

Prüfung im Fach Panel- und Evaluationsverfahren im Sommersemester 2015 Aufgaben

Vorbemerkungen:

- Anzahl der Aufgaben:**
- Die Klausur besteht aus 4 Aufgaben, die alle bearbeitet werden müssen.
- Bewertung:**
- Es können maximal 60 Punkte erworben werden. Die Punktzahl ist für jede Aufgabe in Klammern angegeben. Sie entspricht der für die Aufgabe empfohlenen Bearbeitungszeit in Minuten.
- Erlaubte Hilfsmittel:**
- Taschenrechner
 - Fremdwörterbuch
- Wichtige Hinweise:**
- Sollte es vorkommen, dass bei einer Berechnung eine erforderliche Annahme oder Angabe fehlt, machen Sie dies kenntlich und treffen Sie für den fehlenden Wert eine plausible Annahme.

Aufgabe 1 (13,5 Punkte)

Eine Entwicklungsorganisation analysiert die Kindersterblichkeitsrate (y ; gemessen in %) in Abhängigkeit vom Bruttoinlandsprodukt eines Landes (BIP; gemessen in 1000\$). Die folgende Tabelle zeigt den geschätzten Steigungsparameter eines einfachen linearen Modells; die Spalte (1) zeigt die KQ-Ergebnisse, die Spalten (2) bis (6) enthalten die Ergebnisse einer Quantilsregression.

	(1)	(2)	(3)	(4)	(5)	(6)
	KQ	Quantil				
		10%	25%	50%	75%	90%
BIP	-2,29	-2,23	-1,96	-2,40	-2,21	-2,39

- 1.1 Interpretieren Sie die geschätzten Koeffizienten der KQ-Regression und der Regression am 25% Quantil. (3 Punkte)
- 1.2 Nennen Sie zwei Vorteile der Quantilsregression gegenüber einem Kleinstquadrateschätzer. (2 Punkte)
- 1.3 Welches Minimierungsproblem lösen beide in der Tabelle verwendeten Schätzverfahren? Beschreiben Sie verbal und gehen Sie dabei jeweils auf die Gewichtung der Vorhersagefehler ein. (4 Punkte)
- 1.4 Erläutern Sie den Zusammenhang zwischen dem LAD-Schätzer und der Quantilsregression. Stellen Sie formal die Verlustfunktionen des LAD- und des KQ-Schätzers auf. (3 Punkte)
- 1.5 Weisen die Schätzergebnisse der Quantilsregression darauf hin, dass die Störterme der KQ-Schätzung heteroskedastisch sein könnten? Begründen Sie. (1,5 Punkte)

Aufgabe 2 (16,5 Punkte)

Mit Paneldaten von 400 Unternehmen, die jährlich von 2010 bis einschließlich 2015 beobachtet wurden, wird der Zusammenhang zwischen der Absatzmenge (y_i) und den eingesetzten Ausgaben für Marketingmaßnahmen geschätzt (x_i).

- 2.1 Insgesamt enthält der Datensatz 2084 Beobachtungen. Handelt es sich um ein Balanced Panel? Begründen Sie Ihre Aussage. (2 Punkte)
- 2.2 Erläutern Sie knapp das Vorgehen bei einer *least-squares-dummy-variables* (LSDV) Schätzung. Wie viele Parameter würden im vorliegenden Fall insgesamt geschätzt? (2 Punkte)
- 2.3 Erläutern Sie verbal die Vorgehensweise der Within-Schätzung in der vorliegenden Aufgabenstellung. (3 Punkte)
- 2.4 Prüfen Sie mit einem Hausman-Test am 10%-Signifikanzniveau, ob ein Fixed-Effects- oder ein Random-Effects-Modell für die Schätzung der Absatzmenge verwendet werden sollte. (Außer den Ausgaben auf Marketingmaßnahmen stehen keine weiteren erklärenden Variablen zur Verfügung.) Gehen Sie von einem empirischen Wert der Teststatistik von 3,58 aus. Geben Sie die Null- und Alternativhypothese, die Anzahl der Freiheitsgrade sowie den kritischen Wert an. Treffen und erklären Sie die Testentscheidung. (4 Punkte)
- 2.5 Diskutieren Sie für den vorliegenden Fall die Annahmen, die für Konsistenz des Random-Effects-Schätzers notwendig sind. (3 Punkte)
- 2.6 Im Folgenden wird eine dynamische Modellierung erster Ordnung der Absatzmenge zugrunde gelegt. Nehmen Sie die dafür benötigte Within-Transformation der Daten vor. Erklären Sie, welches Problem bei einer Schätzung dieses Modells besteht. (2,5 Punkte)

