx 4,4\%$.

b)

A) Die t-Statistik für die genannte Nullhypothese ist dem Output zu entnehmen: $t = 4,98$ und liegt damit weit über dem kritischen Wert der t-Verteilung mit 801 Freiheitsgraden und $\alpha = 1\%$. Daher kann die Nullhypothese für $\alpha = 1$ zurückgewiesen werden. → [A)] falsch.

B) richtig, da der P-Wert für den globalen F-Test laut Output weit unter 1% liegt.

C) Der Schätzer $\hat{\sigma}$ für die Standardabweichung σ der Störterme hat den Wert $13,45573$. → [C)] falsch.

D) Das mittlere Jahreseinkommen ist dem Output nicht zu entnehmen. Die Zahl 4,13093e-05 bezieht sich auf den Schätzer des Parameters des Jahreseinkommens. → [D)] falsch

c) Der Unterschied im Erklärungsgehalt der Modelle 1 und 2 wird durch die unterschiedlichen Bestimmtheitsmaße ($R_1^2 = 0,044331$ vs. $R_2^2 = 0,043582$) oder durch die unterschiedlichen Summen der quadrierten Residuen ($SSR_1 = 145026,3$ vs $SSR_2 = 145140$) berichtet. Allerdings fallen diese Unterschiede augenscheinlich sehr schwach aus.

d) Hier findet die F-Statistik Anwendung. Die Berechnung anhand SSR lautet:

$$F = \frac{(SSR_2 - SSR_1)/J}{SSR_1/(n-K-1)} = \frac{(145140 - 145026,3)/2}{145026,3/801} = 0,3160$$

Die Berechnung anhand R^2 lautet:

$$F = \frac{(R_1^2 - R_2^2)/J}{(1-R_1^2)/(n-K-1)} = \frac{(0,044331 - 0,043582)/2}{(1-0,044331)/801} = 0,3139$$

e) Die F-Statistik ist F-Verteilt; hier mit 2 und 801 Freiheitsgraden. Der kritische Wert für $\alpha = 1\%$ liegt zwischen 4,79 und 4,61 und damit über dem Wert der F-Statistik.

f) Der F-Test überprüft, ob eine oder mehrere lineare Restriktionen an die Parameter halten. Hierfür werden zwei Regressionen durchgeführt, eine ohne Beachtung der Restriktionen (hier Modell 1) und eine mit Beachtung der Restriktionen (hier Modell 2). Die genannte F-Statistik ist unter der Hypothese der Restriktionen F-verteilt. Die Nullhypothese wird verworfen, falls der Wert der F-Statistik oberhalb des entsprechenden kritischen Wertes der F-Verteilung liegt. Dies ist hier nicht der Fall (F≈ 0,32 ; $c \geq 4,61$). Also kann die Nullhypothese, dass die zu `cigpric` und `income` gehörigen Parameter beide gleich null sind nicht verworfen werden.

Lösung zu Aufgabe 5

a) *MLR 5 und 6 sind verletzt.*
Die Folge ist, dass die Schätzung für den Standardfehler der OLS-Schätzung $se(\hat{\beta})$ *nicht mehr valide ist und p-Werten des* t*-Tests nicht mehr vertraut werden kann.*

b) AR(1) steht für autoregressiver Prozess erster Ordnung, $u_t = \rho u_{t-1} + \epsilon$, wobei Stationärität $|\rho| < 1$ verlangt.

c) *Name: Durbin Watson.*
Dieser testet die Hypothesen:

$$H_0 : \rho = 0 \quad H_1 : \rho > 0$$

Die Berechnung erfolgt über die OLS-Residuen:

$$DW = \frac{\sum_{t=2}^{T} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{T} \hat{u}_t^2}$$

d) *Hier wird auf den approximativen Zusammenhang zwischen der Schätzung* $\hat{\rho}$ *und* DW *angespielt. Die korrekte Antwort lautet*:

$$DW \approx 2(1 - \hat{\rho}) = 2(1 - .54) = 0.92$$

e) *Die korrekte Transformation lautet*:

$$rp_t - .54rp_{t-1} = \beta_0 + \beta_1(c_t - .54c_{t-1}) + u_t$$

f) *Seriell unkorreliert.*