Lehrstuhl für Statistik und empirische Wirtschaftsforschung

Prof. Regina T. Riphahn, Ph.D.

Diplomprüfung im Fach Ökonometrie im SS 06 - Aufgabenteil

Name, Vorname	
Matrikelnr.	
Studiengang	
Semester	
Datum	**31.07.2006**
Raum, Sitzplatz-Nr.	**H6,**
Unterschrift	

Vorbemerkungen:

Anzahl der Aufgaben: Die Klausur besteht aus 7 Aufgaben, von denen alle bearbeitet werden müssen.

Bewertung: Die Prüfung dauert 120 Minuten, es können maximal 120 Punkte erworben werden. Die Punktzahl ist für jede Aufgabe in Klammern angegeben.

Erlaubte Hilfsmittel:

- 2 DIN A4-Blätter mit Notizen (Vorder- und Rückseite, also max. 4 DIN A4-Seiten)
- Tabellen der statistischen Verteilungen (sind der Klausur beigefügt)
- 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 Annahme fehlt, machen Sie dies kenntlich und treffen Sie für den fehlenden Wert eine plausible Annahme.
- Die Aufgaben 6 und 7 sind im Aufgabenteil zu beantworten, die restlichen Aufgaben im Lösungsteil. Verwenden Sie hierbei für jede Aufgabe ein neues Blatt.

Aufgabe 1: **[22 Punkte]**

Eine US-amerikanische Umweltbehörde untersucht die Determinanten der Luftqualität in Kalifornien. Die Behörde sammelt dazu für 21 Küsten- und 9 Nichtküstenregionen Daten über die folgenden Indikatoren:

airq:	Index für Luftverschmutzung (Wert=200 bei maximaler, =0 bei minimaler Verschmutzung)
vala:	Wertschöpfung von Unternehmen (in 1000 US-$)
rain:	Niederschlag (in Kubik-Zoll)
coas:	Dummyvariable über regionale Lage (1 = Küstenregion, 0 = sonst)
dens:	Bevölkerungsdichte (pro Quadrat-Meile)
medi:	durchschnittliches Pro-Kopf-Einkommen (in US-$)

Die Behörde unterstellt folgendes Modell:

$$airq_i = \beta_0 + \beta_1 \cdot vala_i + \beta_2 \cdot rain_i + \beta_3 \cdot coas_i + \beta_4 \cdot dens_i + \beta_5 \cdot medi_i + e_i$$

Die Auswertung der Daten mit R ergibt folgenden Output:

```
Call:
lm(formula = airq ~ vala + rain + coas + dens + medi)

Coefficients:
              Estimate     Std. Error    t value    Pr(>|t|)
(Intercept)   111.9347        15.3318       ?       1.53e-07 ***
vala            0.0009         0.0023     0.3915    0.6989
rain            0.2507         0.3435     0.7298    0.4726
coas          -33.3983         ?         -3.1937    0.0039 **
dens           -0.0011         0.0016    -0.6612    0.5148
medi            0.6211         0.3881     1.6003    0.1226
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 24.2 on 24 degrees of freedom
Multiple R-Squared: 0.3829,     Adjusted R-squared: ?
F-statistic: ?  on 5 and 24 DF,  p-value: 0.03133
```

a) Berechnen Sie unter Angabe des Rechenwegs (6 Punkte)

a1) den t-Wert für b_0;

a2) den Standardfehler für b_3;

a3) ein 95% Konfidenzintervall für b_2;

a4) die F-Statistik;

a5) das korrigierte R^2.

b) Interpretieren Sie die Koeffizienten b_3 und b_5 inhaltlich und statistisch. (2 Punkte)

c) In einer zusätzlichen Schätzung wird statt des absoluten Wertes der logarithmierte Wert der Bevölkerungsdichte verwendet. Der geschätzte Koeffizient beträgt -0.0039, der dazugehörige t-Wert -2.061. Interpretieren Sie inhaltlich und statistisch. (2 Punkte)

d) In einer weiteren Schätzung wird dem Modell aus a) ein Interaktionsterm `coas*medi` hinzugefügt. Die geschätzte Regressionsgerade lautet: `airq = 112 + 0.0025*vala + 0.2881*rain – 37.27*coas – 0.0011*dens – 0.55*medi + 0.079*(coas*medi)`. Berechnen Sie den marginalen Effekt des Durchschnittseinkommens für Küstenregionen und für Nichtküstenregionen und interpretieren Sie diese inhaltlich. (3 Punkte)

