

Bachelorprüfung

Fach: Empirische Wirtschaftsforschung II

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

Name, Vorname	
Matrikelnr.	
Studiengang	
Semester	
Datum	
Raum	
E-Mail Adresse	
Unterschrift	

Vorbemerkungen:

Anzahl der Aufgaben: Die Klausur besteht aus 7 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:**[19 Punkte]**

Sie möchten prüfen, ob Gewerkschaftmitgliedschaft mit dem Lohn zusammenhängt. Ihnen liegen dazu 1327 Beobachtungen von 500 Vollzeit arbeitenden Personen aus den Jahren 2002 bis 2004 vor. Folgendes Modell soll zunächst mittels KQ geschätzt werden:

$$w_{it} = \beta_0 + \beta_1 \text{gewerkschaft}_{it} + \beta_2 \text{mann}_{it} + \beta_3 \text{erf}_{it} + a_i + \varepsilon_{it}$$

mit i als Personenindikator und t als Zeitindikator.

w = Nettostundenlohn gemessen in €

gewerkschaft = 1, wenn Person i zum Zeitpunkt t Gewerkschaftmitglied ist, sonst 0.

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

erf = Berufserfahrung der Person gemessen in Jahren

a, ε = unbeobachtete Faktoren (Störterm)

- Was ist der Unterschied zwischen einem balanced und einem unbalanced Panel? Welche der beiden Alternativen trifft auf Ihren Datensatz zu? Erläutern Sie Ihre Antwort kurz. (3 Punkte)
- Erklären Sie, weshalb β_1 im vorliegenden Beispiel mit KQ bei Nichtberücksichtigung der Panelstruktur inkonsistent geschätzt werden könnte. (2 Punkte)
- 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)
- Schreiben Sie für das gegebene Beispiel die Schätzgleichung für einen Within-Schätzer sowie für einen First Difference Schätzer auf. Erklären Sie, inwieweit Sie mit Hilfe dieser Schätzansätze das in Teilaufgabe b) angesprochene Problem lösen können. (6 Punkte)
- Gehen Sie davon aus, dass ε_{it} einem Random Walk folgt. Welches Schätzverfahren würden Sie vorziehen, Within oder First Difference Schätzer? Erläutern Sie Ihre Antwort ausführlich. (5 Punkte)

Aufgabe 2:**[6 Punkte]**

Sie möchten prüfen, ob sich die Anzahl an Geschwistern negativ auf den Bildungserfolg einer Person auswirkt, beispielsweise weil Kinder die bildungsrelevanten elterlichen Ressourcen mit ihren Geschwistern teilen müssen. Sie schätzen folgendes Modell mit i als Personenindikator:

$$\text{abi}_i = \beta_0 + \beta_1 \text{geschwister}_i + \beta_2 \text{mann}_i + \varepsilon_i$$

abi = 1, wenn Person Abitur hat, sonst 0.

geschwister = Anzahl Geschwister

mann = 1, wenn Person männlich, sonst 0.

- Erklären Sie unter Angabe eines Beispiels, warum geschwister eine endogene Variable sein könnte. (2 Punkte)
- Sie haben zusätzlich die Variable abtreibung zur Verfügung. Diese Variable gibt an, ob die Mutter der Person schon einmal eine Schwangerschaft abgebrochen hat. Diskutieren Sie, inwieweit abtreibung ein geeignetes Instrument für geschwister darstellt. (4 Punkte)

Aufgabe 3:**[8 Punkte]**

- Erläutern sie knapp was unter schwacher Abhängigkeit einer Zeitreihe zu verstehen ist. (2 Punkte)
- Zeigen Sie, dass ein MA(1) Prozess der Art $x_t = e_t + 0,5e_{t-1}$ 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 $\text{Var}[e_t] = 1$ ist. (6 Punkte)

