

Bachelorprüfung im Sommersemester 2013

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:**[21 Punkte]**

Sie interessieren sich für die Determinanten der Lohnhöhe. Sie verfügen über Umfragedaten zu 1010 Vollzeit arbeitenden westdeutschen Männern aus dem Jahre 2005. Ihr Datensatz enthält folgende Variablen:

- $\log\text{lohn}$ = logarithmierter Bruttostundenlohn
 $y\text{edu}$ = Schul- und Hochschulausbildung in Jahren
 $w\text{orkexp}$ = Arbeitserfahrung in Jahren
 $s\text{enior}$ = Betriebszugehörigkeitsdauer in Jahren

Sie möchten mit folgendem Modell die Bildungsrendite mit KQ schätzen:

$$\log\text{lohn}_i = \beta_0 + \beta_1 y\text{edu}_i + \beta_2 w\text{orkexp}_i + \beta_3 w\text{orkexp}_i^2 + \beta_4 s\text{enior}_i + \beta_5 s\text{enior}_i^2 + u_i,$$

mit i als Personenindikator.

- Vor der Schätzung möchten Sie testen, ob die funktionale Form des Modells korrekt spezifiziert ist. Beschreiben Sie ein geeignetes Testverfahren. (4 Punkte)
- 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 ist 12,37 (p-Wert 0,0004).
 - Beschreiben Sie die Vorgehensweise des durchgeführten Tests am Beispiel. Nennen Sie dabei Null- und Alternativhypothese, definieren Sie die Teststatistik und nennen Sie deren Freiheitsgrade und Verteilung. (7 Punkte)
 - Deutet das Testergebnis auf Heteroskedastie hin? (1 Punkt)
- Nennen Sie zwei Verfahren, mit Heteroskedastie umzugehen. Welche der in b) angesprochenen Probleme werden durch die beiden Verfahren jeweils gelöst, welche nicht? (6 Punkte)

Aufgabe 2:**[11 Punkte]**

- Definieren Sie Endogenität und erklären Sie unter Angabe eines Beispiels, warum das Bildungsniveau im Modell aus Aufgabe (1) eine endogene Variable sein könnte. (3 Punkte)
- Sie möchten die Bildungsrendite im Rahmen einer Instrumentvariablenschätzung bestimmen. Nennen Sie die für eine konsistente Schätzung erforderliche Momentenbedingung und schreiben Sie die Formel für den IV Schätzer für β_1 auf. (3 Punkte)
- Ihnen liegt als zusätzliche Variable die Bildung des Vaters vor und Sie erwägen, diese Variable als Instrument für die Bildung zu benutzen.
 - Welche Eigenschaften muss eine Variable erfüllen, um als Instrumentvariable geeignet zu sein? (2 Punkte)
 - Diskutieren Sie die Eignung der Instrumentvariable „Bildung des Vaters“ am Beispiel der Aufgabenstellung. (3 Punkte)

Aufgabe 3:**[12 Punkte]**

Anstatt eines Querschnitts wie in Aufgabe (1) verwenden Sie nun Paneldaten zu den gleichen Variablen für Ihre Analyse. Ihr Modell lautet nun:

$$\log\text{lohn}_{it} = \beta_0 + \beta_1 \text{yedu}_{it} + \beta_2 \text{workexp}_{it} + \beta_3 \text{workexpsq}_{it} + \beta_4 \text{senior}_{it} + \beta_5 \text{seniorsq}_{it} + u_{it}$$

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

- Schreiben Sie für das gegebene Beispiel die Schätzgleichung für einen Within-Schätzer sowie für einen First-Difference-Schätzer auf. Unter welchen Umständen können Sie mit Hilfe dieser Schätzansätze Koeffizienten konsistent schätzen? (5 Punkte)
- Erläutern Sie den Vorteil und zwei Nachteile des Within-Schätzers gegenüber dem Random-Effects Schätzer. (4 Punkte)
- Welche Probleme ergeben sich bei der Schätzung der Bildungsrendite mit dem Within-Schätzer, wenn Bildung für Erwerbstätige wenig zeitvariant ist? Diskutieren Sie Konsequenzen. (3 Punkte)

Aufgabe 4:**[15 Punkte]**

Sie analysieren das Bruttoinlandsprodukt Deutschlands mit Hilfe von Quartalsdaten über mehrere Jahre und benutzen folgenden Variablen:

$\log\text{GDP}$ = logarithmierte Bruttowertschöpfung

$\log N$ = logarithmierte Anzahl von Beschäftigten

$\log K$ = logarithmierter Kapitalstock

trend = fortlaufender Zeitindikator (1. Quartal $\text{trend}=1$... letztes Quartal $\text{trend}=T$).

