Pobeklausur: "Einführung in die Ökonometrie"

Gesamtpunktzahl: 96

1. (20 Punkte)

Wir betrachten 2 (in den Koeffizienten) lineare Modelle mit folgenden Variablen:

- *cigs*: gerauchte Zigaretten pro Tag
- *educ*: Bildung (in Jahren)
- *lcigprice*: logarithmierter Zigarettenpreis (in cent pro Schachtel)
- *white*: =1 für Menschen mit weißer Hautfarbe
- *income*: Jahreseinkommen in €
- *age*: Alter (in Jahren)
- *restaurn*: =1 wenn Rauchverbot in Restaurants

Modell 1: OLS estimates using the 807 observations 1–807

Abhängige Variable: *cigs*

Variable	Koeffizient	Std. Fehler	<i>t</i> -Statistik	P-Wert
const	14,2155	24,1449	0,5888	0,5562
educ	−0,367797	0,169280	−2,1727	0,0301
lcigprice	−0,199496	5,87322	−0,0340	0,9729
white	−0,117611	1,48056	−0,0794	0,9367
income	0,000130676	5,60173e-05	2,3328	0,0199
age	−0,0438875	0,0287328	−1,5274	0,1270
restaurn	−2,9789	1,13847	−2,6166	0,0090

Summe der quadrierten Residuen	148864,
Standardfehler der Residuen ($\hat{\sigma}$)	13,6411
Unkorrigiertes R^2	0,0190413
Korrigiertes $\bar{R}^2$	0,0116842
$F(6, 800)$	2,58813
P-Wert für $F()$	0,0172411
Log-Likelihood	−3250,3
Akaike Informations-Kriterium	6514,66
Schwarz' Bayes-Kriterium	6547,51

Modell 2: OLS estimates using the 807 observations 1–807
 Abhängige Variable: *cigs*

Variable	Koeffizient	Std. Fehler	<i>t</i> -Statistik	P-Wert
const	7,0946	24,0109	0,2954	0,7677
educ	−0,324134	0,166878	−1,9423	0,0524
lcigpric	−0,484920	5,86470	−0,0827	0,9341
income	0,000130887	5,60285e-05	2,3361	0,0197
restaurn	−2,9133	1,13276	−2,5719	0,0103

Summe der quadrierten Residuen	149298,
Unkorrigiertes R^2	0,0161793
Korrigiertes $\bar{R}^2$	0,0112725
$F(4, 802)$	3,29730
P-Wert für $F()$	0,0108058
Log-Likelihood	−3251,5
Akaike Informations-Kriterium	6513,01
Schwarz' Bayes-Kriterium	6536,48

- Interpretieren Sie unabhängig von den Signifikanzen die Koeffizienten für *educ*, *lcigprice* und *restaurn* im Modell 1!
- Auf welchem Level (10%, 5%, 1%) sind die Koeffizienten aus (a) individuell signifikant?
- Erklärt Modell 2 signifikant einen Teil des Zigarettenkonsums und wie hoch ist das dazugehörige Signifikanzniveau (10%, 5%, 1%)?
- Führen Sie einen Test durch, um Modell 1 und 2 zu vergleichen, wählen Sie dazu ein geeignetes Signifikanzniveau!
- * Beschreiben Sie, wie Sie den Test aus (d) als Lagrange Multiplikator-Test (LM-Test) durchführen würden!
- Vergleichen Sie die Modelle 1 und 2 mit einem anderen (als in d) und e) beschriebenen) Kriterium.
- Berechnen Sie ein 95%iges Konfidenzintervall für den Koeffizienten vor *income* in Modell 1.

