

Bachelorprüfung

Fach: Empirische Wirtschaftsforschung

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

Name, Vorname	
Matrikelnr.	
E-Mail-Adresse	
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 Verteilungen 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**[29 Punkte]**

Sie interessieren sich für die Determinanten von Zufriedenheit. Ihnen liegen Befragungsdaten von 2000 Personen aus dem Jahr 2005 vor, die ihre „allgemeine Lebenszufriedenheit“ auf einer Skala von 0 bis 10 angegeben haben. Daneben enthält der Datensatz einige andere Merkmale, mit denen Sie das Zufriedenheitsniveau erklären möchten:

- Zuf: allgemeine Lebenszufriedenheit (0 bis 10, 0= „sehr unzufrieden“ 10= „sehr zufrieden“)
 Alt: Alter in Jahren
 Mann: Geschlechtsdummy (1= männlich, 0= weiblich)
 Ehe: Familienstandsdummy (1= verheiratet, 0= nicht verheiratet)
 Eink: verfügbares Jahreseinkommen in 2004 (in €)
 Beh: Grad einer amtlich anerkannten Behinderung (0 bis 100)
 AL: Dummy, zeigt an ob eine Person zum Befragungszeitpunkt arbeitslos ist (1= ja, 0= nein)

Sie schätzen das Modell: $Zuf_i = \beta_0 + \beta_1 \cdot Alt_i + \beta_2 \cdot Mann_i + \beta_3 \cdot Ehe_i + u_i$

mittels der KQ-Methode und erhalten folgenden SPSS-Output:

Koeffizienten(a)

Modell		Koeffizienten		T	Signifikanz
		B	Standardfehler		
1	(Konstante)	8,436	2,349	3,591	,000
	Alt	-,053	,176	-,301	,763
	Mann	-,038	,059	-,643	,520
	Ehe	,431	,128	3,367	,000

a Abhängige Variable: Zuf

- a) Interpretieren Sie die Koeffizienten $\hat{\beta}_1$, $\hat{\beta}_2$ und $\hat{\beta}_3$ inhaltlich und im Hinblick auf Signifikanz jeweils unter der Annahme, dass die KQ-Annahmen gültig sind. (4,5 Punkte)
- b) Ein Kollege zweifelt die interne Validität Ihrer Schätzung an, da weder der Gesundheitszustand noch der Erwerbsstatus der Personen im Modell enthalten sind, diese nach bisherigen empirischen Erkenntnissen für die Lebenszufriedenheit aber bedeutsam sind. Außerdem sei davon auszugehen, dass die Lebenszufriedenheit mit dem Alter zunächst ab- und dann wieder zunähme. (6 Punkte)
- (1) Auf welche möglichen Ursachen für eine Verletzung interner Validität zielt der Kollege ab?
 - (2) Diskutieren Sie in Bezug auf den Ausschluss von Gesundheitszustand und Erwerbsstatus, ob Sie die Kritik für gerechtfertigt halten oder nicht.
- c) Sie ändern Ihr Modell wie folgt:
- $$Zuf_i = \beta_0 + \beta_{1a} \cdot Alt_i + \beta_{1b} \cdot Alt_i^2 + \beta_2 \cdot Mann_i + \beta_3 \cdot Ehe_i + \beta_4 \cdot Health_i + u_i$$
- (1) Welche Vorzeichen müssen die Koeffizienten $\hat{\beta}_{1a}$ und $\hat{\beta}_{1b}$ jeweils haben, wenn die These Ihres Kollegen über den Zusammenhang zwischen Alter und Zufriedenheit zutrifft? (1 Punkt)
 - (2) Da der Gesundheitszustand „Health“ nicht direkt beobachtbar ist, verwenden Sie die Variable „Beh“ als Proxy-Variable. Welche formalen Bedingungen muss die beobachtete Variable Beh erfüllen, damit dieses Vorgehen sinnvoll ist? (3 Punkte)

- d) In einem weiteren Schritt fügen Sie mit der Variable „Al“ auch den Erwerbsstatus Ihrem Modell hinzu.

$$Zuf_i = \beta_0 + \beta_{1a} \cdot Alt_i + \beta_{1b} \cdot Alt_i^2 + \beta_2 \cdot Mann_i + \beta_3 \cdot Ehe_i + \beta_4 \cdot Beh_i + \beta_5 \cdot Al_i + u_i$$

Betrachten Sie die Variablen Alt, Mann, Ehe und Beh als exogen.

