

Klausur zu Ökonometrie (Master)

Technische Universität Dortmund

Fakultät Wirtschaftswissenschaften

28. Juli 2023

Bitte tragen Sie Ihre Daten sorgfältig und leserlich ein:

Matrikelnummer

Nachname _____

Studiengang _____

Vorname _____

Bearbeitungshinweise:

Diese Klausur besteht aus fünf Aufgaben, welche alle zu bearbeiten sind.

Alle Antworten sind zu begründen.

Bitte verwenden Sie einen Kugelschreiber oder nicht zu starken Filzstift.

Nicht-programmierbare Taschenrechner sind als Hilfsmittel für diese Klausur zugelassen.

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

Bei 36 von maximal 90 erreichbaren Punkten ist die Klausur bestanden.

Die Bearbeitungszeit beträgt 90 Minuten.

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 / 90

Note:

Aufgabe 1

Für ein lineares Regressionsmodell

MLR 1

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

MLR 4

mit $\mathbf{y}, \mathbf{u} \in \mathbb{R}^n$, $X \in \mathbb{R}^{n \times (k+1)}$ und $\boldsymbol{\beta} \in \mathbb{R}^{k+1}$ gelten die in der Vorlesung getroffenen Annahmen $E[\mathbf{u}|X] = 0$ und $rk(X) = k + 1$. MLR 3

Für einen Vektor $\mathbf{b} \in \mathbb{R}^{k+1}$ sei $\hat{\mathbf{u}}(\mathbf{b})$ definiert durch $\hat{\mathbf{u}}(\mathbf{b}) = \mathbf{y} - X\mathbf{b}$.

$$\rightarrow \sqrt{\sum_{i=1}^n (\hat{u}_i(\mathbf{b}))^2}$$

a) Geben Sie denjenigen Vektor $\mathbf{b}^* \in \mathbb{R}^{k+1}$ an, welcher die Länge von $\hat{\mathbf{u}}(\mathbf{b})$, also $\|\hat{\mathbf{u}}(\mathbf{b})\|$, minimiert.

Betrachten Sie nun eine beliebige lineare Funktion f von \mathbf{y} , dargestellt als $f(\mathbf{y}) = A\mathbf{y}$, wobei $A \in \mathbb{R}^{(k+1) \times n}$ (A ist eine Matrix mit $k + 1$ Zeilen und n Spalten).

b) Welche Bedingung für das Matrixprodukt AX muss gelten, damit $A\mathbf{y}$ ein unverzerrter Schätzer für $\boldsymbol{\beta}$ ist?

c) Können Sie den in Aufgabenteil a) gefundenen Vektor \mathbf{b}^* als lineare Funktion $f(\mathbf{y}) = A\mathbf{y}$ darstellen? Falls ja: Wie lautet die Matrix A ? Falls nein: Wie lautet die Matrix A für den OLS-Schätzer $\hat{\boldsymbol{\beta}}$?

d) Begründen Sie ob die in b) identifizierte Bedingung für Erwartungstreue für die in c) gefundene Matrix A erfüllt ist!

a) $\mathbf{b}^* = \hat{\boldsymbol{\beta}}$

b)
$$E[A\mathbf{y}|X] = A E[\mathbf{y}|X] = A E[X\boldsymbol{\beta} + \mathbf{u}|X] = AX\boldsymbol{\beta} + \underbrace{A E[\mathbf{u}|X]}_{=0}$$
$$= \underbrace{AX}_{I} \boldsymbol{\beta} \stackrel{!}{=} \boldsymbol{\beta} \text{ für alle } \boldsymbol{\beta} \in \mathbb{R}^{k+1}$$

$$\Rightarrow AX = I$$

c)
$$\hat{\boldsymbol{\beta}} = \underbrace{(X'X)^{-1}X'}_A \mathbf{y} \quad A = (X'X)^{-1}X'$$

d) prüfe $AX = I$: $(X'X)^{-1}X'X = I$

Aufgabe 2

Betrachten Sie folgende OLS-Regressionen in Bezug auf den Datensatz `bwght.xls`, wobei

- `bwght` das Geburtsgewicht in Unzen (1 Unze \approx 28 Gramm),
- `cigs` die Anzahl der gerauchten Zigaretten pro Tag während der Schwangerschaft und
- `faminc` das Familieneinkommen in 1000\$ von 1988 bezeichnen:

$$\text{Modell 1: } \widehat{\text{bwght}} = 119,772 - 0,513772 \text{ cigs} \quad R^2 = 0,022729$$

(0,57234) (0,090491)

$$\text{Modell 2: } \widehat{\text{bwght}} = 116,974 - 0,463408 \text{ cigs} + 0,0927647 \text{ faminc} \quad R^2 = 0,029805$$

(1,0490) (0,091577) (0,029188)

$$\text{Modell 3: } \widehat{\text{cigs}} = 3,68811 - 0,0551538 \text{ faminc} \quad R^2 = 0,029945$$

(0,29130) (0,0084321)

Der Stichprobenumfang beträgt für alle Regressionen $n = 1388$, die Standardfehler werden jeweils in Klammern angegeben.