2. (20 Punkte)

- a) Auf welchem Intervall liegt der unkorrigierte R^2 und was sagt er aus?
- b) Nennen Sie 2 mögliche Gründe für verzerrte Schätzungen von Koeffizienten!
- c) Welche Auswirkungen hat Heteroskedastizität auf die Konsistenz der Schätzung mit OLS?
- d) Angenommen wir hätten folgendes (in den Koeffizienten) lineares Modell:

$$wage = \beta_0 + \beta_1 male + \beta_2 white + u$$

(*wage*: Stundenlohn, *male*: Dummy, =1 für Männer, *white*: Dummy, =1 für Menschen mit weißer Hautfarbe.

- I.) Entwickeln Sie ein Modell, in dem es möglich ist zu vergleichen, ob es signifikante Unterschiede im Lohn zwischen weißen Männern und nichtweißen Frauen gibt.
 - II.) Wie würden Sie den Lohnunterschied zwischen den beiden Gruppen in dem von Ihnen entwickelten Modell berechnen?
- e) Beschreiben Sie, wie Sie zwei in den Koeffizienten lineare **nichtverschachtelte** (non nested) Modelle mithilfe des
- I.) Non nested F-Test
 - II.) Davidson-MacKinnon-Test
- vergleichen würden. Nennen Sie bei einem von beiden Tests die möglichen Ergebnisse und die jeweilige Interpretation!

3. (8 Punkte)

Gegeben sei Y als abhängige Variable und X , A als erklärende Variablen. Die Variable A sei endogen und die Variable X exogen.

- a) Nennen Sie die Bedingungen, die ein valides Instrument (Z) erfüllen muss.
- b) Wie würde man eine der beiden Bedingungen für ein valides Instrument überprüfen?
- c) Beschreiben Sie, wie man mithilfe einer Zweistufenschätzung (mit Instrument Z) konsistente Koeffizientenschätzungen erhält.

4. (10 Punkte)

Gegeben sei folgendes in den Koeffizienten lineares Modell:

$$lwage = \beta_0 + \beta_1 age + \beta_2 exper + \beta_3 sibs + \beta_4 female + \beta_5 abil + u$$

Die Variablen in diesem Modell sind:

- lwage: logarithmierter Stundenlohn in €
- age: Alter in Jahren
- exper: Berufserfahrung in Jahren
- sibs: Anzahl der Geschwister
- female: =1 für Frauen
- abil: individuelle Fähigkeiten
- KWW: Punkte in einem Test auf Allgemeinbildung

Modell 1: OLS estimates using 852 observations from 1–935

Fehlende oder unvollständige Beobachtungen entfernt: 83

Abhängige Variable: lwage

Variable	Koeffizient	Std. Fehler	t-Statistik	P-Wert
const	6,14381	0,154839	39,6787	0,0000
age	0,0251372	0,00505755	4,9702	0,0000
exper	0,00691641	0,00357690	1,9336	0,0535
sibs	-0,00530847	0,00770614	-0,6889	0,4911
female	-0,246300	0,0450996	-5,4612	0,0000

Modell 2: OLS estimates using 852 observations from 1–935

Fehlende oder unvollständige Beobachtungen entfernt: 83

Abhängige Variable: lwage

Variable	Koeffizient	Std. Fehler	t-Statistik	P-Wert
const	6,07964	0,151970	40,0055	0,0000
age	0,0107521	0,00548169	1,9614	0,0502
exper	0,00290067	0,00356321	0,8141	0,4158
sibs	0,00252821	0,00765325	0,3303	0,7412
female	-0,184569	0,0452963	-4,0747	0,0001
KWW	0,0128536	0,00210057	6,1191	0,0000

[4.] Aufgaben

- Diskutieren und interpretieren Sie die Koeffizienten β_1 und β_4 in Modell 1.
- Man weiß, dass *abil* für die Erklärung des Lohnes relevant ist. Welche Konsequenz könnte dies für die Koeffizientenschätzung in Modell 1 haben?
- Wenn man davon ausgeht, dass der Koeffizient vor *abil* und die Korrelation zwischen *age* und *abil* erwartungsgemäß größer als Null sind, in welche Richtung ist der Schätzer $\hat{\beta}_1$ in Modell 1 verzerrt?
- Im Modell 2 wurde *KWW* als Proxy für *abil* verwendet. Welche Eigenschaften muss *KWW* erfüllen, damit sie sich als Proxy eignet?
- Welches Problem lässt sich mit der Verwendung einer Proxy im Gegensatz zur IV-Schätzung nicht beheben?