- (1) Könnte ein Endogenitätsproblem in Bezug auf die Variable Al bestehen? Erläutern Sie Ihre Antwort inhaltlich. (2,5 Punkte)
 - (2) Welche formale Annahme des KQ-Modells wäre verletzt? Welche Konsequenzen hätte dies für die Eigenschaften des KQ-Schätzers? (2 Punkte)
- e) Sie haben noch eine weitere Variable „Clos“ im Datensatz zur Verfügung. Diese beschreibt, ob die Befragungsperson seit Beginn des Jahres 2004 in einem Betrieb beschäftigt war, der bis zum Befragungszeitpunkt geschlossen wurde. Sie möchten im Modell aus d) diese Variable als Instrument für „Al“ nutzen. (10 Punkte)
- (1) Welche formalen Bedingungen muss „Clos“ erfüllen, um ein gutes Instrument für „Al“ zu sein? Halten Sie diese Bedingungen für erfüllt? Erläutern Sie Ihre Überlegungen.
 - (2) Nehmen Sie an, die Bedingungen seien erfüllt. Wie lauten die Schätzgleichungen im Rahmen einer 2SLS-Schätzung auf der ersten und zweiten Stufe?

Aufgabe 2

[15,5 Punkte]

Sie wollen prüfen, wie das Zinsniveau das Wirtschaftswachstum beeinflusst. Sie haben dazu für 10 Jahre monatliche Beobachtungen für den Euroraum. Ihr Datensatz enthält folgende Variablen:

$gGDP_t$ monatliche prozentuale Änderung des Bruttoinlandsprodukts im Euroraum

int_t kurzfristiger Zinssatz in Prozent, gesetzt durch die Europäische Zentralbank

- a) Sie schätzen das Modell $gGDP_t = \alpha_0 + \delta_0 int_t + \delta_1 int_{t-1} + \delta_2 int_{t-2} + u_t$ und erhalten folgenden Regressionsoutput: (6,5 Punkte)

Koeffizienten(a)

Modell	Koeffizienten		T	Signifikanz
	B	Standardfehler		
1 (Konstante)	4,682	3,208	1,459	,147
int[t]	-,083	,120	-0,692	,490
int[t-1]	-,238	,058	-4,103	,000
int[t-2]	,094	,051	1,843	,068

a Abhängige Variable: $gGDP_t$

- (1) Interpretieren Sie die Steigungsparameter einzeln inhaltlich und statistisch.
 - (2) Wie hoch sind der langfristige und der kurzfristige Multiplikator von Zinsänderungen?
- b) Wie würden Sie im Modell aus a) jährliche Trends in der abhängigen Variable und jahreszeitliche Einflüsse berücksichtigen? Welche Kodierungen müssen Sie vornehmen? Stellen Sie das neue Modell dar. (3 Punkte)

- c) Gehen Sie davon aus, dass die Europäische Zentralbank den kurzfristigen Zins anhand folgender Formel festlegt: $int_t = \gamma_0 + \gamma_1(gGDP_{t-1} - 3) + v_t$ und $\gamma_1 > 0$ ist. Kann das Modell in b) bei gegebener Datenlage mit KQ i) erwartungstreu und ii) konsistent geschätzt werden? Begründen Sie und gehen Sie auf die erforderlichen Annahmen ein. (Es seien TS.1; TS.2 sowie TS.1' und TS.2' (siehe Tabellenanhang) erfüllt.) (6 Punkte)

Aufgabe 3

[11,5 Punkte]

Betrachten Sie erneut die Situation aus Aufgabe 2. Sie verwenden nun ein verkürztes Modell $gGDP_t = \alpha_0 + \delta_0 int_t + u_t$ und vermuten, dass positive Autokorrelation erster Ordnung ein Problem darstellen könnte.

