

Bachelorprüfung im Wintersemester 2013/14

Fach: Empirische Wirtschaftsforschung II

Prüfer: Prof. Regina T. Riphahn, Ph.D.

Name, Vorname	
Matrikelnr.	
Studiengang	
Semester	
Datum	
Raum	
Unterschrift	

Vorbemerkungen:

Anzahl der Aufgaben: Die Klausur besteht aus 5 Aufgaben, die alle bearbeitet werden müssen.

Bewertung: Es können maximal 90 Punkte erworben werden. Die Punktzahl ist für jede Aufgabe in Klammern angegeben. Sie entspricht der für die Aufgabe empfohlenen Bearbeitungszeit in Minuten.

Erlaubte Hilfsmittel:

- Tabellen der statistischen Verteilung und Liste der Annahmen (sind der Klausur beigelegt)
- 1 DIN-A4-Seite mit Notizen
- Taschenrechner
- Fremdwörterbuch

Wichtige Hinweise:

- Sollte es vorkommen, dass die statistischen Tabellen, die dieser Klausur beiliegen, den exakten Wert der Freiheitsgrade nicht ausweisen, machen Sie dies kenntlich und verwenden Sie den nächstgelegenen Wert.
- Sollte es vorkommen, dass bei einer Berechnung eine erforderliche Information fehlt, machen Sie dies kenntlich und treffen Sie für den fehlenden Wert eine plausible Annahme.

Aufgabe 1:**[25 Punkte]**

Sie interessieren sich für die Determinanten der Lohnhöhe. Ihr Querschnittsdatensatz enthält folgende Variablen:

lohn = Bruttostundenlohn in €

alter = Alter in Jahren

mann = 1 wenn männlich, 0 wenn weiblich

ausb = Schul- und Hochschulausbildung in Jahren

single = Familienstand: 1, wenn Person alleinstehend ist, sonst 0.

attrak = misst auf einer Skala von 1 - 5, wie attraktiv diese Person von anderen bewertet wird.

demog = misst den Anteil des jeweils eigenen Geschlechts an allen Personen im heiratsfähigen Alter auf Landkreisebene.

Sie schätzen das Modell: $lohn_i = \beta_0 + \beta_1 alter_i + \beta_2 single_i + \beta_3 mann_i + \beta_4 ausb_i + u_i$.

- Was ist Heteroskedastie und welche beiden ungünstigen Auswirkungen hat Heteroskedastie auf KQ-Schätzergebnisse? (3 Punkte)
- Sie führen nach der Schätzung des Modells einen Breusch Pagan Test auf Heteroskedastie durch. Die LM – Teststatistik beträgt 10,0 (p-Wert < 0,01).
 - Beschreiben Sie die Vorgehensweise des durchgeführten Tests am Beispiel. Nennen Sie dabei Null- und Alternativhypothese, definieren Sie die Teststatistik und ihre Bestandteile, nennen Sie die Anzahl der Freiheitsgrade und die Verteilung der Teststatistik. (7 Punkte)
 - Deutet das Testergebnis auf Heteroskedastie hin? (1 Punkt)
- Sie interessieren sich für den kausalen Effekt von *single*. Diskutieren Sie am Beispiel warum *single* endogen sein könnte. (2 Punkte)
- Sie möchten den Effekt von *single* im Rahmen einer Instrumentvariablenschätzung bestimmen. Welche Eigenschaften muss ein Instrument erfüllen? (2 Punkte)
- Diskutieren Sie am Beispiel, unter welchen Bedingungen sich die Variablen *attrak* und *demog* als Instrumentvariablen für *single* eignen. (6 Punkte)
- Erläutern Sie die zwei Stufen der 2SLS Schätzung unter Nutzung der Variablen *attrak* und *demog*. Schreiben Sie dazu alle relevanten Schätzgleichungen auf. (4 Punkte)