Sie schätzen eine Cobb-Douglas Produktionsfunktion:

$$\log\text{GDP}_t = \beta_0 + \beta_1 \log N_t + \beta_2 \log K_t + \beta_3 \text{trend}_t + u_t$$

mit $t = 1, \dots, T$ als Zeitindikator und erhalten für $\hat{\beta}_3$ Wert 0,005.

- Interpretieren Sie $\hat{\beta}_3$ und fertigen Sie eine Skizze des Zusammenhangs zwischen GDP_t und t an (Achsenbeschriftung nicht vergessen). (3 Punkte)
- Sie identifizieren saisonale Muster in der Bruttowertschöpfung und vermuten, dass der Einfluss der Zeit auf die Bruttowertschöpfung in Ihrem Modell noch nicht ausreichend spezifiziert ist. Erläutern Sie zwei Lösungsvorschläge. (4 Punkte)
- Nach der Berücksichtigung saisonaler Muster vermuten Sie einen AR(1) Prozess im Störterm u_t .
 - Beschreiben Sie einen stabilen AR(1) Prozess formal. (2 Punkte)
 - Welche beiden ungünstigen Auswirkungen hat ein AR(1) Prozess im Störterm auf KQ-Schätzergebnisse? (2 Punkte)
- Ein Durbin-Watson Test ergibt eine Teststatistik von 1,5.
 - Welche Testentscheidung treffen Sie, wenn der untere kritische Wert (d_L) 1,57 beträgt? Würde es sich um einen positiven oder negativen AR(1) Prozess handeln? (2 Punkte)
 - Wie würden Sie das Testergebnis bewerten, wenn Sie zusätzlich von einer Korrelation zwischen u_t und $\log K_{t+1}$ ausgehen? (2 Punkte)

Aufgabe 5:

[31 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 konsistent.

2.	Dynamische Vollständigkeit eines Zeitreihenmodells führt zu Problemen,	
a	<input type="checkbox"/>	wenn die verzögerte abhängige Variable als Regressor enthalten ist.
b	<input type="checkbox"/>	wenn Heteroskedastie auftritt.
c	<input type="checkbox"/>	schwache Abhängigkeit vorliegt.
d	<input type="checkbox"/>	Keine der genannten Antworten.

3.	Endogene Stichprobenselektion	
a	<input type="checkbox"/>	bedeutet systematische Stichprobenselektion auf Basis endogener Variablen.
b	<input type="checkbox"/>	erhöht die Stichprobengröße.
c	<input type="checkbox"/>	ist asymptotisch unproblematisch.
d	<input type="checkbox"/>	erhöht die Effizienz der Schätzung.

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

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

6.	Der Between Schätzer für Paneldaten mit Störterm $u_{it} = a_i + e_{it}$, $t = 1, \dots, T$.	
a	<input type="checkbox"/>	ist eine Querschnittsschätzung (d.h. $T = 1$) über erste Differenzen.
b	<input type="checkbox"/>	ist eine Querschnittsschätzung (d.h. $T = 1$) über individuenspezifische Mittelwerte.
c	<input type="checkbox"/>	ist eine Längsschnittsschätzung (d.h. $T > 1$) über erste Differenzen.
d	<input type="checkbox"/>	ist eine Längsschnittsschätzung (d.h. $T > 1$) über individuenspezifische Mittelwerte.

7.	Ist MLR.1 bis MLR.5 (siehe Anhang) erfüllt, so ist	
a	<input type="checkbox"/>	KQ effizient.
b	<input type="checkbox"/>	ist KQ erwartungstreu.
c	<input type="checkbox"/>	ist KQ konsistent.
d	<input type="checkbox"/>	Alle genannten Antworten.