5. (12 Punkte)

Gegeben Sei ein in den Koeffizienten lineares Modell mit folgenden Variablen:

- wage: Stundenlohn in €
- black: =1 für Menschen mit schwarzer Hautfarbe
- bigcity: =1 für Menschen, die in einer großen Stadt leben
- smllcity: =1 für Menschen, die in einer kleinen Stadt leben
- exper: Berufserfahrung in Jahren
- expersq: quadrierte Berufserfahrung in Jahren
- educ: Bildung in Jahren

Modell: OLS estimates using the 1260 observations 1–1260

Abhängige Variable: wage

Variable	Koeffizient	Std. Fehler	t-Statistik	P-Wert
const	-2,4483	0,707320	-3,4614	0,0006
black	-0,622070	0,471354	-1,3198	0,1872
bigcity	1,94202	0,344954	5,6298	0,0000
smllcity	0,691616	0,281079	2,4606	0,0140
exper	0,297736	0,0393457	7,5672	0,0000
expersq	-0,00445055	0,000886301	-5,0215	0,0000
educ	0,377585	0,0487693	7,7423	0,0000

Summe der quadrierten Residuen	22965,5
Standardfehler der Residuen ($\hat{\sigma}$)	4,28117
Unkorrigiertes R^2	0,160232
$F(6, 1253)$	39,8464
Log-Likelihood	-3616,6

RESET-Test für Spezifikation –

Nullhypothese: Spezifikation ist angemessen

Teststatistik: $F(2, 1251) = 4,16068$

mit p-Wert = $P(F(2, 1251) > 4,16068) = 0,0158133$

White's Test für Heteroskedastizität –

Teststatistik: $nR^2 = 41,5266$

mit p-Wert = $P(\chi_{22}^2 > 41,5266) = 0,0071291$

- a) Berechnen Sie den marginalen Einfluss der Berufserfahrung! Was ist der Umkehrpunkt und wie würden Sie ihn interpretieren?
- b) Interpretieren Sie das Testergebnis des RESET-Tests!
- c) Betrachten Sie den unter dem Modell gegebenen White-Test:
 - I.) Welches Modell wird für den sogenannten Alternativen White-Test geschätzt?
 - II.) Wie sehen die linearen Restriktionen (formelle Nullhypothese) für den alternativen White-Test aus?
 - III.) Haben wir Homo- oder Heteroskedastizität vorliegen?
 - IV.) Angenommen es ist Heteroskedastizität, ist die Schätzung (in)konsistent?
 - V.) Wie erhalten wir unter Heteroskedastizität einen effizienten Schätzer? Was ist die Grundidee dieses Verfahrens?

6. (10 Punkte)

Betrachtet werden 2 binary response-Modelle mit folgenden Variablen:

- arr06: =1 wenn im Jahr 2006 inhaftiert
- black: =1 bei Menschen mit schwarzer Hautfarbe
- hispan: =1 wenn Lateinamerikaner
- inc06: legales Einkommen im Jahr 2006 in 1000 €
- durat: Dauer von Arbeitslosigkeit in der Vergangenheit (vor 2006) in Monaten

Modell 1: Logit estimates using the 2725 observations 1–2725
Abhängige Variable: arr06

Variable	Koeffizient	Std. Fehler	t-Statistik	Steigung*
const	-0,939793	0,0792563	-11,8576	
black	0,766025	0,115345	6,6412	0,148065
hispan	0,448444	0,107716	4,1632	0,0866799
inc06	-0,00659921	0,000883824	-7,4667	-0,00127556
durat	0,0200870	0,00936375	2,1452	0,00388263

