

**Prüfung im Fach Ökonometrie im SS 2012**

## Aufgabe 1 (20 Punkte)

Sei  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$  ein  $N \times 1$  Vektor und  $\mathbf{X}$  eine  $N \times K$  Matrix.

- 1.1 Beschreiben Sie formal das Minimierungsproblem zur Bestimmung von  $\hat{\boldsymbol{\beta}}_{KQ}$ . Erläutern Sie kurz verbal. (2 Punkte)
- 1.2 Es lässt sich zeigen, dass die Bedingung erster Ordnung für das Minimierungsproblem  $\mathbf{X}' \cdot [\mathbf{y} - \mathbf{X}\mathbf{b}] = \mathbf{0}$  lautet. Zeigen Sie, dass  $\mathbf{b} = \hat{\boldsymbol{\beta}}_{KQ}$  diese Bedingung erfüllt. (2 Punkte)
- 1.3 Sie interessieren sich für den Zusammenhang zwischen Monatseinkommen  $y$  und Bierkonsum  $x$ . Sie haben folgende Informationen: Person 1 verdient 3000 Euro und trinkt 3 Liter Bier pro Tag. Person 2 verdient 1000 Euro und trinkt 1 Liter Bier pro Tag. Person 3 verdient 500 Euro und trinkt 2 Liter Bier pro Tag. Person 4 verdient 3500 Euro und trinkt ebenfalls 2 Liter Bier pro Tag.
  - 1.3.1 Zeichnen Sie die Werte  $(y_i, x_i)$  in ein Koordinatensystem ein. Welcher Zusammenhang zwischen Einkommen und Bierkonsum ist erkennbar? (kurze Antwort genügt) (3 Punkte)
  - 1.3.2 Berechnen Sie für diese Stichprobe  $\hat{\boldsymbol{\beta}}_{KQ}$ . Unterstellen Sie hierbei  $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$  und verwenden Sie  $(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 2.25 & -1 \\ -1 & 0.5 \end{bmatrix}$ . Interpretieren Sie kurz den Bierkoeffizienten (Sie brauchen keine Aussage zur statistischen Signifikanz machen). (9 Punkte)
- 1.4 Sei  $N$  die Anzahl der Beobachtungen und  $Var(\boldsymbol{\varepsilon}) = \boldsymbol{\Omega} = \sigma^2\boldsymbol{\Psi} = \sigma^2(P'P)^{-1} \neq \sigma^2\mathbf{I}$ . Gehen Sie von  $E(\boldsymbol{\varepsilon}) = \mathbf{0}$  sowie nicht-stochastischen  $\mathbf{X}$  aus.
  - 1.4.1 Welche KQ-Annahme(n) könnten in diesem Fall verletzt sein? (1 Punkt)
  - 1.4.2 Führen Sie die GLS-Transformation für das Modell  $y_i = \mathbf{x}'_i\boldsymbol{\beta} + \varepsilon_i$  durch. Zeigen Sie, dass die Störterme des transformierten Modells homoskedastisch und nicht-autokorreliert sind. (3 Punkte)

## Aufgabe 2 (18.5 Punkte)

Sie interessieren sich für die Determinanten von Hauspreisen. Ihr Datensatz enthält für 506 Häuser folgende Variablen:

Variable:	Beschreibung:
<i>lnprice</i>	logarithmierter Hauspreis
<i>lnconc</i>	Feinstaub-Konzentration in der Umgebung (logarithmiert)
<i>rooms</i>	Anzahl Zimmer
<i>noise</i>	Fluglärm, gemessen in Dezibel (db)

Sie schätzen folgendes Modell mit KQ:

$$\ln price = \beta_0 + \beta_1 \ln conc + \beta_2 rooms + \beta_3 noise + \varepsilon$$

STATA liefert folgenden output:

Source	SS	df	MS	Number of obs = 506		
Model	48.7165934	3	16.2388645	F( 3, 502)	=	227.29
Residual	35.8656316	502	.071445481	Prob > F	=	0.0000
				R-squared	=	0.5760
				Adj R-squared	=	0.5734
Total	84.582225	505	.167489554	Root MSE	=	.26729

  

lnprice	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lnconc	-.6453292	.0625799	-10.31	0.000	-.76828	-.5223784
rooms	.256727	.0186769	13.75	0.000	.2200325	.2934216
noise	-.0508733	.0059261	-8.58	0.000	-.0625164	-.0392303
_cons	10.35946	.2190695	47.29	0.000	9.929052	10.78986