Aufgabe 2:**[9 Punkte]**

- Erläutern sie knapp was unter schwacher Abhängigkeit einer Zeitreihe zu verstehen ist. (2 Punkte)
- Prüfen Sie, ob ein MA(1) Prozess der Art $x_t = e_t + 0,99e_{t-1}$ schwach abhängig ist. Gehen Sie dabei davon aus, dass e_t eine unabhängig und identisch verteilte (i.i.d.) Zufallsvariable mit $E[e_t] = 0$ und $Var[e_t] = 1$ ist. (7 Punkte)

Aufgabe 3**[12 Punkte]**

- i) Beschreiben Sie im Detail Ziel und Vorgehen des Cochrane-Orcutt Verfahrens für den Fall, dass für den Störterm e_t im einfachen linearen Regressionsmodell mit exogenem Regressor gilt: $e_t = \rho e_{t-1} + v_t$. Dabei entspricht v_t den Gauss Markov Annahmen und der Wert von $|\rho| < 1$ sei bekannt. (4 Punkte)

- ii) Ist der Schätzer BLUE? Begründen Sie. (1 Punkt)
- b) Beeinflusst das Vorliegen eines stabilen AR(1) Prozesses im Störterm für eine KQ Schätzung des Modells $y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_t + e_t$ die Konsistenzeigenschaft des Schätzers? Begründen Sie! (2 Punkte)
- c) Beschreiben Sie ein Verfahren mit dem Sie im Modell $y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_t + e_t$ das Vorliegen eines AR(1) Prozesses im Störterm testen können. Nennen Sie Null- und Alternativhypothese, Teststatistik und Schlusslogik. (5 Punkte)

Aufgabe 4:

[14 Punkte]

Sie möchten prüfen, ob es Lohnunterschiede zwischen dem öffentlichen Sektor und dem Privatsektor gibt. Ihnen liegen dazu Paneldaten mit 42199 Beobachtungen von 7842 arbeitenden Personen vor. Folgendes Modell soll zunächst mittels KQ geschätzt werden:

$$\log w_{it} = \beta_0 + \beta_1 \text{öffDie}_{it} + \beta_2 \text{mann}_{it} + \beta_3 \text{erf}_{it} + u_{it}$$

mit $u_{it} = a_i + \varepsilon_{it}$, i als Personenindikator und t als Zeitindikator.

$\log w$ = logarithmierter Bruttostundenlohn gemessen in €

öffDie = 1, wenn Person i zum Zeitpunkt t im öffentlichen Dienst beschäftigt ist, sonst 0.

mann = 1, wenn Person männlich ist, sonst 0.

erf = Berufserfahrung der Person gemessen in 10 Jahren

a, ε = unbeobachtete Faktoren (Störterm)

Eine Schätzung des Modells mit verschiedenen Schätzverfahren liefert folgende Ergebnisse (Standardfehler in Klammern):

	KQ	Fixed Effects
öffDie	0.145 (0.005)	0.014 (0.006)
mann	0.246 (0.004)	(omitted)
erf	0.057 (0.002)	0.375 (0.004)
cons	2.145 (0.005)	1.759 (0.074)

- a) Sie überlegen, statt KQ einen Random Effects Schätzer zu verwenden. Erklären Sie, warum dieser bei Paneldaten gegenüber KQ einen Vorteil hat und worin der Vorteil besteht. (3 Punkte)
- b) Erklären Sie, weshalb β_j im vorliegenden Beispiel mit KQ bei Nichtberücksichtigung der Panelstruktur inkonsistent geschätzt werden könnte. (2 Punkte)
- c) Schreiben Sie für das gegebene Beispiel die Schätzgleichung für einen Within-Schätzer auf. Erklären Sie, inwieweit Sie mit Hilfe dieses Schätzansatzes das in Teilaufgabe b) angesprochene Problem lösen können. (3 Punkte)
- d) Interpretieren Sie den $\widehat{\beta}_1$ der Fixed Effects Schätzung inhaltlich und statistisch. (3 Punkte)
- e) Wie lässt sich der zwischen KQ und Fixed Effects Schätzung stark unterschiedliche Schätzwert für β_1 erklären? Geben Sie zwei mögliche Erklärungen an. (3 Punkte)

Aufgabe 5:**[30 Punkte]**

Welche Antwort ist richtig? Kreuzen Sie nur **eine Antwort** pro Aufgabe an. Falls mehrere Aussagen korrekt sind, kreuzen Sie **nur** die entsprechende **Antwortkombination** an. Für jede richtige Antwort gibt es 1 Punkt. Für falsche Antworten werden keine Punkte abgezogen.

