

Prüfung im Fach Ökonometrie im SS 2011 – Aufgabenteil

Name, Vorname	
Matrikelnr.	
Studiengang	
E-Mail-Adresse	
Unterschrift	

Vorbemerkungen:

Anzahl der Aufgaben:

- Die Klausur besteht aus 5 Aufgaben.

Bewertung:

- Die Prüfungsdauer beträgt für alle Studierenden 90 Minuten, es können maximal 90 Punkte erworben werden.
- Die Punktzahl für jede Aufgabe ist in Klammern angegeben.

Erlaubte Hilfsmittel:

- Tabellen der statistischen Verteilungen und Formelsammlung (sind der Klausur beigelegt)
- Taschenrechner
- Fremdwörterbuch

Wichtige Hinweise:

- Sollte es vorkommen, dass die statistischen Tabellen, die dieser Klausur beiliegen, den exakten Wert der gesuchten Freiheitsgrade nicht ausweisen, machen Sie dies kenntlich und verwenden Sie den nächstgelegenen Wert.
- Sollte es vorkommen, dass bei einer Berechnung eine erforderliche Annahme oder Angabe fehlt, machen Sie dies kenntlich und treffen Sie für den fehlenden Wert eine plausible Annahme.
- Aufgabe 5 ist im Aufgabenteil zu beantworten, die restlichen Aufgaben im Lösungsteil. Verwenden Sie hierbei für jede Aufgabe ein neues Blatt.

Aufgabe 1 (17.5 Punkte)

Mit einer Regression wird die durchschnittliche (logarithmierte) Entfernung zwischen Wohnung und Arbeitsstätte in Abhängigkeit von erklärenden Merkmalen geschätzt. Tabelle 1 zeigt die deskriptiven Statistiken und eine Beschreibung der Variablen. Die Regressionsergebnissen sind in Tabelle 2 ausgewiesen.

Tabelle 1: Deskriptive Statistiken

Variable	Mittelwert	Std. Abw.	Min.	Max.	Beschreibung
ln_commdist	2.25	1.26	0	6.91	Entfernung zw. Wohnung und Arbeit in km (logarithmiert)
ln_hwage	2.53	0.61	-0.06	4.06	Stundenlohn (logarithmiert)
female	0.50	0.50	0	1	=1, falls Frau; =0 sonst
ln_hwageXfemale	1.21	1.27	-0.06	4.06	Interaktionsterm: ln_hwage×female
educ	12.75	2.74	7	18	Dauer der Ausbildung (in Jahren)
tenure	11.49	10.4	0	52.7	Dauer der Betriebszugehörigkeit (in Jahren)
tenuresq	240	366	0	2777	Dauer der Betriebszugehörigkeit (quadriert)

Tabelle 2: Regressionsergebnisse

Source	SS	df	MS	Number of obs = 9510		
Model	912.560347	6	152.093391	F(6, 9503) =	102.18	
Residual	14144.7644	9503	1.48845253	Prob > F =	0.0000	
				R-squared =		
				Adj R-squared =		
Total	15057.3248	9509	1.58348141	Root MSE =	1.22	

ln_commdist	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ln_hwage	.2781026	.0331323	8.39	0.000	.2131562	.343049
female	-.7213241	.1097058	-6.58	0.000	-.9363709	-.5062773
ln_hwageXfemale	.1342272	.0422104	3.18	0.001	.0514858	.2169685
educ	.0155981	.0050494	3.09	0.002	.0057002	.025496
tenure	-.0242696	.0038524	-6.30	0.000	-.0318212	-.016718
tenuresq	.0003455	.0001059	3.26	0.001	.0001379	.0005531
_cons	1.746348	.0885472	19.72	0.000	1.572776	1.919919

- 1.1 Betrachten Sie die Schätzgleichung insgesamt und interpretieren Sie den Zusammenhang zwischen Entfernung und Stundenlohn inhaltlich. (2.5 Punkte)
- 1.2 Berechnen und interpretieren Sie den Wert für das Bestimmtheitsmaß. (2 Punkte)
- 1.3 Tabelle 3 enthält Ergebnisse der Regression $y_i = \mathbf{x}_i' \boldsymbol{\beta} + \alpha_2 \hat{y}_i^2 + \alpha_3 \hat{y}_i^3 + v_i$, wobei y_i die Variable *ln_commdist* bezeichnet, \hat{y}_i ist ihr auf Basis der Schätzergebnisse in Tabelle 2 vorhergesagter Wert. Der Vektor \mathbf{x}_i enthält erklärende Variablen, v_i ist ein Störterm. Verwenden Sie die Regressionsergebnisse in Tabelle 3 für einen RESET-Test bei einem Signifikanzniveau von 5%. Geben Sie Null- und Alternativhypothesen, Teststatistik, Entscheidungsregel und Testentscheidung an. (8.5 Punkte)