8.	Das LAD Verfahren	
a	<input type="checkbox"/>	maximiert die Summe der Absolutwerte der Residuen.
b	<input type="checkbox"/>	minimiert die Summe der Absolutwerte der Residuen.
c	<input type="checkbox"/>	minimiert die Summe der quadrierten Absolutwerte der Residuen.
d	<input type="checkbox"/>	a und b.
9.	Im Fall seriell korrelierter Störterme	
a	<input type="checkbox"/>	kann $\hat{\beta}$ fälschlicherweise als insignifikant ausgewiesen werden.
b	<input type="checkbox"/>	kann gleichzeitig Heteroskedastie vorliegen.
c	<input type="checkbox"/>	ist $\hat{\beta}$ verzerrt.
d	<input type="checkbox"/>	a und b.
10.	Der Prozess $u_t = e_t - 0,99e_{t-1}$ mit $e_t \sim i. i. d. (0, \sigma^2)$	
a	<input type="checkbox"/>	ist schwach abhängig.
b	<input type="checkbox"/>	definiert einen AR(1) Prozess.
c	<input type="checkbox"/>	ergibt $\text{cov}(u_t; u_{t-1}) = 0$.
d	<input type="checkbox"/>	Alle genannten Antworten.
11.	Gepoolte Querschnittsdaten können	
a	<input type="checkbox"/>	Aussagen über die Änderungen marginaler Effekte über die Zeit ermöglichen.
b	<input type="checkbox"/>	die Anwendung eines Between Schätzers ermöglichen.
c	<input type="checkbox"/>	die Anwendung eines First Difference Schätzers ermöglichen.
d	<input type="checkbox"/>	die Anwendung eines Random Effects Schätzers ermöglichen.
12.	Bei einer in zwei Stufen durchgeführten 2SLS Schätzung muss	
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
13.	Der Hausman Test verwirft die Nullhypothese, wenn	
a	<input type="checkbox"/>	LSDV und Within Schätzer unterschiedliche Koeffizienten ergeben.
b	<input type="checkbox"/>	Random Effects- und Within Schätzer ähnliche Koeffizienten ergeben.
c	<input type="checkbox"/>	Random Effects- und Within Schätzer unterschiedliche Koeffizienten ergeben.
d	<input type="checkbox"/>	a und c.
14.	Lassen sich in einem Querschnittsdatensatz Geschwister identifizieren, erlaubt dies	
a	<input type="checkbox"/>	die Anwendung des Random Effects Schätzers.
b	<input type="checkbox"/>	die Anwendung des Fixed Effects Schätzers.
c	<input type="checkbox"/>	die Anwendung des Instrumentvariablenschätzers.
d	<input type="checkbox"/>	Alle der genannten Antworten.
15.	Schwache Abhängigkeit	
a	<input type="checkbox"/>	gilt für stabile AR Prozesse.
b	<input type="checkbox"/>	impliziert Stationarität.
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.

16.	Sei $y_i = \beta x_i + u_i$. Das Auslassen der Variable z_i führt stets zu einem verzerrten $\hat{\beta}$, falls	
a	<input type="checkbox"/>	z_i mit y_i korreliert ist.
b	<input type="checkbox"/>	z_i mit x_i korreliert ist.
c	<input type="checkbox"/>	z_i mit u_i und x_i korreliert ist.
d	<input type="checkbox"/>	a und c.

17.	Beim Davidson-McKinnon Test	
a	<input type="checkbox"/>	wird auch auf Autokorrelation höherer Ordnung getestet.
b	<input type="checkbox"/>	werden nicht genestete Modelle verglichen.
c	<input type="checkbox"/>	ist die abhängige Variable für beide Modelle unterschiedlich.
d	<input type="checkbox"/>	Keine der genannten Antworten.

18.	Eine Difference-in-Difference Schätzung	
a	<input type="checkbox"/>	ist nur bei Vorliegen von Paneldaten möglich.
b	<input type="checkbox"/>	kann 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"/>	Keine der genannten Antworten.

19.	Bei einem balanced Panel	
a	<input type="checkbox"/>	werden die gleichen Merkmalsträger wiederholt befragt.
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 c.