- a) Was versteht man unter Autokorrelation? Welche Konsequenzen ergeben sich für den KQ-Schätzer und nachfolgende Tests? (4 Punkte)
- b) Wie können Sie den Autokorrelationskoeffizienten im AR(1)-Modell bestimmen? Erläutern Sie Ihre Vorgehensweise. (2 Punkte)
- c) Gehen Sie davon aus, dass Sie den Autokorrelationskoeffizienten kennen und dieser zwischen -1 und 1 liegt. Man kann die Schätzgleichung so transformieren, dass ein Modell ohne serielle Korrelation resultiert. Stellen Sie am Beispiel diese Transformation dar und benennen Sie die Auswirkung auf den Störterm. (Ignorieren Sie dabei die gesonderte Behandlung der ersten Beobachtung.) (5,5 Punkte)

Aufgabe 4

[24 Punkte]

Wahr oder falsch? Tragen Sie für jede der folgenden Aussagen ein „w“ für „wahr“ oder ein „f“ für „falsch“ ein. Für jede richtige Antwort gibt es 0,75 Punkte, für jede falsche Antwort werden 0,75 Punkte abgezogen. Die Gesamtpunktzahl kann nicht negativ werden.

	Beim Schätzen mit Paneldaten in ersten Differenzen werden konstante, unbeobachtete Effekte kontrolliert.
	Beim Schätzen mit Paneldaten in ersten Differenzen werden die Effekte von konstanten Variablen oft nur unpräzise geschätzt.
	Das Weighted Least Squares Verfahren wird bei Heteroskedastie angewendet, da es dann effizienter sein kann als das Kleinstquadrat-Verfahren.
	KQ-Schätzer sind konsistent, wenn die abhängige Variable stationär ist.
	Der Davidson-MacKinnon-Test gilt auch für nicht genestete Modelle.
	Bei T=2 ergibt die within-Schätzung dieselben Ergebnisse wie eine Schätzung in ersten Differenzen.
	Heteroskedastierobuste t-Werte sind größer als die herkömmlichen.
	Bei Verwendung eines Finite Distributed Lag Modells zweiter Ordnung wird unterstellt, dass Änderungen eines Regressors, die länger als 2 Perioden zurückliegen, keinen Einfluss auf die abhängige Variable haben.
	Für die Eigenschaften des Kleinstquadrat-Schätzers ist es bedeutsam, ob erklärende Variablen und Störterm korreliert sind.
	Statische Modelle können nur mit Zeitreihendaten geschätzt werden.
	Schwache Abhängigkeit erfordert einen Erwartungswert von Null.
	Das difference-in-differences-Verfahren wird häufig in Verbindung mit natürlichen Experimenten verwendet.
	Externe Validität kann durch Unterschiede in Grundgesamtheiten eingeschränkt werden.