Aufgabe 3 (15 Punkte)

Im Land Teutonia wird empfohlen, dass Kinder, die vor dem 31.12. eines Jahres 3 Jahre alt werden, den Kindergarten am 01.01. des nächsten Jahres beginnen. Sie interessieren sich für den Effekt des Alters bei Kindergarteneintritt auf den späteren Schulerfolg einer Person. Sie stellen zu diesem Zweck folgendes Modell auf:

$$abitur_i = \alpha + \beta kgalter_i + \varepsilon_i,$$

wobei Ihnen folgende Daten aus dem Land Teutonia zur Verfügung stehen:

- abitur* =1, falls Person Abitur hat; =0 sonst.
- kgalter* Alter bei Kindergarteneintritt (in Jahren).
- monat* Geburtsmonat der Person.
- spacet* =1, falls Person zwischen Juli und Dezember geboren ist; =0 sonst.

- 3.1 Welche für die Konsistenz des KQ-Schätzers benötigte Annahme könnte in diesem Modell verletzt sein? Erläutern Sie die Annahme kurz an einem Beispiel. (3 Punkte)
- 3.2 Nennen und skizzieren Sie kurz verbal ein Verfahren (außer two-stage-least-squares), das in dieser Situation geeignet sein könnte, um den kausalen Effekt des Alters bei Kindergarteneintritt auf den späteren Schulerfolg der Kinder zu identifizieren. (3 Punkte)
- 3.3 Sie verwenden ein 2-stage-least-squares (2SLS) Verfahren und instrumentieren die Variable *kgalter* mit der binären Variable *spacet*: $kgalter_i = \theta + \gamma spacet_i + v_i$

Stata liefert folgende Regressionsergebnisse auf der ersten Stufe:

Source	SS	df	MS			
Model	72.1222404	1	72.1222404	Number of obs =	3185	
Residual	1690.2859	3183	.531035469	F(1, 3183) =	135.81	
Total	1762.40814	3184	.553520144	Prob > F =	0.0000	
				R-squared =	0.0409	
				Adj R-squared =	0.0406	
				Root MSE =	.72872	

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
spacet	-.3009611	.0258248	-11.65	0.000	-.3515961	-.2503261
_cons	3.514357	.0182466	192.60	0.000	3.478581	3.550134

Stata liefert folgende Regressionsergebnisse auf der zweiten Stufe:

Source	SS	df	MS			
Model	2.3302865	1	2.3302865	Number of obs =	3185	
Residual	778.37615	3183	.244541674	F(1, 3183) =	9.53	
Total	780.706436	3184	.245196745	Prob > F =	0.0020	
				R-squared =	0.0030	
				Adj R-squared =	0.0027	
				Root MSE =	.49451	

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
kgalter	-.1797505	.0582294	-3.09	0.002	-.2939214	-.0655797
_cons	1.034842	.196086	5.28	0.000	.6503747	1.41931

- 3.3.1 Interpretieren Sie den Koeffizienten für *spacet* auf der ersten Stufe inhaltlich. (1 Punkt)