2.1 Interpretieren Sie den Koeffizienten von  $\ln conc$  inhaltlich und statistisch. (1.5 Punkte)

2.2 Leiten Sie für  $E(\ln y_i) = \beta \cdot \ln x_i$  den marginalen Effekt  $\frac{dE(y_i)}{dx_i}$  her (Sie dürfen  $E(\ln y_i) \approx \ln E(y_i)$  verwenden).  
Formen Sie  $\frac{dE(y_i)}{dx_i}$  weiter um, so dass  $\beta$  als Elastizität interpretierbar ist. (4 Punkte)

2.3 Sie vermuten, dass der Effekt von  $noise$  durch die ausgelassene Variable  $lowinc$  verzerrt sein könnte.  $lowinc \in [0, 1]$  gibt den Anteil an Personen mit niedrigem Einkommen in der unmittelbaren Nachbarschaft eines Hauses an.

2.3.1 Unter welcher Bedingung wäre  $\hat{\beta}_{noise}$  durch Auslassen der Variable  $lowinc$  unterschätzt? Wann wäre  $\hat{\beta}_{noise}$  unverzerrt? (5 Punkte)

2.3.2 Glücklicherweise enthält Ihr Datensatz  $lowinc$ . Wie lautet die Schätzgleichung, wenn Sie vermuten, dass der Einfluss von  $noise$  nicht konstant ist, sondern mit  $lowinc$  variiert? Geben Sie zusätzlich  $\frac{dE(\ln price_i)}{dnoise_i}$  für das erweiterte Modell an. (3 Punkte)

2.3.3 Das erweiterte Modell hat ein  $R^2$  von 0.71. Testen Sie, ob die zusätzliche(n) Variable(n) gemeinsam einen signifikanten Einfluss haben. Geben Sie hierfür Hypothesen, Teststatistik, kritischen Wert (1% Signifikanzniveau) und Testentscheidung an. (5 Punkte)

### Aufgabe 3 (9 Punkte)

Sei  $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ . Unterstellen Sie im Folgenden, dass Heteroskedastie vorliegt und die restlichen Gauss-Markov Annahmen weiterhin erfüllt sind.

3.1 Welche Konsequenzen hat Heteroskedastie in Bezug auf Erwartungstreue und geschätzte Varianz von  $\hat{\beta}_{KQ}$ ? (Es müssen keine Formeln angegeben werden) (2 Punkte)

3.2 Welche Einträge hat die Matrix  $Var(\varepsilon)$  in diesem Fall? (Gehen Sie von  $N = 3$  aus) (3 Punkte)

3.3 Sie möchten den Zusammenhang zwischen Wirtschaftswachstum und Entwicklungshilfe untersuchen. Ihre Daten enthalten für 71 Länder folgende Variablen:

Variable:	Beschreibung:
$gdpgrowth$	jährliches BIP-Wachstum (in %). Bsp.: $gdpgrowth = 4$ bedeutet 4%
$aid$	empfangene Entwicklungshilfe (in Millionen \$)
$gdpcap$	BIP pro Einwohner (in 10000 \$)
$coast$	Land hat direkten Zugang zum Meer

Sie schätzen folgendes Modell mit KQ:

$$gdpgrowth = \beta_0 + \beta_1 aid + \beta_2 gdpicap + \beta_3 coast + \varepsilon$$

STATA liefert folgenden Output:

Source	SS	df	MS			
Model	54.818528	3	18.2728427	Number of obs =	71	
Residual	73.9011368	67	1.10300204	F( 3, 67) =	16.57	
Total	128.719665	70	1.83885235	Prob > F =	0.0000	
				R-squared =	0.4259	
				Adj R-squared =	0.4002	
				Root MSE =	1.0502	

  

gdpgrowth	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
aid	.0456781	.0016207	2.82	0.006	.0013328	.0078027
gdpicap	-.978649	.7752667	-1.26	0.211	-2.526088	.5687899
coast	1.04722	.1871015	5.60	0.000	.6737636	1.420676
_cons	-.1350778	.1272177	-1.06	0.292	-.3890054	.1188498