e) Beschreiben Sie die Vorgehensweise des Breusch-Pagan Tests. Die aus dem ersten Modell (siehe Output) resultierende Teststatistik ist BP=3.1416. Testen Sie auf dem 5%-Signifikanzniveau, ob Heteroskedastie vorliegt. (5 Punkte)

f) Es besteht weiterhin die Vermutung, dass die Fehlertermvarianz mit steigendem Einkommen variiert: $var(e_i) = \sigma_i^2 = \sigma^2 * medi_i$, was zu Heteroskedastie führen kann. (4 Punkte)

f1) Beschreiben Sie eine Vorgehensweise, das zugrunde liegende Modell so zu transformieren, dass ein Modell mit homoskedastischen Fehler resultiert und zeigen Sie letzteres.

f2) Wie unterscheidet sich die Interpretation der Schätzergebnisse des Modells mit transformierten Daten von derjenigen der Originalschätzung?

Aufgabe 2: **[8 Punkte]**

Eine Unternehmensberatung untersucht die relative Effizienz der Weinproduktion in 75 kalifornischen Winzereien und legt hierfür die folgende Produktionsfunktion zugrunde:

$$y_i = \beta_1 + \beta_2 \cdot xper_i + \beta_3 \cdot cap_i + \beta_4 \cdot lab_i + e_i,$$

wobei

y_i: (Quantitäts- und Qualitäts-)Index der i-ten Winzerei

$xper_i$: Erfahrung des Managers der i-ten Winzerei

cap_i: Kapitaleinsatz der i-ten Winzerei

lab_i: Arbeitseinsatz der i-ten Winzerei

Zusätzlich steht noch Information über das Alter des Managers der i-ten Winzerei (age_i) zur Verfügung.

In der nachfolgenden Tabelle sind die Ergebnisse einer Kleinstquadrate-Schätzung (KQ), einer Hilfsregression, einer IV-Schätzung sowie einer KQ-Schätzung einer reduzierten Gleichung dargelegt:

	(1)	(2)	(3)	(4)
	KQ	Hilfsregression	IV	KQ
Abh. Variable	y	y	y	xper
Constant	1.762*	-2.487	-2.487	4.716*
	(1.06)	(2.19)	(2.72)	(2.57)
cap	0.438***	0.332***	0.332**	0.407*
	(0.12)	(0.12)	(0.15)	(0.21)
lab	0.239**	0.240**	0.240*	-0.115
	(0.100)	(0.097)	(0.12)	(0.18)
xper	0.147**	0.512***	0.512**	–
	(0.063)	(0.18)	(0.22)	
age	–	–	–	0.166***
				(0.053)
ν	–	-0.416**	–	–
		(0.19)		
Observations	75	75	75	75
R^2	0.56	0.59	–	0.17
Anmerkungen: Standardfehler in Klammern; *** p<0.01, ** p<0.05, * p<0.1				

a) Die Unternehmensberatung vermutet, dass *xper* mit *e* korreliert ist. Was wäre die Konsequenz?(2 Punkte)

b) Für den Durbin-Wu-Hausman Test wurde die Hilfsregression in Spalte (2) durchgeführt. Beschreiben Sie den Test inklusive der Null- und Alternativhypothese und führen Sie ihn durch. (4 Punkte)

c) Ist *age* ein gutes Instrument für *xper*? Diskutieren Sie kurz die beiden relevanten Aspekte. (2 Punkte)

Aufgabe 3: **[16 Punkte]**

Unterstellen Sie das ‚wahre' Modell $y_i = x_i'\beta + z_i'\gamma + \varepsilon_i$.

a) Erläutern Sie ausführlich die Auswirkungen für b, wenn statt des ‚wahren' Modells nur $y_i = x_i'\beta + \varepsilon_i$ geschätzt wird. (3 Punkte)

b) Erläutern Sie die Konsequenzen, wenn das Modell aus a) das ‚wahre' wäre, jedoch das eingangs unterstellte Modell geschätzt würde. (1 Punkte)

c) Wie können Sie allein mit Hilfe der Schätzergebnisse aus Teilaufgabe a) empirisch die funktionale Form des Modells überprüfen? Stellen Sie Ihre Vorgehensweise kurz dar. (5 Punkte)

d) Die nachfolgende Tabelle zeigt KQ Schätzergebnisse einer Lohnfunktionsgleichung, in der zunächst nur für Jahre der Bildung (*educ*) kontrolliert wird. Die Schätzungen in den Spalten (2) und (3) kontrollieren für Testergebnisse eines IQ-Tests (*IQ*) als Proxy für nicht beobachtbare Fähigkeiten sowie für einen Interaktionsterm zwischen *educ* und *IQ*. In der Stichprobe liegen die Mittelwerte (Standardabweichungen) für *educ* bei 13.48 (2.20) und für *IQ* bei 101.28 (15.05).