1.	Ist die zeitkonstante, unbeobachtete Heterogenität a_i unkorreliert mit der erklärenden Variablen x_{it} , d.h. $Cov(a_i, x_{it}) = 0$, dann	
a	<input type="checkbox"/>	ergeben der KQ Schätzer und der Between Schätzer identische Ergebnisse.
b	<input type="checkbox"/>	ergeben der Fixed Effects Schätzer und der Between Schätzer identische Ergebnisse.
c	<input type="checkbox"/>	kann der Effekt zeitkonstanter Merkmale konsistent geschätzt werden.
d	<input type="checkbox"/>	ist der Instrumentvariablenschätzer effizient.

2.	Der LSDV Schätzer	
a	<input type="checkbox"/>	ist BLUE, falls Paneldaten verwendet werden.
b	<input type="checkbox"/>	ist inkonsistent, wenn ein Regressor mit dem zeitinvarianten Teil des Störterms korreliert.
c	<input type="checkbox"/>	ist inkonsistent, wenn ein y mit dem zeitvarianten Teil des Störterms korreliert.
d	<input type="checkbox"/>	Keine der genannten Antworten.

3.	Attenuation bias	
a	<input type="checkbox"/>	ist eine Folge von Heteroskedastie.
b	<input type="checkbox"/>	kann mit Hilfe von Fixed Effects Schätzern verhindert werden.
c	<input type="checkbox"/>	ist eine Folge von Autokorrelation.
d	<input type="checkbox"/>	kann bei Messfehlern in der erklärenden Variable auftreten.

4.	Im linearen Wahrscheinlichkeitsmodell	
a	<input type="checkbox"/>	ist die abhängige Variable binär kodiert.
b	<input type="checkbox"/>	tritt zwangsläufig Autokorrelation auf.
c	<input type="checkbox"/>	können vorhergesagte Werte für x außerhalb des gültigen Wertebereichs liegen.
d	<input type="checkbox"/>	können bei einer FGLS Schätzung zur Behebung von Heteroskedastie nie alle Beobachtungen verwendet werden.

5.	Ist MLR.1 bis MLR.5 (siehe Anhang) erfüllt, so ist	
a	<input type="checkbox"/>	KQ normalverteilt.
b	<input type="checkbox"/>	sind KQ und LAD äquivalent.
c	<input type="checkbox"/>	ist KQ konsistent.
d	<input type="checkbox"/>	Alle genannten Antworten.

6.	Das LAD Verfahren	
a	<input type="checkbox"/>	maximiert die Summe der Absolutwerte der Residuen.
b	<input type="checkbox"/>	minimiert die Differenz der Absolutwerte der Residuen.
c	<input type="checkbox"/>	minimiert die Summe der quadrierten Absolutwerte der Residuen.
d	<input type="checkbox"/>	modelliert den auf x bedingten Median von y.

7.	Im Fall seriell korrelierter Störterme	
a	<input type="checkbox"/>	kann $\hat{\beta}$ fälschlicherweise als insignifikant ausgewiesen werden.
b	<input type="checkbox"/>	ist das R^2 falsch berechnet.
c	<input type="checkbox"/>	ist $\hat{\beta}$ verzerrt.
d	<input type="checkbox"/>	b und c.