*Evaluated at the mean

Mittelwert von arr06 = 0,277

Number of cases 'correctly predicted' = 1985 (72,8 percent)

$f(\beta'x)$ at mean of independent vars = 0,193

Log-Likelihood = -1526,73

Likelihood-Quotienten-Test: $\chi_4^2 = 162,917$ (P-Wert 0,000000)

Akaike Informations-Kriterium (AIC) = 3063,45

Schwarz' Bayes-Kriterium (BIC) = 3093

Hannan-Quinn-Kriterium (HQQ) = 3074,13

McFaddens Pseudo- $R^2 = 0,0506524$

Modell 2: Logit estimates using the 2725 observations 1–2725
Abhängige Variable: arr06

Variable	Koeffizient	Std. Fehler	t-Statistik	Steigung*
const	-0,665366	0,0660867	-10,0681	
durat	0,0199908	0,00922612	2,1668	0,00389321
inc06	-0,00727792	0,000879843	-8,2718	-0,00141738

*Evaluated at the mean

Mittelwert von $arr06 = 0,277$

Number of cases 'correctly predicted' = 1970 (72,3 percent)

$f(\beta'x)$ at mean of independent vars = 0,195

Log-Likelihood = -1551,48

Likelihood-Quotienten-Test: $\chi^2_2 = 113,412$ (P-Wert 0,000000)

Akaike Informations-Kriterium (AIC) = 3108,95

Schwarz' Bayes-Kriterium (BIC) = 3126,69

Hannan-Quinn-Kriterium (HQQ) = 3115,36

McFaddens Pseudo- $R^2 = 0,035261$

Grundlegende Statistiken, mit Beobachtungen 1–2725

Variable	arith. Mittel	Median	Minimum	Maximum
black	0,161101	0,000000	0,000000	1,00000
hispan	0,217615	0,000000	0,000000	1,00000
inc06	54,9670	29,0000	0,000000	541,000
durat	2,25138	0,000000	0,000000	25,0000

- Verwenden Sie Modell 1, um die Wahrscheinlichkeit einer Inhaftierung (im Jahr 2006) im Mittel der Variablen vorherzusagen!
- Testen Sie die Variablen *black* und *hispan* auf gemeinsame Signifikanz! Wieviele Freiheitsgrade hat die entsprechende Verteilung der Teststatistik? Interpretieren Sie das Ergebnis im Hinblick auf einem Modellvergleich!
- Wie würden Sie allgemein den marginalen Effekt einer erklärenden Variablen berechnen? Berechnen und interpretieren Sie den marginalen Effekt der Variable *inc06* in Modell 2, wenn die erklärenden Variablen den Wert ihres Medians annehmen!

7. (8 Punkte)

- a) Nennen Sie zwei Ansätze zur Trendbestimmung im Rahmen der klassischen Zeitreihenanalyse.
- b) Beschreiben Sie kurz eine Möglichkeit zur Modellierung der Saisonkomponente.
- c) Schreiben Sie die Gleichung für einen AR(1)-Prozess auf. Was passiert, wenn $\rho = 1$ ist?

8. (8 Punkte)

- a) Worin liegt der Unterschied zwischen gepoolten Querschnittsdaten und echten Paneldaten?
- b) Gegeben sei folgendes Paneldatenmodell für zwei Perioden t :

$$y_{it} = \beta_0 + \delta_0 d_t + \beta_1 x_{1,it} + a_i + u_{it}$$

wobei a_i ein unbeobachteter fester Effekt ist.

- I.) Beschreiben Sie kurz die Vorgehensweise beim Differenzenansatz und stellen Sie die entsprechende Gleichung auf.
- II.) Beschreiben Sie kurz einen alternativen Ansatz zum Differenzenansatz.