(7 Punkte)

Abh. Variable: log(wage)	(1)	(2)	(3)
Constant	5.395	5.176	5.648
	(0.113)	(0.128)	(0.546)
educ	0.065	0.054	0.018
	(0.006)	(0.007)	(0.041)
IQ	–	0.0036	-0.009
		(0.0010)	(0.0052)
educ*IQ	–	–	0.00034
			(0.00038)
Observations	935	935	935
R^2	0.253	0.263	0.263
Anmerkungen: Standardfehler in Klammern			

d1) Welchen Effekt hat die Berücksichtigung von *IQ* auf b_{educ}? Interpretieren Sie die Ergebnisse in Spalten (1) und (2) inhaltlich.

d2) Spalte (2): Wie hoch ist der Effekt einer Steigerung in *IQ* um eine Standardabweichung, ausgedrückt in Einheiten von *educ*?

d3) Spalte (3): Wie interpretieren Sie hier b_{educ}?

Aufgabe 4: **[8 Punkte]**

Unterstellen Sie das Modell $y_{it} = x_{it}'\beta + \alpha_i + \varepsilon_{it}$, mit i als Index für die Beobachtungseinheit i und t als Index für die Beobachtungsperiode t.

a) Durch welche Annahmen unterscheiden sich random effects und fixed effects Schätzer. (3 Punkte)

b) Ein Hausman Test prüft, ob die Ergebnisse eines random effects oder fixed effects Schätzers zu bevorzugen sind. (5 Punkte)

b1) Skizzieren Sie die Idee des Hausman Tests.

b2) Die Teststatistik eines Hausman Tests sei 2.33 bei 2 Freiheitsgraden. Testen Sie auf dem 5% Signifikanzniveau, ob der random effects oder der fixed effects Schätzer heranzuziehen ist.

Aufgabe 5: **[11 Punkte]**

Gegeben sei ein Störtermprozess mit folgender Struktur:

$$\varepsilon_t = \upsilon_t + \upsilon_{t-1} + \upsilon_{t-2},$$

wobei $\upsilon_t \sim N\left(0, \sigma_\upsilon^2\right)$ und $cov\left(\upsilon_t, \upsilon_{t-s}\right) = 0 \;\forall$ s.

a) Bestimmen Sie die Varianz von ε und die Kovarianz benachbarter Störterme ε_t und ε_{t-1}. (3 Punkte)

b) Stellen Sie die Varianz-Kovarianz-Matrix von ε_t dar. (4 Punkte)

c) Beschreiben Sie einen empirischen Zusammenhang, bei dem eine solche Fehlertermstruktur vorkommen kann. (4 Punkte)

Aufgabe 6: **[40 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.

	Das angepasste R^2 entspricht der Korrelation zwischen der abhängigen Variablen einer linearen Regressionsgleichung und ihrem vorhergesagten Wert.
	Wird beim GMM Schätzer nicht die optimale Gewichtungsmatrix W gewählt, so sind die Parameterschätzer verzerrt.
	Autokorrelierte Störtermprozesse sind dann stationär, wenn der Einfluss vergangener Schocks auf die laufenden Störterme mit der Zeit abnimmt.
	Im linearen Regressionsmodell wird unterstellt, dass die abhängige Variable keine Zufallsvariable ist.
	Bei Gültigkeit des Gauss-Markov Theorems gibt es keinen Schätzer des linearen Regressionsmodells mit kleinerer Varianz als die des Kleinstquadrateschätzers.
	Beim Least Squares Dummy Variables Schätzer (LSDV) wird ein Kleinstquadrateschätzer auf ein Modell angewendet, bei dem sowohl von der abhängigen Variable, wie von den erklärenden Variablen die beobachtungsspezifischen Mittelwerte abgezogen wurden.
	AR(1) Störterme sind homoskedastisch.
	Bei endogenen erklärenden Variablen ist eine GMM Schätzung dann effizient, wenn die Anzahl der Momentenbedingungen (R) kleiner ist als die Anzahl der zu schätzenden Parameter (K).
	Der Kleinstquadrateschätzer minimiert die Summe der quadrierten horizontalen Abweichungen von der Regressionsgerade.
	Wenn statt eines Cochrane-Orcutt Schätzers ein iterativer Cochrane-Orcutt Schätzer verwendet wird, steigt die Effizienz der Schätzung.
	Die Normalverteilung ist eine einparametrige Verteilungsfunktion.
	Der Durbin-Wu-Hausman Test auf Endogenität einer erklärenden Variablen wird durchgeführt, indem der Regressionsgleichung eine zusätzliche erklärende Variable hinzugefügt wird.
	Bei negativer Autokorrelation zweiter Ordnung ist der Durbin-Watson Test nicht durchführbar.
	Man wählt den GIVE (generalized instrumental variables estimator) Schätzer, wenn mehr endogene erklärende Variablen als Instrumente vorliegen.
	Die Nullhypothese H_0: $\beta \geq c$ wird bei 1500 Freiheitsgraden am 5 Prozentniveau verworfen, wenn als Teststatistik der t-Wert kleiner als – 1,645 ist.