8.	Bei einer in zwei Stufen durchgeführten 2SLS Schätzung muss stets	
a	<input type="checkbox"/>	Heteroskedastie bereinigt werden.
b	<input type="checkbox"/>	die Schätzung der Standardfehler angepasst werden.
c	<input type="checkbox"/>	für Saisonalität kontrolliert werden.
d	<input type="checkbox"/>	Autokorrelation bereinigt werden
9.	Der Hausman Test verwirft die Nullhypothese, wenn	
a	<input type="checkbox"/>	die Störterme heteroskedastisch sind.
b	<input type="checkbox"/>	die Störterme einem AR(1) Prozess folgen.
c	<input type="checkbox"/>	Random Effects- und Within Schätzer unterschiedliche Koeffizienten ergeben.
d	<input type="checkbox"/>	die Störterme nicht normalverteilt sind.
10.	Lassen sich in Querschnittsdaten von Schülern Familien identifizieren, erlaubt dies	
a	<input type="checkbox"/>	die Schätzung eines Finite Distributed Lag Modells.
b	<input type="checkbox"/>	die Anwendung des Fixed Effects Schätzers.
c	<input type="checkbox"/>	die Anwendung des Prais-Winsten Schätzers.
d	<input type="checkbox"/>	Alle der genannten Antworten.
11.	Schwache Abhängigkeit	
a	<input type="checkbox"/>	gilt für stabile AR Prozesse.
b	<input type="checkbox"/>	führt zu ungenauen IV Schätzungen.
c	<input type="checkbox"/>	gilt für $u_t = u_{t-1} + e_t$, wobei $e_t \sim i. i. d. (0, \sigma^2)$.
d	<input type="checkbox"/>	Keine der genannten Antworten.
12.	Beim Davidson-McKinnon Test	
a	<input type="checkbox"/>	wird auch auf Autokorrelation höherer Ordnung getestet.
b	<input type="checkbox"/>	werden genestete Modelle verglichen.
c	<input type="checkbox"/>	ist die abhängige Variable für beide Modelle gleich.
d	<input type="checkbox"/>	Keine der genannten Antworten.
13.	Eine Difference-in-Difference Schätzung	
a	<input type="checkbox"/>	erfordert unterschiedliche Trends in der abhängigen Variable.
b	<input type="checkbox"/>	kann nicht zur Evaluation von Politikmaßnahmen verwendet werden.
c	<input type="checkbox"/>	ist inkonsistent, wenn sich Treatment- und Kontrollgruppe vor dem Treatment in beobachtbaren Merkmalen unterscheiden.
d	<input type="checkbox"/>	ist mit gepoolten Querschnittsdaten möglich.
14.	Ist die abhängige Variable integriert mit Ordnung eins	
a	<input type="checkbox"/>	ist sie schwach abhängig.
b	<input type="checkbox"/>	ist ihre erste Differenz schwach abhängig.
c	<input type="checkbox"/>	folgt sie einem stabilen AR(1) Prozess.
d	<input type="checkbox"/>	muss der Breusch-Godfrey Test durchgeführt werden.
15.	Bei einem balanced Panel	
a	<input type="checkbox"/>	werden ausscheidende Merkmalsträger sofort durch neue ersetzt.
b	<input type="checkbox"/>	entspricht die Anzahl der Beobachtungen der Anzahl der Befragungsjahre.
c	<input type="checkbox"/>	ist die Beobachtungszahl ein ganzzahliges Vielfaches der Anzahl der Merkmalsträger.
d	<input type="checkbox"/>	a und b.

16.	Finite Distributed Lag Modelle	
a	<input type="checkbox"/>	sind dynamisch vollständig.
b	<input type="checkbox"/>	sind statisch.
c	<input type="checkbox"/>	können verzögerte Werte der abhängigen Variable als Regressor beinhalten.
d	<input type="checkbox"/>	Keine der genannten Antworten.