Aufgabe 4:**[14 Punkte]**

- a) Was ist mit serieller Korrelation im Störterm einer linearen Regression gemeint? (1 Punkt)
- b) Welche negativen Konsequenzen hat ein autoregressiver Prozess erster Ordnung im Störterm des Modells $y_t = \beta_0 + \beta_1 x_t + e_t$ für eine Schätzung mit KQ, wenn die Annahmen TS.1' bis TS.3' erfüllt sind? (2 Punkte)
- c) Die Hilfsregression $\hat{r}_t = \alpha_0 + \alpha_1 \hat{r}_{t-1} + \alpha_2 \hat{r}_{t-2} + \alpha_3 x_t + u_t$, wobei \hat{r}_t das Residuum aus der Schätzung in Teilaufgabe 4b) ist, liefert Ihnen für α_1 den Wert 0.6 mit einem Standardfehler von 0.2 und für α_2 den Wert 0.02 mit einem Standardfehler von 0.1. Auf welche Form der seriellen Korrelation weist dieses Ergebnis hin? (1 Punkt)
- d) Gehen Sie nun davon aus, dass u_t aus der Schätzgleichung in Teilaufgabe c) heteroskedastisch ist. Was ist Heteroskedastie und welche Folgen hat es für die Schätzung in Teilaufgabe c) und ihre dort getroffene Schlussfolgerung? Wie können Sie dieses Problem möglichst einfach beheben. (5 Punkte)
- e) Der Störterm der Regressionsgleichung $y_t = \beta_0 + \beta_1 x_t + e_t$ folge folgendem Prozess: $e_t = \rho e_{t-1} + v_t$ (v_t erfüllt die Gauss-Markov Annahmen). Schreiben Sie die transformierte Schätzgleichung für das Cochrane-Orcutt Verfahren auf. Erläutern Sie hierbei kurz das Ziel des Verfahrens. (3 Punkte)
- f) Welches Problem wird durch das Cochrane-Orcutt Verfahren behoben? (1 Punkt)
- g) Welchen Nachteil hat das Cochrane-Orcutt Verfahren im Vergleich zum Prais-Winsten Verfahren? (1 Punkt)

Aufgabe 5:**[11 Punkte]**

Nach der Fertigstellung eines Atomkraftwerks (AKW) im Jahre 1984 behaupten die im Umkreis von 5km wohnenden Hausbesitzer, dass der Preis Ihrer Häuser aufgrund des Baus des AKW gesunken sei. Ihnen liegen Daten über regionale Hauspreise aus dem Jahr 1978 (vor Bekanntgabe der Bauabsicht für das AKW) und aus dem Jahr 1985 vor und sie schätzen die Regression:

$$\ln(\text{preis}) = \beta_0 + \beta_1 d85 + \beta_2 \text{radius} + \beta_3 (d85 * \text{radius}) + u.$$

preis = Preis des Hauses in D-Mark

$d85$ = 1, wenn die Beobachtung aus dem Jahr 1985 ist, sonst = 0

radius = 1, wenn die Beobachtung zu einem Haus im Umkreis von 5km um das AKW gehört

Sie erhalten folgendes Regressionsergebnis (Standardfehler in Klammern):

$$\beta_0 = 11,51 (2,22) \quad \beta_2 = -0,28 (0,08)$$

$$\beta_1 = 0,11 (0,12) \quad \beta_3 = -0,04 (0,03)$$

- a) Interpretieren Sie die geschätzten Parameter β_1 und β_2 inhaltlich und statistisch. (6 Punkte)
- b) Hat der Bau des AKW die Hauspreise im 5km Radius um das Kraftwerk negativ beeinflusst? Begründen Sie. (3 Punkte)
- c) Wie heißt das hier verwendete KQ-Schätzverfahren und welche zentrale Annahme muss neben den Gauss-Markov Annahmen gelten, damit Sie den Effekt des AKW identifizieren können? (2 Punkte)

Aufgabe 6:**[22 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 0,5 Punkte, für jede falsche Antwort werden 0,5 Punkte abgezogen. Die Gesamtpunktzahl kann nicht negativ werden.