Tabelle 3: Regressionsergebnisse

ln_commdist	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ln_hwage	-1.807968	.6703899	-2.70	0.007	-3.122076	-.4938608
female	4.935471	1.751507	2.82	0.005	1.502142	8.368799
ln_hwageXfemale	-.9429534	.3266221	-2.89	0.004	-1.583203	-.3027043
educ	-.1019698	.0382517	-2.67	0.008	-.1769513	-.0269883
tenure	.1593845	.0586165	2.72	0.007	.0444837	.2742854
tenuresq	-.0022692	.0008391	-2.70	0.007	-.003914	-.0006245
yhat_2	3.834189	1.142614	3.36	0.001	1.594421	6.073957
yhat_3	-.6239351	.1783697	-3.50	0.000	-.9735778	-.2742924
_cons	-6.793467	2.567201	-2.65	0.008	-11.82573	-1.761204

Für $\alpha = (\alpha_2, \alpha_3)'$ lautet die geschätzte Varianz-Kovarianz-Matrix $\widehat{\mathbf{V}}(\hat{\alpha})$ bzw. ihre Inverse $\widehat{\mathbf{V}}(\hat{\alpha})^{-1}$:

$$\widehat{\mathbf{V}}(\hat{\alpha}) = \begin{pmatrix} 1.306 & \\ -0.203 & 0.032 \end{pmatrix}, \quad \widehat{\mathbf{V}}(\hat{\alpha})^{-1} = \begin{pmatrix} 63.782 & \\ 406.12 & 2617.30 \end{pmatrix}$$

1.4 Überprüfen Sie anhand der Regressionsergebnisse in Tabelle 4 mit einem Goldfeld-Quandt-Test bei einem Signifikanzniveau von 5%, ob die Varianz der Störgröße für Männer größer ist als für Frauen. Geben Sie Hypothesen, Teststatistik, Entscheidungsregel, Testentscheidung und Interpretation des Ergebnisses an. (4.5 Punkte)

Tabelle 4: Regressionsergebnisse der nach Geschlecht getrennten Schätzungen

. reg ln_commdist ln_hwage educ tenure tenuresq if female == 0

Source	SS	df	MS	Number of obs = 4735		
Model	180.216704	4	45.0541761	F(4, 4730) =	28.59	
Residual	7453.502	4730	1.57579323	Prob > F =	0.0000	
				R-squared =	0.0236	
				Adj R-squared =	0.0228	
Total	7633.71871	4734	1.61253036	Root MSE =	1.2553	

ln_commdist	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ln_hwage	.3389154	.0375557	9.02	0.000	.2652889	.412542
educ	-.0054693	.0072344	-0.76	0.450	-.0196522	.0087135
tenure	-.029778	.0055148	-5.40	0.000	-.0405896	-.0189664
tenuresq	.0004404	.000146	3.02	0.003	.0001541	.0007267
_cons	1.896735	.0985583	19.24	0.000	1.703515	2.089955

. reg ln_commdist ln_hwage educ tenure tenuresq if female == 1

Source	SS	df	MS	Number of obs = 4775		
Model	304.363798	4	76.0909496	F(4, 4770) =	54.48	
Residual	6661.57318	4770	1.39655622	Prob > F =	0.0000	
				R-squared =	0.0437	
				Adj R-squared =	0.0429	
Total	6965.93698	4774	1.45914055	Root MSE =	1.1818	

ln_commdist	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ln_hwage	.3562289	.0348714	10.22	0.000	.2878648	.424593
educ	.0379863	.0070361	5.40	0.000	.0241923	.0517802
tenure	-.0185156	.00544	-3.40	0.001	-.0291806	-.0078506
tenuresq	.0002364	.0001571	1.50	0.132	-.0000716	.0005443
_cons	.8369844	.0947656	8.83	0.000	.6512001	1.022769

Aufgabe 2 (14 Punkte)

Die Determinanten der Partizipation von Frauen am Arbeitsmarkt werden mit einem binären Logit-Modell geschätzt. Tabelle 5 enthält die deskriptiven Statistiken und eine Beschreibung der Variablen. Die Regressionsergebnisse sind in Tabelle 6 ausgewiesen.