17.	Der Fixed Effects Schätzer für $y_{it} = \beta_0 + \beta_1 x_{it} + \alpha_i + u_{it}$	
a	<input type="checkbox"/>	ist effizient wenn der zeitinvariante Teil des Störterms unabhängig von x_{it} ist.
b	<input type="checkbox"/>	erfordert serielle Korrelation in den zeitvarianten Variablen.
c	<input type="checkbox"/>	liefert umso präzisere Schätzungen je höher die Within-Variation der Regressoren x_{it} ist.
d	<input type="checkbox"/>	Keine der genannten Antworten.

18.	Kovarianzstationarität eines Prozesses x_t impliziert, dass	
a	<input type="checkbox"/>	die Varianz von x_t für alle Zeitpunkte gleich ist.
b	<input type="checkbox"/>	x_t strikt stationär ist.
c	<input type="checkbox"/>	x_t schwach abhängig ist.
d	<input type="checkbox"/>	b und c.

19.	Folgt der Störterm in einem Zeitreihenmodell einem Random Walk mit Drift	
a	<input type="checkbox"/>	ist der Störterm I(2).
b	<input type="checkbox"/>	Handelt es sich um einen stabilen AR(1) Prozess.
c	<input type="checkbox"/>	folgt der Störterm einem langfristigen Trend.
d	<input type="checkbox"/>	a und b.

20.	Der White Test	
a	<input type="checkbox"/>	basiert auf einer Regression der abhängigen Variable auf ein Polynom der vorhergesagten abhängigen Variable.
b	<input type="checkbox"/>	testet auf Autokorrelation 1. Ordnung.
c	<input type="checkbox"/>	vergleicht die Spezifikation nicht genesteter Modelle.
d	<input type="checkbox"/>	ist allgemeiner als der Breusch-Pagan Test.

21.	Endogene Stichprobenselektion	
a	<input type="checkbox"/>	bedeutet systematische Stichprobenselektion auf Basis exogener Variablen.
b	<input type="checkbox"/>	kann zu verzerrten Parameterschätzern führen.
c	<input type="checkbox"/>	ist asymptotisch unproblematisch.
d	<input type="checkbox"/>	erhöht die Effizienz der Schätzung.

22.	Instrumentvariablen	
a	<input type="checkbox"/>	können bei Modellen mit Interaktionsterm nicht verwendet werden.
b	<input type="checkbox"/>	dürfen nicht mit den zu instrumentierenden Regressoren korreliert sein.
c	<input type="checkbox"/>	dürfen nicht mit den exogenen Regressoren korreliert sein.
d	<input type="checkbox"/>	dürfen nicht mit dem Störterm korreliert sein.

23.	Proxyvariablen müssen	
a	<input type="checkbox"/>	dieselben Eigenschaften wie Instrumentvariablen aufweisen.
b	<input type="checkbox"/>	stetig sein.
c	<input type="checkbox"/>	mit der ausgelassen Variable korreliert sein.
d	<input type="checkbox"/>	mit den Regressoren korreliert sein.

24.	Die Annahme TS.5: $\text{corr}(u_t, u_s X) = 0$ mit $t \neq s$,	
-----	--	--

a	<input type="checkbox"/>	ist im Zeitreihenfall erforderlich, da die Annahme der Zufallsstichprobe nicht gilt.
b	<input type="checkbox"/>	ist im Querschnittsfall erforderlich, da die Annahme der Zufallsstichprobe nicht gilt.
c	<input type="checkbox"/>	ist für eine konsistente Schätzung der Parameter erforderlich.
d	<input type="checkbox"/>	Alle der genannten Antworten.

25.	Ist die zeitkonstante, unbeobachtete Heterogenität a_i unkorreliert mit der erklärenden Variablen x_{it} , d.h. $Cov(a_i, x_{it}) = 0$, dann	
a	<input type="checkbox"/>	ist der KQ-Schätzer effizient.
b	<input type="checkbox"/>	ist der Random Effects Schätzer effizienter als der Fixed Effects Schätzer.
c	<input type="checkbox"/>	ist der KQ Schätzer inkonsistent.
d	<input type="checkbox"/>	ist der Random Effects Schätzer inkonsistent.