	Wenn alle erklärenden Variablen x strikt exogen sind, ist der within Schätzer konsistent.
	Jede Variable z_i kann als Instrumentvariable genutzt werden, wenn Sie mit dem Störterm des Regressionsmodells unkorreliert ist.
	Die statistische Signifikanz eines Steigungsparameters lässt sich mittels eines F-Tests testen.
	Der Durbin-Watson Test verallgemeinert den White Test.
	Um k Parameter zu identifizieren, benötigt man mindestens k Momentenbedingungen.
	Der random effects Schätzer kann die between und fixed effects Schätzer an Effizienz übertreffen.
	Bei moving average Störtermprozessen kann die Korrelation zwischen länger auseinanderliegenden Störtermen Null betragen.
	Der Goldfeld-Quandt Test ist ein F-Test auf die Gleichheit der Varianz zweier Teilstichproben.
	Wenn gilt $E\{x_{i2}\varepsilon_i\} = 0$, sagen wir, dass x_{i2} eine endogene erklärende Variable ist.
	Um die Unverzerrtheit des Kleinstquadrateschätzers zu beweisen, braucht man stärkere Annahmen, als zum Nachweis seiner Konsistenz.
	Newey-West Standardfehler korrigieren sowohl für Heteroskedastie unbekannten Ursprungs als auch für Autokorrelationsmuster, die auf H Perioden beschränkt sind.
	Der F-Test auf gemeinsame Signifikanz einer Gruppe von erklärenden Variablen im Rahmen eines linearen Modells kann mittels R^2 Werten durchgeführt werden.
	White Standardfehler korrigieren für Autokorrelation beliebiger Ordnung.
	Ohne Kenntnis der Kovarianz zwischen den beiden Schätzern kann der Hausman Test zur Entscheidung zwischen fixed und random effects Schätzung nicht durchgeführt werden.
	Bei Modellen in struktureller Form ist es möglich, dass erklärende Variablen mit dem Störterm korreliert sind.
	Durch das Hinzufügen weiterer erklärender Variablen kann der angepasste R^2 Wert nicht sinken.
	Wenn der p-Wert größer ist als das Signifikanzniveau eines Tests, wird die Nullhypothese verworfen.
	Wenn heteroskedastische Störterme vorliegen, ist der Feasible GLS Schätzer BLUE.
	F-Tests lassen sich im Rahmen des linearen Modells in Wald Tests überführen, die χ^2 verteilte Teststatistiken erzeugen.
	Modelle in reduzierter Form enthalten auf der rechten Seite keine endogenen erklärenden Variablen.
	Je nachdem, ob die Schätzung mit oder ohne Konstante durchgeführt wird, spricht man bei Paneldaten von fixed oder random effects Schätzungen.
	Mithilfe des Akaike Information Criterion (AIC) und des Schwarz Bayesian Information Criterion (BIC) wird auf Strukturbruch in den Daten getestet.
	Multikollinearitätsprobleme können durch Vergrößerung der Stichprobe reduziert werden.
	Wurden Steigungsparameter mit dem verallgemeinerte Kleinstquadrateschätzer geschätzt, so ist der F Test nicht anwendbar.
	Ist eine erklärende Variable mit Messfehler gemessen, so beschreibt der KQ Schätzer ihres Steigungsparameters im linearen Modell konsistent den Effekt der gemessenen Variable.