	Bei Heteroskedastie ist der White-Schätzer dem KQ-Schätzer zur Bestimmung der Standardfehler vorzuziehen.
	Finite Distributed Lag Modelle weisen mindestens einen verzögerten Wert von x auf.

	Sind die Regressoren eines Zeitreihenmodells nicht strikt exogen, so sind die KQ Schätzer stets inkonsistent.
	Die Relevanz eines Instruments lässt sich empirisch testen.
	Das Spurious Regression Problem bei Zeitreihenregressionen tritt nur dann auf, wenn abhängige und erklärende Variablen einem gegenlaufenden Trend folgen.
	Sind die Störterme autokorreliert, ist der KQ-Schätzer effizient.
	Für einen autoregressiven Prozess x_t der Ordnung a gilt $Cov[x_t, x_{t+h}] = 0$ sobald $h > a$.
	Ein Random Walk Prozess ist kovarianzstationär.
	Die erste Differenz eines Random Walk Prozesses ist schwach abhängig.
	Einheitswurzelprozesse können durch Wurzelziehen in schwach abhängige Prozesse überführt werden.
	Bei positiver Autokorrelation sind die von KQ ausgewiesenen Standardfehler falsch.
	Tritt in einem Modell $y_t = \beta_0 + \beta_1 y_{t-1} + u_t$ Autokorrelation erster Ordnung in u_t auf, liefert die KQ Schätzung inkonsistente Koeffizienten.
	Das LAD Verfahren minimiert die Summe der quadrierten Absolutwerte der Residuen.
	In dynamisch vollständigen Modellen sind die Störterme nicht autokorreliert.
	Liefern White-Schätzer und Prais-Winsten Schätzer stark unterschiedliche Koeffizienten, sollte das Schätzverfahren mit der niedrigeren Residuenquadratsumme verwendet werden.
	Sind die Regressoren eines Modells nicht contemporär exogen, liefern sowohl Cochrane-Orcutt als auch Prais-Winsten Schätzer inkonsistente Schätzergebnisse.
	Beim Schätzen mit Paneldaten in ersten Differenzen werden zeitkonstante, unbeobachtete Effekte herausgerechnet.
	Eine wichtige Annahme beim Difference-in-Differences Verfahren ist, dass sich die Treatmentgruppe zu beiden Beobachtungszeitpunkten identisch zusammensetzt.
	Eine Trendbereinigung kann auch dann Sinn machen, wenn gleichzeitig Saisoneffekte bereinigt werden.
	Das White-Schätzverfahren dient zur Korrektur der Standardfehler bei Vorliegen von Heteroskedastie.
	Newey-West t-Werte sind stets größer als die bei KQ Ausgewiesenen.
	Stichprobenselektion anhand der Ausprägungen der abhängigen Variablen ist für die Parameterschätzer stets unproblematisch.
	Eine stetige Variable kann als Proxy-Variable verwendet werden.
	Im Fall einer binären abhängigen Variable liegt bei einer KQ Schätzung perfekte Multikollinearität vor.
	Kann die H_0 im Breusch-Godfrey Test verworfen werden, schließen wir, dass Heteroskedastie vorliegt.
	Treten in den Daten fehlende Werte zufällig auf, führt das Weglassen der Beobachtungen mit fehlenden Werten nicht zu verzerrten KQ Schätzern.
	Die Verwendung von verzögerten endogenen Variablen auf der rechten Seite eines Zeitreihenmodells führt zwangsläufig zu inkonsistenten Schätzergebnissen.
	Ausreißer werden beim LAD Verfahren stärker gewichtet als beim KQ Verfahren.
	Der Durbin-Watson Test unterstellt strikt exogene Regressoren.
	Ist der $Median(y) < E(y)$, so könnte dies an positiven Ausreißern in y liegen.
	Externe Validität ist gegeben, wenn die Ergebnisse einer Studie auch auf andere Grundgesamtheiten übertragen werden können.
	Interne Validität ist gegeben, sobald die Koeffizienten unverzerrt geschätzt wurden.
	Man spricht von einem Balanced Panel, wenn die Anzahl der Querschnittsbeobachtungen der Anzahl der Wiederholungsbeobachtungen entspricht.
	Der Breusch-Pagan Test kann auch als F-Test durchgeführt werden.
	Es ist nicht möglich, mittels FGLS Schätzung Heteroskedastie und Autokorrelation gleichzeitig zu berücksichtigen.
	Wenn $T=2$, liefern der Within Schätzer und der Schätzer in ersten Differenzen identische Koeffizienten und Standardfehler.
	Damit Kovarianzstationarität vorliegt, muss der Erwartungswert der Zufallsvariable konstant