Nehmen Sie an, dass für Modell 2 die in der Vorlesung getroffenen Annahmen MLR 4 und MLR 6 an die unbeobachtbaren Störterme erfüllt seien: $E[\mathbf{u}|X] = \mathbf{0}$ und $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_n)$

- a) Mit welchem Test würden Sie die Hypothese, dass ein Parameter in Modell 2 gleich null ist, untersuchen?
- b) Interpretieren Sie die Regressionsergebnisse für Modell 2.
Können Sie die Hypothese aus a) für die einzelnen Parameter zurückweisen, ohne exakt zu rechnen?
- c) Sind die Regressoren `cigs` und `faminc` korreliert?
- d) Wird in Modell 1 der Parameter für `cigs` erwartungstreu geschätzt?

Aufgabe 3

Betrachten Sie beruhend auf dem gleichen Datensatz wie Aufgabe 2, `bwght.xls`, folgende Regression:

$$\begin{aligned} \text{Modell 4: } \widehat{\text{bwght}} &= 115,998 - 0,470802 \text{ cigs} + 0,0925442 \text{ faminc} \\ &\quad \begin{matrix} (12,407) & (0,092782) & (0,032898) \end{matrix} \\ &- 0,0180842 \text{ motheduc} - 0,0121839 \text{ cigprice} + 0,144697 \text{ cigtax} \quad R^2 = 0,032271 \\ &\quad \begin{matrix} (0,25861) & (0,11073) & (0,14517) \end{matrix} \end{aligned}$$

Wie in Aufgabe 2 ist der Stichprobenumfang $n = 1388$, die Standardfehler werden jeweils in Klammern angegeben.

Die Bedeutungen der drei neuen Regressoren lauten:

`motheduc`: Bildungsjahre der Mutter

`ciprice`: Zigarettenpreis des jeweiligen Bundesstaates von 1988

`cigtax`: Zigarettensteuer des jeweiligen Bundesstaates von 1988

Nehmen Sie wie in Aufgabe 2 an, dass die Störterme homoskedastisch und seriell unkorreliert normalverteilt sind.

- Ist einer der Parameter der drei neuen Regressoren **einzeln** signifikant von null verschieden? Beschreiben Sie das anzuwendende Testverfahren. Die Testentscheidung können Sie mit der in der Vorlesung besprochenen Daumenregel treffen.
- Welchen Test wenden Sie an, um die **gemeinsame** Signifikanz der Parameter der drei neuen Regressoren zu prüfen?
- Beschreiben Sie das Testverfahren und geben Sie an, welche Informationen Sie hierfür benötigen.
- Treffen und begründen Sie eine Testentscheidung

Hinweis: Sie finden geeignete statistische Tabellen im Anhang dieser Klausur.

Aufgabe 4

- a) Was bedeuten OLS und GLS und worin besteht der Hauptunterschied zwischen diesen beiden Konzepten?
- b) Was sind Querschnittsdaten und inwiefern ist der Unterschied zwischen OLS und GLS vor allem bei Querschnittsdaten relevant?
- c) Was sind Zeitreihen und inwiefern ist der Unterschied zwischen OLS und GLS vor allem bei Zeitreihen relevant?
- d) Beschreiben Sie für Querschnittsdaten oder für Zeitreihendaten einen Test, ob OLS anwendbar ist.

Aufgabe 5

Betrachten Sie auf Grundlage des Datensatzes `card.xls` folgende Regressionen:

$$\begin{aligned} \text{Modell 1: } \widehat{\text{lwage}} &= 4,66603 + 0,0931680 \text{ educ} + 0,0406574 \text{ exper} \\ &\quad (0,063790) \quad (0,0036118) \quad (0,0023344) \\ \text{Modell 2: } \widehat{\text{educ}} &= 16,5549 - 0,417871 \text{ exper} + 0,600044 \text{ nearc4} \\ &\quad (0,10442) \quad (0,0088591) \quad (0,078779) \\ \text{Modell 3: } \widehat{\text{lwage}} &= 1,79498 + 0,262043 \widehat{\text{educ}} + 0,111928 \text{ exper} \\ &\quad (0,48685) \quad (0,028615) \quad (0,012229) \end{aligned}$$

Die Stichprobengröße beträgt jeweils $n = 3010$, Standardfehler sind jeweils in Klammern.

Die Variablen haben folgende Bedeutungen:

- `lwage`: Logarithmus des Stundenlohns in US-cents von 1976
- `educ`: Schuljahre in 1976
- `exper`: Alter - educ - 6
- `nearc4`: Kindheit in der Nähe einer Uni mit vierjährigen Studiengängen

- a) Diskutieren Sie in Bezug zu Modell 1 welche der MLR-Annahmen wahrscheinlich verletzt ist.
- b) Nehmen Sie an, einer der Regressoren in Modell 1 wäre endogen. Welche Annahmen müsste eine Instrumentalvariable für den endogenen Regressor erfüllen?
- c) Diskutieren Sie, ob `nearc4` ein geeignetes Instrument für `educ` ist. Ist es besser geeignet als das in der Vorlesung vorgeschlagene Instrument `fatheduc` (Bildungsjahre des Vaters)?
- d) Berechnen Sie den IV-Schätzer $\hat{\beta}_{\text{educ}}^{IV}$, wobei Sie `educ` durch `nearc4` instrumentieren. Erklären Sie Ihren Rechenweg.

Kritische Werte der t -Verteilung

		Signifikanzniveau				
		10%	5%	2,5%	1%	0,5%
einseitig:		10%	5%	2,5%	1%	0,5%
zweiseitig:		20%	10%	5%	2%	1%
Freiheits- grade	1	3,078	6,314	12,706	31,821	63,657
	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- Grade	1	2,71	3,84	6,63
	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