Tabelle 5: Deskriptive Statistiken

Variable	Mittelwert	Std. Abw.	Min.	Max.	Beschreibung
inlf	0.57	0.5	0	1	=1, falls Frau erwerbstätig ist, =0 sonst.
otherinc	20.13	11.63	-0.03	96	Sonstiges Einkommen der Frau (in 1000 USD pro Jahr)
exper	10.63	8.07	0	45	Berufserfahrung in Jahren
expersq	178.04	249.63	0	2025	Berufserfahrung (quadriert)
kidslt6	0.24	0.52	0	3	Anzahl der Kinder jünger als 6 Jahre
kidsge6	1.35	1.32	0	8	Anzahl der Kinder mit 6 Jahren oder älter
motheduc	9.25	3.37	0	17	Ausbildung der Mutter (in Jahren)
fatheduc	8.81	3.57	0	17	Ausbildung des Vaters (in Jahren)

Tabelle 6: Regressionsergebnisse für binäres Logit-Modell

Logistic regression	Number of obs	=	753			
	LR chi2(5)	=	147.53			
	Prob > chi2	=	0.0000			
Log likelihood = -441.10884	Pseudo R2	=	0.1433			

inlf	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
otherinc	-.0138466	.0073218	-1.89	0.059	-.0281971	.000504
exper	.2217742	.0299974	7.39	0.000	.1629803	.280568
expersq	-.0043733	.0009582	-4.56	0.000	-.0062514	-.0024953
kidslt6	-.7245464	.1650951	-4.39	0.000	-1.048127	-.4009659
kidsge6	.188833	.0663582	2.85	0.004	.0587733	.3188928
_cons	-1.068482	.277402	-3.85	0.000	-1.61218	-.5247842

- 2.1 Bestimmen und interpretieren Sie den marginalen Effekt für ein weiteres Jahr Berufserfahrung bei einer bestehenden Berufserfahrung von 10 Jahren. Unterstellen Sie Stichprobenmittelwerte der anderen Kovariaten. *Hinweis:* Für den marginalen Effekt im Logit-Modell gilt allgemein: $\partial E(y_i | \mathbf{x}_i) / \partial x_{ik} = \exp(\mathbf{x}'_i \boldsymbol{\beta}) / (1 + \exp(\mathbf{x}'_i \boldsymbol{\beta}))^2 \beta_k$. (5 Punkte)
- 2.2 Testen Sie am 10% Signifikanzniveau, ob sich der Erklärungsgehalt signifikant verbessert, wenn die Bildungsvariablen der Eltern (*motheduc* und *fatheduc*) als weitere Kovariablen in das Modell aufgenommen werden. Die Regressionsergebnisse des erweiterten Modells finden sich in Tabelle 7. Geben Sie Hypothesen, Teststatistik, Entscheidungsregel und Testentscheidung an. (4.5 Punkte)

Tabelle 7: Regressionsergebnisse für Modellerweiterung

Logistic regression	Number of obs =	753
	LR chi2(7) =	164.90
	Prob > chi2 =	0.0000
Log likelihood = -432.42332	Pseudo R2 =	0.1601

inlf	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
otherinc	-.0187896	.0075904	-2.48	0.013	-.0336666 - .0039127
exper	.2204833	.0304128	7.25	0.000	.1608752 .2800913
expersq	-.0041626	.0009732	-4.28	0.000	-.0060701 -.0022552
kidslt6	-.819102	.1702708	-4.81	0.000	-1.152827 -.4853773
kidsge6	.2009833	.0676692	2.97	0.003	.0683541 .3336125
motheduc	.0815068	.0305369	2.67	0.008	.0216557 .141358
fatheduc	.033261	.0286873	1.16	0.246	-.022965 .0894871
_cons	-2.02696	.3700785	-5.48	0.000	-2.752301 -1.30162

2.3 Erläutern Sie, wie man die Hypothese aus Aufgabe 2.2 nur mit den Schätzergebnissen aus Tabelle 6 überprüfen kann (d. h. ohne Verwendung der Regressionsergebnisse aus Tabelle 7). *Hinweis:* Sie müssen den Test nicht durchführen. (3 Punkte)