	sein.
	Liegt negative Autokorrelation erster Ordnung vor, ist die DW Statistik < 2 .
	Der Durbin Watson Test lässt sich nur auf AR(1) Prozesse anwenden.
	Im Random Effects Modell werden den Beobachtungseinheiten zufällige Effekte der erklärenden Variablen auf die Konstante zugewiesen.
	Der Between Schätzer kann den Effekt zeitkonstanter erklärender Variablen nicht bestimmen.
	Das Fixed Effects Modell kann auch auf gepoolte Querschnittsdaten angewendet werden.
	Das Random Effects Modell eignet sich nicht für gepoolte Querschnitte.
	Im Trendmodell $\log(y_t) = \beta_0 + \beta_1 t + e_t$ kann β_0 approximativ als mittlere periodische Wachstumsrate von y interpretiert werden.

Aufgabe 7:

[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.	Das Least Absolute Deviations Verfahren	
a	<input type="checkbox"/>	liefert präzisere Ergebnisse als KQ, wenn die erklärenden Variablen multikollinear sind.
b	<input type="checkbox"/>	minimiert die Summe der absoluten Abweichungen von der Regressionsgerade.
c	<input type="checkbox"/>	wird zur Schätzung des auf x bedingten Medians verwendet.
d	<input type="checkbox"/>	ist vorteilhaft, wenn sich von der übrigen Verteilung stark abweichende Beobachtungswerte im Datensatz befinden.
e	<input type="checkbox"/>	b, c und d.
f	<input type="checkbox"/>	Keine der Antworten.

2.	Eine Proxy-Variable	
a	<input type="checkbox"/>	sollte mit der ausgelassenen Variable möglichst hoch korreliert sein.
b	<input type="checkbox"/>	sollte über die Korrelation mit der ausgelassenen Variable hinaus keinen eigenständigen Effekt auf die abhängige Variable haben.
c	<input type="checkbox"/>	sollte möglichst hoch mit den übrigen erklärenden Variablen des Modells korreliert sein.
d	<input type="checkbox"/>	sollte mit allen erklärenden Variablen des ursprünglichen Modells möglichst hoch korreliert sein.
e	<input type="checkbox"/>	a und b.
f	<input type="checkbox"/>	b und c.

3.	Bei Messfehlern	
a	<input type="checkbox"/>	in der abhängigen Variable werden die Koeffizienten stets verzerrt geschätzt.
b	<input type="checkbox"/>	in der abhängigen Variable werden die Koeffizienten inkonsistent geschätzt.
c	<input type="checkbox"/>	in der abhängigen Variable sind die Standardfehler der Koeffizienten beeinträchtigt.
d	<input type="checkbox"/>	in der erklärenden Variable werden die R^2 Werte zu hoch ausgewiesen.
e	<input type="checkbox"/>	a und b.
f	<input type="checkbox"/>	Keine der genannten Antworten.

4.	Attenuation bias	
a	<input type="checkbox"/>	tritt auf bei Messfehlern in der abhängigen Variable.
b	<input type="checkbox"/>	ist eine Folge von Heteroskedastie.
c	<input type="checkbox"/>	ist eine Folge von Autokorrelation.
d	<input type="checkbox"/>	ist eine Folge von Multikollinearität.
e	<input type="checkbox"/>	a und b.
f	<input type="checkbox"/>	Keine der genannten Antworten.