- 3.3.2 Interpretieren Sie den geschätzten Koeffizienten für die erklärende Variable $kgalter$ inhaltlich und statistisch. (2 Punkte)
- 3.4 Welchen Effekt identifizieren Sie in diesem Fall mit dem 2SLS Verfahren? Für welche Gruppe ist dieser Effekt gültig? (2 Punkte)
- 3.5.1 Welche Bedingungen müssen erfüllt sein, damit die Instrumentalvariable als geeignet betrachtet werden kann? (2 Punkte)
- 3.5.2 Erläutern Sie knapp die Konsequenzen eines schwachen Instruments für die 2SLS Schätzung. Liegt in diesem Fall das Problem eines schwachen Instrument vor? Begründen Sie. (2 Punkte)

Aufgabe 4 (15 Punkte)

- 4.1 Zeigen Sie formal, was ein Average Treatment Effect (ATE) und ein Average Treatment Effect on the Treated (ATT) ist. Verwenden Sie die Notation des Potential Outcome Frameworks. (Dieser Ansatz beschreibt den Treatmentindikator mit D_i , die Outcomevariable mit Treatment mit $y_i(1)$ und die Outcomevariable ohne Treatment mit $y_i(0)$.) (3 Punkte)
- 4.2 Leiten Sie die Selektionsverzerrung bei der Schätzung des ATTs mittels beobachtbarer Größen formal unter Verwendung des Potential Outcome Frameworks her. Erklären Sie kurz verbal, wann Sie den ATT unverzerrt schätzen können. (3 Punkte)
- 4.3 Sie möchten anhand von Querschnittsdaten überprüfen, ob sich die Teilnahme an einer Weiterqualifizierungsmaßnahme für Arbeitslose positiv auf das Einkommen der Teilnehmer ausgewirkt hat. Insgesamt nahmen 20% aller Arbeitslosen an der Maßnahme teil. Sie führen dazu ein nearest-neighbour-matching mit Zurücklegen durch.
- 4.3.1 Wie wirkt sich das Zurücklegen bei dem Matching Algorithmus auf Verzerrung und Varianz des Schätzers aus? (2 Punkte)
- 4.3.2 Bei Durchführung des Matchings erhalten Sie bezüglich der Mittelwerte der Einkommen folgende Ergebnisse:

	Treatmentgruppe (<i>Teilnahme=1</i>)	Kontrollgruppe (<i>Teilnahme=0</i>)
Unmatched Sample	6.349,15	21.553,9
Matched Sample	6.349,15	4.944,36

Berechnen Sie den ATT und nutzen Sie die Mittelwerte, um eine Aussage zu der Selektion der Teilnehmer zu treffen. Unterscheiden sich Treatment und Kontrollgruppen in ihren potential outcomes ohne Treatment? (3 Punkte)

- 4.3.3 Definieren Sie die SUTVA (stable unit treatment value assumption). Ist diese Annahme in dem vorliegenden Beispiel problematisch? Begründen Sie knapp. (4 Punkte)

Tabelle 3: Perzentile der χ^2 -Verteilung

Zelleneintrag: c, sodass $\text{Prob}[\chi_n^2 \leq c] = P$, mit n Freiheitsgraden

P \ n	0.005	0.01	0.025	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.975	0.99	0.995
1	0.00004	0.0002	0.001	0.004	0.016	0.102	0.455	1.323	2.706	3.842	5.024	6.635	7.879
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	16.34	20.34	24.93	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	19.04	23.34	28.24	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.31	46.93
30	13.79	14.95	16.79	18.49	20.60	24.48	29.34	34.80	40.26	43.77	46.98	50.89	53.67
35	17.19	18.51	20.57	22.47	24.80	29.05	34.34	40.22	46.06	49.80	53.20	57.34	60.27
40	20.71	22.16	24.43	26.51	29.05	33.66	39.34	45.62	51.81	55.76	59.34	63.69	66.77
45	24.31	25.90	28.37	30.61	33.35	38.29	44.34	50.98	57.51	61.66	65.41	69.96	73.17
50	27.99	29.71	32.36	34.76	37.69	42.94	49.33	56.33	63.17	67.50	71.42	76.15	79.49

Quelle: In R generiert