2.4 Nennen Sie drei Eigenschaften des ML-Schätzers, die sich bei korrekt spezifizierter Likelihoodfunktion zeigen lassen. (1.5 Punkte)

Aufgabe 3 (18 Punkte)

Sie interessieren sich für den Zusammenhang zwischen dem Konsum von Sandspielzeug für Kinder und dem durchschnittlichen Familieneinkommen. Ihnen liegen dazu deutsche Daten für 32 Sommer- und Winterhalbjahre vor. Ihr Datensatz enthält folgende Variablen:

- C_t : Konsumausgaben für Sandspielzeug in 1000 EUR
- W_t : Durchschnittliches Familieneinkommen
- O_t : Anteil der Bevölkerung, der an der Küste wohnt
- t : Halbjahresindex von 1995(H1): 1 bis 2010(H2): 32

Sie schätzen folgendes Modell in Stata:

$$C_t = \beta_0 + \beta_1 \ln W_t + \beta_2 O_t + \varepsilon_t$$

Tabelle 8: Regressionsergebnisse

Source	SS	df	MS	Number of obs = 32		
Model	.096863744	2	.048431872	F(2, 29) =	94.52	
Residual	.014860225	29	.000512422	Prob > F =	0.0000	
Total	.111723969	31	.003603999	R-squared =	0.8670	
				Adj R-squared =	0.8578	
				Root MSE =	.02264	

C	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lnW	.8992026	.0660467	13.61	0.000	.7641219 1.034283
O	-.0585477	.0748908	-0.78	0.441	-.2117165 .0946211
_cons	.5075768	.5914838	0.86	0.398	-.7021434 1.717297

3.1 Interpretieren Sie inhaltlich und statistisch den Zusammenhang zwischen Familieneinkommen und dem Konsum von Sandspielzeug. (1.5 Punkte)

- 3.2 Ihr Kollege vermutet, dass in Ihrem Modell Autokorrelation vorliegt. Erläutern Sie den Begriff. Welche Folgen hat ein Vorliegen von Autokorrelation für Ihre Ergebnisse? (2.5 Punkte)
- 3.3 Begründen Sie anhand eines Beispiels, warum im vorliegenden Fall Autokorrelation vorliegen könnte. (2.5 Punkte)
- 3.4 Beschreiben Sie die Testidee des Breusch-Godfrey-Tests auf Autokorrelation 1. Ordnung. Nennen Sie Hypothesen, Teststatistik und Schlusslogik. Skizzieren Sie knapp die Durchführung des Tests anhand des vorliegenden Modells. (5 Punkte)
- 3.5 Erläutern Sie die Idee des Cochrane-Orcutt Schätzers. Zeigen Sie formal die Vorgehensweise. Setzen Sie voraus, dass $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$ gilt und ρ bekannt ist. Wodurch unterscheidet sich dieses Verfahren vom Prais-Winsten Schätzer und welches Verfahren würden Sie im vorliegenden Fall bevorzugen? Begründen Sie kurz. (5 Punkte)
- 3.6 In den meisten Fällen ist ρ unbekannt. Können Sie trotzdem die Verfahren aus 3.5 anwenden? Begründen Sie Ihre Antwort. (1.5 Punkte)

Aufgabe 4 (12 Punkte)

- 4.1 Erläutern Sie, was die Aussage "Der KQ-Schätzer ist unverzerrt" bedeutet. (1 Punkt)
- 4.2 Erläutern Sie den Unterschied zwischen Unverzerrtheit und BLUE-Eigenschaft des KQ-Schätzers. (2 Punkte)
- 4.3 Zeigen Sie formal die Unverzerrtheit des KQ-Schätzers. Nehmen Sie an, dass die Gauss-Markov Annahmen gelten. (3 Punkte)
- 4.4 Welche Annahmen benötigen Sie, um diese Eigenschaft herzuleiten? Benennen Sie diese konkret und erläutern Sie an welcher Stelle der Herleitung Sie sie anwenden. (2 Punkte)
- 4.5 Welche Auswirkungen ergeben sich bei Vorliegen von heteroskedastischen Störtermen für die Unverzerrtheit des KQ-Schätzers. Erläutern Sie Ihre Antwort. (2 Punkte)
- 4.6 Welche Konsequenzen ergeben sich bei heteroskedastischen Störtermen für die BLUE Eigenschaft? Erläutern Sie Ihre Antwort knapp. (2 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.