20.	Finite Distributed Lag Modelle	
a	<input type="checkbox"/>	können nur mit Regressionskonstante geschätzt werden.
b	<input type="checkbox"/>	können auch auf Datenstrukturen ohne zeitliche Dimension angewendet werden, z.B. Clusterstichproben.
c	<input type="checkbox"/>	sind nur für Balanced Panels konsistent.
d	<input type="checkbox"/>	Keine der genannten Antworten.

21.	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 Regressor mit dem zeitvarianten Teil des Störterms korreliert.
d	<input type="checkbox"/>	Keine der genannten Antworten.

22.	Der Fixed Effects Schätzer	
a	<input type="checkbox"/>	ist ineffizient wenn der zeitinvariante Teil des Störterms unabhängig von X ist.
b	<input type="checkbox"/>	erfordert serielle Korrelation in den zeitvarianten Variablen.
c	<input type="checkbox"/>	liefert umso präzisere Schätzungen je geringer die Within-Variation der Regressoren ist.
d	<input type="checkbox"/>	Keine der genannten Antworten.

23.	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"/>	die Kovarianz zwischen Ausprägungen gleich weit entfernter Zeitpunkte konstant ist.
d	<input type="checkbox"/>	a und c.

24.	Folgt der Störterm in einem Zeitreihenmodell einem Random Walk mit Drift	
a	<input type="checkbox"/>	sollte das Modell in ersten Differenzen geschätzt werden.
b	<input type="checkbox"/>	Handelt es sich um einen stabilen AR(1) Prozess.
c	<input type="checkbox"/>	ist der KQ Schätzer BLUE.
d	<input type="checkbox"/>	a und b.
25.	Eine zur Lösung des Problems ausgelassener Variablen verwendete Proxy-Variable sollte	
a	<input type="checkbox"/>	mit der ausgelassenen Variable möglichst hoch korreliert sein.
b	<input type="checkbox"/>	möglichst hoch mit den übrigen erklärenden Variablen des Modells korreliert sein.
c	<input type="checkbox"/>	mit allen Regressoren des ursprünglichen Modells möglichst hoch korreliert sein.
d	<input type="checkbox"/>	b und c.
26.	Heteroskedastie	
a	<input type="checkbox"/>	kann bei i.i.d. Störtermen auftreten.
b	<input type="checkbox"/>	kann zusammen mit Autokorrelation auftauchen.
c	<input type="checkbox"/>	tritt im linearen Wahrscheinlichkeitsmodell auf.
d	<input type="checkbox"/>	b und c.
27.	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"/>	Keine der genannten Antworten.
28.	$E(u_i X) = 0$ impliziert	
a	<input type="checkbox"/>	$E(x_i) = 0$.
b	<input type="checkbox"/>	$cov(u_i, X) = 0$.
c	<input type="checkbox"/>	$E(u_i X = 1) = 1$.
d	<input type="checkbox"/>	Alle genannten Antworten.
29.	Instrumentvariablen	
a	<input type="checkbox"/>	können bei Modellen mit Interaktionsterm nicht verwendet werden.
b	<input type="checkbox"/>	eignen sich auch als Proxyvariablen.
c	<input type="checkbox"/>	dürfen mit den exogenen Regressoren korreliert sein.
d	<input type="checkbox"/>	a und c.
30.	Bei Fehlspezifikation der funktionalen Form	
a	<input type="checkbox"/>	kann das Einfügen von Interaktionstermen helfen.
b	<input type="checkbox"/>	ist keine der Gauss-Markov-Annahmen verletzt.
c	<input type="checkbox"/>	sollte ein Finite-Distributed-Lag Modell geschätzt werden.
d	<input type="checkbox"/>	Keine der genannten Antworten.
31.	$Cov(x_t, x_{t+1}) = Cov(x_t, x_{t-1})$ gilt stets, sobald	
a	<input type="checkbox"/>	strikte Stationarität vorliegt.
b	<input type="checkbox"/>	Kovarianzstationarität vorliegt.
c	<input type="checkbox"/>	schwache Abhängigkeit vorliegt.
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 $t \neq s$.