26.	Eine gepoolte KQ Schätzung für Paneldaten mit Störterm $u_{it} = a_i + e_{it}$	
a	<input type="checkbox"/>	ist unverzerrt, wenn die erklärenden Variablen und a_i korreliert sind.
b	<input type="checkbox"/>	und der Between Schätzer liefern identische Schätzergebnisse.
c	<input type="checkbox"/>	ist effizienter als eine Between Schätzung.
d	<input type="checkbox"/>	a und b.

27.	Periodendummies in Fixed Effects Schätzungen	
a	<input type="checkbox"/>	kontrollieren für Zeittrends in der abhängigen Variable.
b	<input type="checkbox"/>	haben keine Within-Variation.
c	<input type="checkbox"/>	führen nur bei einem Balanced Panel zu Autokorrelation im Störterm.
d	<input type="checkbox"/>	erlauben die Schätzung individuenspezifischer Achsenabschnitte.

28.	Der Vorteil des Durbin-Watson Tests besteht in	
a	<input type="checkbox"/>	seiner Fähigkeit auf Autokorrelation höherer Ordnung testen zu.
b	<input type="checkbox"/>	seiner exakten Gültigkeit auch in kleinen Stichproben.
c	<input type="checkbox"/>	seiner exakten Gültigkeit auch bei endogenen Regressoren.
d	<input type="checkbox"/>	a und c.

29.	Als Impact Multiplier bezeichnet man	
a	<input type="checkbox"/>	Interaktionsterme mit der Konstanten.
b	<input type="checkbox"/>	das Vorgehen zur Bereinigung multiplikativer Heteroskedastie.
c	<input type="checkbox"/>	die Summe aller Koeffizienten in einem Finite Distributed Lag Modell.
d	<input type="checkbox"/>	den Effekt des kontemporären Regressors in einem Finite Distributed Lag Modell.

30.	Die Treatmentgruppe sollte	
a	<input type="checkbox"/>	aus weniger Personen als die Kontrollgruppe bestehen.
b	<input type="checkbox"/>	genauso groß wie die Kontrollgruppe sein.
c	<input type="checkbox"/>	sich nicht systematisch in unbeobachteten Merkmalen von der Kontrollgruppe unterscheiden.
d	<input type="checkbox"/>	a und b.

Annahmen im linearen Regressionsmodell

Einfaches Modell

SLR.1 Das Bevölkerungsmodell ist linear in den Parametern

SLR.2 Die Stichprobe ist zufällig

SLR.3 Die Realisationen von x_i in der Stichprobe sind nicht alle identisch

SLR.4 $E(u|x) = 0$

SLR.5 $\text{Var}(u|x) = \sigma^2$

Multiples Modell

MLR.1 Das Bevölkerungsmodell ist linear in den Parametern

MLR.2 Die Stichprobe ist zufällig

MLR.3 Keine perfekte Kollinearität

MLR.4 $E(u|x_1, \dots, x_k) = E(u) = 0$

MLR.5 $\text{Var}(u|x_1, \dots, x_k) = \text{Var}(u) = \sigma^2$

MLR.6 $u \sim \text{Normal}(0, \sigma^2)$

Zeitreihenmodelle

TS.1 Der stochastische Prozess ist linear in den Parametern

TS.2 Keine perfekte Kollinearität

TS.3 $E(u_t | X) = 0, t=1, 2, \dots, n$

TS.4 $\text{Var}(u_t | X) = \text{Var}(u_t) = \sigma^2, t = 1, 2, \dots, n$

TS.5 $\text{corr}(u_t, u_s | X) = 0, t \neq s$

TS.6 $u_t \sim N(0, \sigma^2)$

TS.1' Der stochastische Prozess ist stationär, schwach abhängig und linear in den Parametern

TS.2' Keine perfekte Kollinearität

TS.3' $E(u_t | x_t) = 0$

TS.4' $\text{Var}(u_t | x_t) = \sigma^2$

TS.5' $E(u_s u_t | x_t, x_s) = 0$ für alle $t \neq s$.