	In einer Regressionsgleichung kann nicht gleichzeitig für saisonale Schwankungen und polynomiale Trends kontrolliert werden.
	Das difference-in-differences-Verfahren kann nicht bei logarithmierten abhängigen Variablen verwendet werden, da $\log(\bar{x}_{treatment}) - \log(\bar{x}_{control}) \neq \log(\bar{x}_{treatment} - \bar{x}_{control})$.
	Exogene Stichprobenselektion führt zu verzerrten Parameterschätzern.
	Da die Steigungsparameter bei Heteroskedastie falsch berechnet werden, ist auch der Achsenabschnitt falsch.
	Annahme TS.5: $\text{corr}(u_t, u_s X) = 0$ besagt, dass weder u_s noch u_t mit X korreliert sein dürfen. Es gilt dabei $s \neq t$.
	Panelschätzungen mittels einer Dummy-Variablen-Regression (LSDV) und mittels des within-Verfahrens liefern identische Koeffizienten.
	Wenn nicht alle Beobachtungseinheiten i gleich viele Beobachtungsperioden haben, spricht man von „unbalanced panels“.
	Im Fall einer binären abhängigen Variable liegt bei einer Kleinstquadrat-Schätzung Heteroskedastie vor.
	Die erklärenden Variablen x_t sind kontemporär exogen, wenn gilt $E(u_t x_t) = 0$.
	Einer IV-Schätzung können exogene erklärende Variablen als Kontrollvariablen hinzugefügt werden.
	Wenn Exogenität der erklärenden Variablen vorliegt, ist der 2SLS-Schätzer effizienter als KQ.
	Bei Heteroskedastie ist der White-Schätzer dem Kleinstquadrat-Schätzer zur Bestimmung der Standardfehler vorzuziehen.
	Ist die abhängige Variable binär, muss im linearen Modell Autokorrelation vorliegen.
	Damit Kovarianzstationarität vorliegt, muss die Varianz der Zufallsvariable konstant sein.
	Wird der Zusammenhang zwischen abhängiger und erklärenden Variablen nicht korrekt abgebildet, hat man es mit dem Problem der Fehlspezifikation zu tun.
	Für die within-Schätzung muss der Stichprobenmittelwert einer Variablen von den tatsächlichen Beobachtungswerten abgezogen werden: man bildet aus der erklärenden Variable x_{it} die neue Variable $(x_{it} - \bar{x})$.
	Eine Dummy-Variablen kann aufgrund der sich ergebenden Heteroskedastieproblematik nicht als Proxy-Variablen verwendet werden.
	Kann die H_0 im White-Test nicht verworfen werden, liegt Heteroskedastie vor.
	Eine Trendbereinigung macht nur Sinn, wenn gleichzeitig um Saisoneffekte bereinigt wird.

Aufgabe 5

[10 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.	In einem RESET-Test mit \hat{y}^2 und \hat{y}^3 für 66 Beobachtungen und 3 Steigungsparametern ergibt sich ein Wert der Teststatistik von 2,8. Daher
a	<input type="checkbox"/> kann die H_0 auf dem 5% Signifikanzniveau nicht verworfen werden.
b	<input type="checkbox"/> kann die H_0 auf dem 10% Signifikanzniveau nicht verworfen werden.
c	<input type="checkbox"/> ergibt der Test am 5%-Signifikanzniveau, dass Endogenität vorliegt.
d	<input type="checkbox"/> a und b.
e	<input type="checkbox"/> a und c.
f	<input type="checkbox"/> keine der Antworten.

2.	Liegt ein Messfehler in der abhängigen Variable vor,	
a	<input type="checkbox"/>	sind die mit KQ geschätzten Parameter verzerrt.
b	<input type="checkbox"/>	wird die Regressionskonstante verzerrt geschätzt, wenn der Messfehler mit den erklärenden Variablen korreliert ist.
c	<input type="checkbox"/>	ergeben sich unverzerrte Parameterschätzer, wenn der Messfehler nicht mit den erklärenden Variablen korreliert ist.
d	<input type="checkbox"/>	a und b und c.
e	<input type="checkbox"/>	b und c.
f	<input type="checkbox"/>	keine der Antworten.

3.	Panelverfahren	
a	<input type="checkbox"/>	können auch auf Datenstrukturen ohne zeitliche Dimension angewendet werden, z.B. Clusterstichproben.
b	<input type="checkbox"/>	sind nur für balanced panels konsistent und normalverteilt.
c	<input type="checkbox"/>	können nur mit Regressionskonstante geschätzt werden.
d	<input type="checkbox"/>	sind bei Vorliegen einer Panelstruktur der Daten dem KQ-Schätzer in der Regel vorzuziehen.
e	<input type="checkbox"/>	a und d.
f	<input type="checkbox"/>	b und c.