3.3.1 Interpretieren Sie den Koeffizienten von *aid* inhaltlich. (1 Punkt)

3.3.2 Erläutern Sie, weshalb wegen *aid* ein Heteroskedastie-Problem vorliegen könnte. (2 Punkte)

3.3.3 Nennen Sie einen Test, mit dessen Hilfe auf Heteroskedastie getestet werden kann. (1 Punkt)

#### Aufgabe 4 (14 Punkte)

Für den US-Bundesstaat Texas stehen Ihnen Zeitreihendaten für  $T = 22$  Jahre zur Verfügung. Sie verwenden diese Daten, um eine Konsumfunktion zu schätzen. Die Ergebnisse des linearen Regressionsmodells

$$consumption = \beta_0 + \beta_1 income + \varepsilon$$

sind in der Tabelle ausgewiesen. Die Variablen Konsum (*consumption*) und Einkommen (*income*) sind jeweils in Mrd. US-Dollar gemessen.

Source	SS	df	MS			
Model	1045.96374	1	1045.96374	Number of obs =	22	
Residual	87.8111413	20	4.39055706	F( 1, 20) =	238.23	
Total	1133.77488	21	53.9892799	Prob > F =	0.0000	
				R-squared =	0.9225	
				Adj R-squared =	0.9187	
				Root MSE =	2.0954	

  

consumption	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
income	.6502425	.0421286	?	?	?	?
_cons	43.55203	.7762366	56.11	0.000	41.93283	45.17123

4.1 Sie interessieren sich für das Vorhersageintervall für den Konsum. Erläutern Sie verbal, welche Breite das Vorhersageintervall für Einkommenswerte annimmt, die nahe beim Minimum, Maximum sowie beim Mittelwert des Einkommens liegen. (2 Punkte)

4.2 Die Teststatistik für einen Breusch-Godfrey-Test auf Autokorrelation lautet:  $LM = (T - 2) \cdot R^2$ . Erklären Sie die Komponenten der Teststatistik. Führen Sie den Test auf Autokorrelation 2. Ordnung am 1%-Signifikanzniveau durch. Verwenden Sie dazu den folgenden Stata-Output. Geben Sie Null- und Alternativhypothese,

Freiheitsgrade, Schlusslogik und Testergebnis an. *Hinweis:* Die Variablen L1.e bzw. L2.e im Stata Output bezeichnet die verzögerten Residuen  $e_{t-1}$  bzw.  $e_{t-2}$ . (8 Punkte)

Source	SS	df	MS			
Model	379.631141	3	126.543714	Number of obs =	20	
Residual	91.8702516	16	7.06694243	F( 3, 16) =	17.91	
Total	471.501393	19	29.468837	Prob > F =	0.0001	
				R-squared =	?	
				Adj R-squared =	0.7602	
				Root MSE =	2.6584	

  

e	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
L1.	1.420216	.2274218	6.24	0.000	.9289015	1.911531
L2.	-.6480794	.2477851	-2.62	0.021	-1.183386	-.1127723
income	.2190376	.3980387	0.55	0.591	-.6408727	1.078948
_cons	-1.040122	2.290245	-0.45	0.657	-5.987897	3.907652

4.3 Erläutern Sie kurz die Idee des Cochrane-Orcutt Schätzers. Zeigen Sie formal die Vorgehensweise am Beispiel eines autoregressiven Prozesses zweiter Ordnung. D. h., der Störterm von  $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$  folgt seinen Vorläufern gemäß:  $\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + v_t$ . Dabei gilt:  $E(v_t) = 0$  und  $Var(v_t) = \sigma_v^2$ . Nehmen Sie an, dass  $\rho_1$  und  $\rho_2$  bekannt sind. Wieviele Beobachtungen stehen zur Schätzung des transformierten Modells zur Verfügung? (4 Punkte)