Die Normalverteilung ist eine zweiparametrische Verteilungsfunktion.	
Mit steigender Zahl von Freiheitsgraden konvergiert die t-Verteilung zur F-Verteilung.	
Ein Spaltenvektor ergibt sich als die Quadratwurzel eines Zeilenvektors.	
Das angepasste R^2 einer Schätzung kann bei Hinzufügen von erklärenden Variablen nicht sinken.	
Insignifikante Koeffizienten können größer als 100 sein.	
Die Nullhypothese $\beta \leq \beta_k$ wird am 5 Prozent Signifikanzniveau bei 500 Freiheitsgraden verworfen, wenn der empirische t-Wert größer als 1.96 ist.	
Die statistische Signifikanz eines Steigungsparameters lässt sich mittels eines F-Tests testen.	

Um die Unverzerrtheit des Kleinstquadrateschätzers zu beweisen, braucht man stärkere Annahmen, als zum Nachweis seiner Konsistenz.	
Die Varianz eines unverzerrten Schätzers kann höher sein als die Varianz eines inkonsistenten Schätzers.	
An der Stichprobenmitte ist der Vorhersagefehler des KQ-Schätzers am kleinsten.	
Der KQ Schätzer ist konsistent, wenn erklärende Variablen und Störterm unkorreliert sind.	
Die Werte der t-Verteilung sind symmetrisch um den Wert 1.96 verteilt.	
Die Breite des Vorhersageintervalls für die abhängige Variable variiert mit der Streuung des Störterms.	
Bei der Ableitung des KQ-Schätzers im linearen Modell erhält man so viele Normalgleichungen wie Beobachtungen vorliegen.	
Unter einem Interaktionsterm versteht man das Produkt zweier erklärender Variablen.	
Der Chow-Test prüft, ob die lineare oder loglineare Version der abhängigen Variable dem Modell besser entspricht.	
Mithilfe eines linearen Regressionsmodells lassen sich Elastizitäten schätzen.	
Mithilfe des Akaike Information Criterion (AIC) und des Schwarz Bayesian Information Criterion (BIC) wird auf Strukturbruch in den Daten getestet.	
Der F-Test auf gemeinsame Signifikanz einer Gruppe von erklärenden Variablen kann mittels der R^2 Werte durchgeführt werden.	
Bei Autokorrelation in Form von moving average Störprozessen gibt es Fehlerterme, die nicht miteinander korreliert sind.	
Newey-West Standardfehler korrigieren für Heteroskedastie unbekanntem Ursprungs ebenso wie für Autokorrelation.	
Wenn irrelevante erklärende Variablen berücksichtigt werden, dann sinkt die Varianz des KQ-Schätzers.	
Wenn heteroskedastische Störterme vorliegen, ist jeder Feasible GLS Schätzer BLUE.	
Das Problem der perfekten Multikollinearität kann durch Vergrößerung der Stichprobe reduziert werden.	
Ein Parameterschätzer ist konsistent, wenn er die kleinste Varianz hat.	
Die Annahme $\varepsilon_i \sim i.i.d(0, \sigma^2)$ schließt sowohl Heteroskedastie als auch Autokorrelation aus.	
Homoskedastische Störterme haben eine nicht konstante Varianz.	
Wenn man eine konkrete Form der Heteroskedastie unterstellt, kann man FGLS-Schätzer verwenden.	
Auf Basis linearer Modelle geschätzte Koeffizienten können immer als Kausaleffekte interpretiert werden.	
Autokorrelation führt zu Verzerrtheit des KQ-Schätzers.	
Der Durbin-Watson-Test ist bei negativer Autokorrelation nicht anwendbar.	
Der Breusch-Pagan-Test ist eine Verallgemeinerung des White-Tests.	
Berücksichtigt man im Modell irrelevante erklärende Variablen, so sinkt die Streuung der geschätzten Parameter.	
In ein Modell mit logarithmierter abhängiger Variable können keine Dummy-Variablen als erklärende Größen eingefügt werden, da $\ln(0) = -\infty$.	
Die Likelihoodfunktion wird typischerweise in logarithmierter Form maximiert.	
Die χ^2 -Verteilung ist symmetrisch.	
Die Wahrscheinlichkeit eines Typ-II-Fehlers ist umso höher, je höher das Signifikanzniveau α eines Tests ist.	
Der Kleinstquadrateschätzer minimiert die Summe der quadrierten horizontalen Abweichungen von der Regressionsgerade.	