	Die Verteilungsfunktion der Steigungsparameter des linearen Modells in kleinen Stichproben kann nur dann präzise beschrieben werden, wenn die Verteilung des Störterms feststeht.
	Gibt es verschiedene Modelle, die eine abhängige Variable erklären können, so beschreibt der encompassing Test, ob ein Modell den Erklärungsbeitrag des anderen mit enthält.
	Die Annahme $E\{\varepsilon \mid X\} = 0$ lässt zu, dass die Varianz des ε von X abhängt.
	Der nichtlineare Kleinstquadrateschätzer bestimmt diejenigen Parameter, die die quadrierte Summe der Störterme minimieren
	Auch bei exakter Multikollinearität kann der Kleinstquadrateschätzer unverzerrt geschätzt werden.
	Heteroskedastische Störterme bilden die Schocks vergangener Perioden ab.
	Ist eine erklärende Variable mit Messfehler gemessen, so sind die KQ Schätzer des Steigungsparameters des linearen Modells und die der Regressionskonstante inkonsistent.
	Ein Parameterschätzer ist effizient, wenn er gegen seinen wahren Wert konvergiert.
	Das Auslassen einer relevanten erklärenden Variablen führt zu überhöhten Standardfehlern.
	Der Chow-Test überprüft mittels einer F Teststatistik, ob vorhergesagte Werte der abhängigen Variable den Erklärungsgehalt des Modells erhöhen.
	Der verallgemeinerte Kleinstquadrate Schätzer wendet den KQ Schätzer auf transformierte Variablen an.
	Ein Vorhersageintervall für ein y_0 kann bestimmt werden, ohne den Parameterschätzer für die Regressionskonstante zu kennen.
	Enthält das lineare Regressionsmodell eine verzögerte endogene Variable (y_{t-1}), dann sollte für den Test auf Autokorrelation erster Ordnung des Störterms der Durbin-Watson Test verwendet werden.
	In der Paneldatenanalyse wird standardmäßig unterstellt, dass sich die Steigungsparameter über die Zeit ändern.
	Ist die abhängige Variable lognormal verteilt und in logarithmierter Form geschätzt, so kann ihr nicht-logarithmierter Wert nur unter Berücksichtigung der Varianz des Störterms vorhergesagt werden.
	Bei Autokorrelation erster Ordnung gilt das Gauss Markov Theorem nicht mehr.
	Die optimale Gewichtungsmatrix des GMM Modells W entspricht der Varianz-Kovarianzmatrix der Koeffizienten.
	Auch bei verzerrten Koeffizienten führt der Kleinstquadrateschätzer zu unverzerrten Vorhersagen.
	Die Annahme $V\{\varepsilon\} = \sigma^2 I$ gilt unter Autokorrelation, aber nicht unter Heteroskedastie.
	Unabhängig von den Eigenschaften des Störterms einer linearen Regressionsgleichung führt die Berücksichtigung von verzögerten endogenen Variablen (y_{t-1}) zur Inkonsistenz der geschätzten Steigungsparameter.

Aufgabe 7: **[15 Punkte]**

Wahr oder falsch? Begründen Sie Ihre Auffassung (Bsp.: "Stimmt, weil..." bzw. "Stimmt nicht, weil..."). Nur bei korrekter Begründung erhält jede richtige Antwort 1.5 Punkte; Angaben **ohne Begründung** werden **nicht gewertet**.

	Der verallgemeinerte Kleinstquadrateschätzer kann als gewichteter KQ Schätzer interpretiert werden.
	Es ist nicht möglich, im Rahmen eines linearen Regressionsmodells Elastizitäten zu schätzen.
	Die Wahrscheinlichkeit eines Typ I Fehlers ist umso höher, je höher das Signifikanzniveau α eines Tests.
	Am Signifikanzniveau von 5 Prozent weisen ein- und zweiseitige Tests einer Hypothese den gleichen kritischen Wert der Teststatistik aus.
	Die Unverzerrtheit des Kleinstquadrateschätzers lässt sich nicht nachweisen, wenn $E\{X'\varepsilon\}= 0$.
	Ob "schwache Instrumente" vorliegen, lässt sich durch eine Hilfsregression überprüfen.
	Der PE Test testet in zwei Stufen, ob das lineare oder loglineare Modell angemessen ist.
	Alle Fragestellungen der Paneldatenanalyse können auch mit Querschnittsdaten beantwortet werden.
	Identisch und unabhängig verteilte Störterme können heteroskedastisch sein.
	Das GMM Verfahren nutzt Annahmen an die Verteilungsfunktion der abhängigen Variable.