5.	Exogene Stichprobenselektion	
a	<input type="checkbox"/>	erfolgt auf Basis der abhängigen Variable.
b	<input type="checkbox"/>	kann auf Basis einer der erklärenden Variable erfolgen.
c	<input type="checkbox"/>	führt zu inkonsistenten Schätzern.
d	<input type="checkbox"/>	verringert die Stichprobengröße.
e	<input type="checkbox"/>	b und d.
f	<input type="checkbox"/>	b, c und d.

6.	Im linearen Wahrscheinlichkeitsmodell	
a	<input type="checkbox"/>	ist die erklärende Variable eine Dummy Variable.
b	<input type="checkbox"/>	tritt zwangsläufig Autokorrelation auf.
c	<input type="checkbox"/>	tritt zwangsläufig Heteroskedastie auf.
d	<input type="checkbox"/>	können bei einer FGLS Schätzung zur Behebung von Heteroskedastie nie alle Beobachtungen verwendet werden.
e	<input type="checkbox"/>	c und d.
f	<input type="checkbox"/>	Keine der genannten Antworten.

7.	Bei Fehlspezifikation der funktionalen Form	
a	<input type="checkbox"/>	Sind die Schätzer inkonsistent.
b	<input type="checkbox"/>	Ist der Störterm mit den erklärenden Variablen korreliert.
c	<input type="checkbox"/>	Sollte der RESET Test die H0 verwerfen.
d	<input type="checkbox"/>	Kann es vorteilhaft sein, weitere Polynome der erklärenden Variablen mit aufzunehmen.
e	<input type="checkbox"/>	a und b.
f	<input type="checkbox"/>	Alle genannten Antworten.

8.	Der Between Schätzer für Paneldaten mit $e_{it} = a_i + u_{it}$	
a	<input type="checkbox"/>	ist auch dann konsistent, wenn die erklärenden Variablen und a_i korreliert sind.
b	<input type="checkbox"/>	ist auch dann konsistent, wenn die erklärenden Variablen und u_{it} korreliert sind.
c	<input type="checkbox"/>	schätzt KQ für ein Modell individuenspezifischer Mittelwerte.
d	<input type="checkbox"/>	ist inkonsistent bei Heteroskedastie.
e	<input type="checkbox"/>	b und c.
f	<input type="checkbox"/>	Keine der genannten Antworten.

9.	Der LSDV Schätzer für Paneldaten mit $e_{it} = a_i + u_{it}$	
a	<input type="checkbox"/>	ist auch dann konsistent, wenn die erklärenden Variablen und a_i korreliert sind.
b	<input type="checkbox"/>	ist auch dann konsistent, wenn die erklärenden Variablen und u_{it} korreliert sind.
c	<input type="checkbox"/>	und der Within Schätzer für Paneldaten liefern stets identische Schätzergebnisse bzgl. der Steigungsparameter.
d	<input type="checkbox"/>	Berechnet für jede Beobachtungseinheit einen eigenen Achsenabschnitt.
e	<input type="checkbox"/>	a, c und d.
f	<input type="checkbox"/>	Keine der genannten Antworten.

10.	Gepoolte Querschnittsdaten	
a	<input type="checkbox"/>	liegen vor, wenn Querschnittsdaten aus mehreren Jahren vorhanden sind.
b	<input type="checkbox"/>	erlauben die Anwendung des Difference-in-Differences-Schätzers.
c	<input type="checkbox"/>	ermöglichen die Anwendung eines First Difference Schätzers.
d	<input type="checkbox"/>	führen bei der Anwendung von KQ stets zu heteroskedastischen Störtermen.
e	<input type="checkbox"/>	a und b.
f	<input type="checkbox"/>	a, b und c.

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 $t \neq s$.