### Aufgabe 5: Wahr-Falsch Fragen (28.5 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.

Der ML-Schätzer nutzt keine Annahmen über die Verteilung der Störterme $\epsilon$ .	
Bei AR(1) Prozessen im Störterm sind alle Elemente der Varianz-Kovarianz Matrix des Störterms von Null verschieden.	
Ein RESET Test wird verwendet, um auf Exogenität der Regressoren zu testen.	
Der Durbin-Watson-Test verallgemeinert den Breusch-Godfrey-Test.	
Positive Autokorrelation kommt typischerweise in Querschnittsdaten vor.	
Bei Autokorrelation sind die mit den KQ-Schätzern ausgewiesenen p-Werte gültig.	
Eine Gruppe von Vektoren ist linear abhängig, wenn einer der Vektoren als Linearkombination der anderen beschrieben werden kann.	
Im linearen Modell stimmt der ML-Schätzer mit dem KQ-Schätzer für $\sigma^2$ überein.	
Das Auslassen einer relevanten erklärenden Variable aus der Schätzgleichung kann zu verzerrten und inkonsistenten KQ-Schätzern führen.	
Das angepasste $R^2$ einer Schätzung kann bei Hinzufügen von erklärenden Variablen sinken.	
Die Varianz eines unverzerrten Schätzers kann höher sein als die Varianz eines inkonsistenten Schätzers.	
Der KQ-Schätzer ist konsistent, wenn $E(\mathbf{x}_i \epsilon_i) = 0$ .	
Bei der Ableitung des KQ-Schätzers im linearen Modell erhält man so viele Normalgleichungen wie Beobachtungen vorliegen.	
Der F-Test auf gemeinsame Signifikanz einer Gruppe von erklärenden Variablen kann mittels der $R^2$ Werte durchgeführt werden.	
Das Problem der perfekten Multikollinearität kann durch Vergrößerung der Stichprobe reduziert werden.	
In ein Modell mit logarithmierter abhängiger Variable können keine Dummy-Variablen als erklärende Größen eingefügt werden, da $\ln(0) = -\infty$ .	
Bei binären Modellen mit einer erklärenden Variablen entspricht das Vorzeichen des marginalen Effekts dem Vorzeichen des Parameterschätzers.	
Der Durbin-Watson-Test eignet sich zum Testen auf Autokorrelation höherer Ordnung.	
Unter einem Interaktionsterm kann man das Produkt zweier erklärender Variablen verstehen.	
Eine starke Krümmung der Log-Likelihood-Funktion an der Stelle $\hat{\theta}$ führt zu einer unpräzisen Schätzung des Parameters $\theta$ .	
Die Standardnormalverteilung hat einen Erwartungswert von Eins.	
Der Durbin-Watson Test auf Autokorrelation ist nur anwendbar, wenn das Modell mindestens zwei Steigungsparameter enthält.	
Mithilfe eines linearen Regressionsmodells lassen sich Elastizitäten schätzen.	
Die Inverse einer Matrix hat die gleiche Dimension wie die Ausgangsmatrix.	
Bei einer Regression ohne Konstante $\beta_0$ verläuft die Regressionsgerade durch den Koordinatenursprung.	
Die Annahme $E(\epsilon_t   \mathbf{x}_t) = 0$ lässt zu, dass $Cov(\epsilon_t, \epsilon_{t-1}) > 0$ ist.	
Die Parameter $\beta_1$ und $\beta_2$ des Modells $y_i = \beta_0 x_{1i}^{\beta_1} x_{2i}^{\beta_2} e^{\epsilon_i}$ können nicht mit dem KQ-Verfahren geschätzt werden.	
Der RESET-Test dient dazu, die unterstellte funktionale Form der Regressionsgleichung zu überprüfen.	
Je größer der Typ II Fehler eines Tests, umso größer muss der Typ I Fehler sein.	
Heteroskedastische Störterme haben eine Varianz von 1.	

Die im Rahmen des ML-Verfahrens angewendeten Testprinzipien erlauben keinen Test von Linearkombinationen der Parameter.	
Enthält das lineare Regressionsmodell eine verzögerte endogene Variable ( $y_{t-1}$ ), dann kann für den Test auf Autokorrelation erster Ordnung des Störterms der Breusch-Godfrey-Test verwendet werden.	
Wald-, Likelihood-Ratio- und Lagrange-Multiplier-Tests sind asymptotisch äquivalent.	
Die vorhergesagten Wahrscheinlichkeiten eines Probit-Modells liegen immer im Intervall $[0; 1]$ .	
Die Prais-Winsten-Schätzer müssen vor der Interpretation transformiert werden.	
Wenn statt eines Cochrane-Orcutt-Schätzers ein Prais-Winsten-Schätzer verwendet wird, steigt die Effizienz der Schätzung.	
Der Durbin-Watson Test verallgemeinert den White Test.	
Der ML-Schätzer für die Varianz des Störterms im linearen Modell ist verzerrt.	