4.	Voraussetzung für die Gültigkeit der Annahme $E(u_t X) = 0$ ist, dass	
a	<input type="checkbox"/>	die Störterme aller Perioden untereinander unkorreliert sind.
b	<input type="checkbox"/>	die Störterme nicht mit X korreliert sind.
c	<input type="checkbox"/>	$E(u_t)=0$ gilt.
d	<input type="checkbox"/>	die Regressoren keine Linearkombinationen anderer Regressoren sind.
e	<input type="checkbox"/>	b und c.
f	<input type="checkbox"/>	c und d.

5.	$Cov(x_t; x_{t+h}) = Cov(x_{t+2}; x_{t+h+2})$ ist eine notwendige Voraussetzung für	
a	<input type="checkbox"/>	schwache Abhängigkeit.
b	<input type="checkbox"/>	Kovarianzstationarität.
c	<input type="checkbox"/>	strikte Stationarität.
d	<input type="checkbox"/>	Konsistenz von KQ.
e	<input type="checkbox"/>	b und c.
f	<input type="checkbox"/>	a und b.

6.	Im Fall seriell korrelierter Störterme	
a	<input type="checkbox"/>	ist die wahre Varianz von $\hat{\beta}_j$ anders als vom KQ-Schätzer ausgewiesen.
b	<input type="checkbox"/>	kann ein geschätzter Parameter $\hat{\beta}_j$ fälschlicherweise als insignifikant ausgewiesen werden.
c	<input type="checkbox"/>	kann ein geschätzter Parameter $\hat{\beta}_j$ fälschlicherweise als signifikant ausgewiesen werden.
d	<input type="checkbox"/>	a und b.
e	<input type="checkbox"/>	a und b und c.
f	<input type="checkbox"/>	keine der Antworten.

7.	Unter Gültigkeit der Annahmen TS.1 bis TS.3 (siehe Tabellenanhang) ist KQ	
a	<input type="checkbox"/>	BLUE.
b	<input type="checkbox"/>	unverzerrt.
c	<input type="checkbox"/>	effizient.
d	<input type="checkbox"/>	asymptotisch normalverteilt.
e	<input type="checkbox"/>	a und c.
f	<input type="checkbox"/>	c und d.

8.	Ein statisches Modell	
a	<input type="checkbox"/>	kann keine Autokorrelation aufweisen.
b	<input type="checkbox"/>	ist homoskedastisch.
c	<input type="checkbox"/>	misst Effekte, die in derselben Periode wirken, in der sie verursacht werden.
d	<input type="checkbox"/>	beinhaltet verzögerte erklärende Variablen.
e	<input type="checkbox"/>	a und b und c.
f	<input type="checkbox"/>	a und d.

9.	Damit schwache Abhängigkeit für y_t vorliegt, muss gelten:	
a	<input type="checkbox"/>	$\text{cov}(y_t; y_{t+h}) \rightarrow 0$ wenn $h \rightarrow \infty$.
b	<input type="checkbox"/>	$\text{cov}(y_t; y_{t+h}) \rightarrow 0$ wenn $t \rightarrow \infty$.
c	<input type="checkbox"/>	$\text{cov}(y_t; y_{t+h}) \rightarrow 1$ wenn $h \rightarrow \infty$.
d	<input type="checkbox"/>	dass Kovarianzstationarität vorliegt.
e	<input type="checkbox"/>	b und d.
f	<input type="checkbox"/>	a und c.

10.	Exogene Stichprobenselektion	
a	<input type="checkbox"/>	bedeutet systematische Stichprobenselektion auf Basis beobachtbarer, exogener Variablen.
b	<input type="checkbox"/>	führt zu verzerrten Parameterschätzern.
c	<input type="checkbox"/>	führt zu inkonsistenten Parameterschätzern.
d	<input type="checkbox"/>	führt tendenziell zu insignifikanten Parameterschätzern.
e	<input type="checkbox"/>	a und d.
f	<input type="checkbox"/>	b und c.