

Prüfung im Fach Ökonometrie im WS 2011/12 – 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 (11.5 Punkte)

Sie verfügen über einen Datensatz, der Informationen zu den jährlichen individuellen Gesundheitsausgaben und dem sozio-ökonomischen Hintergrund von Individuen enthält. Sie stellen folgendes Modell der logarithmierten jährlichen Gesundheitsausgaben auf:

$$\ln y_i = \beta_0 + \beta_1 \text{private}_i + \beta_2 \text{totchr}_i + \beta_3 \text{female}_i + \varepsilon_i$$

Die Ergebnisse der KQ-Schätzung sind in Tabelle 2 abgebildet.

Tabelle 1: Beschreibung der Variablen

ln_y	Jährliche Gesundheitsausgaben in US-Dollar (logarithmiert)
private	Private Krankenversicherung (privat=1, nicht-privat=0)
totchr	Anzahl der chronischen Erkrankungen
female	Frau (Frau=1, Mann=0)

Tabelle 2: Regressionsergebnisse

Source	SS	df	MS			
Model	?	3	353.369031	Number of obs =	2955	
Residual	?	2951	1.51297265	F(3, 2951) =	233.56	
Total	5524.88938	2954	1.87030785	Prob > F	= 0.0000	
				R-squared	= 0.1919	
				Adj R-squared	= 0.1911	
				Root MSE	= 1.2300	

ln_y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
private	.2420292	.0461862	5.24	0.000	.1514687	.3325897
totchr	.4526897	.0175108	25.85	0.000	.418355	.4870244
female	-.0632034	.0461299	-1.37	0.171	-.1536533	.0272466
_cons	7.134789	.0545876	130.70	0.000	7.027755	7.241822

- 1.1 Interpretieren Sie statistisch und inhaltlich den Zusammenhang zwischen privater Krankenversicherung (*private*) und den jährlichen Gesundheitsausgaben. (1.5 Punkte)
- 1.2 Berechnen Sie die Summe der quadrierten erklärten Abweichungen, die Summe der quadrierten nicht-erklärten Abweichungen, sowie die geschätzte Varianz des Störterms. (4 Punkte)
- 1.3 Nennen Sie zwei Änderungen der Schätzergebnisse, wenn Sie anstatt der Dummy-Variable *female* die Dummy-Variable *male* (=1 falls Mann, 0 sonst) verwenden. *Hinweis*: Es ist keine Berechnung notwendig. (1 Punkt)
- 1.4 Überprüfen Sie für das Modell, ob sich die Koeffizienten der Variablen *totchr* und *female* für privat bzw. nicht-privat versicherte Personen unterscheiden. Führen Sie einen geeigneten Test durch. Geben Sie Null- und Alternativhypothesen, Teststatistik (inkl. Schätzgleichung des unrestringierten Modells), kritischen Wert und Testentscheidung am 5% Signifikanzniveau an. *Hinweis*: Das Bestimmtheitsmaß des unrestringierten Modells hat den Wert $R^2 = 0.1932$. (5 Punkte)

Aufgabe 2 (13.5 Punkte)

Mit Zeitreihendaten wird der durchschnittliche tägliche Umsatz in einer Cafeteria in Abhängigkeit von erklärenden Merkmalen geschätzt. Die Cafeteria ist von Montag bis Freitag geöffnet. Tabelle 3 zeigt die deskriptiven Statistiken und eine Beschreibung der Variablen. Die Regressionsergebnisse sind in Tabelle 4 ausgewiesen.

Tabelle 3: Deskriptive Statistiken

Variable	Mittelwert	Std. Abw.	Min.	Max.	Beschreibung
sales	147.41	46.24	61.94	240.87	täglicher Umsatz in US-Dollar
Tue	0.20	0.40	0	1	=1, falls Dienstag, 0 sonst.
Wed	0.20	0.40	0	1	=1, falls Mittwoch, 0 sonst.
Thu	0.22	0.42	0	1	=1, falls Donnerstag, 0 sonst.
Fri	0.20	0.40	0	1	=1, falls Freitag, 0 sonst.
temp	43.1	15.07	20	80	Temperatur in Fahrenheit

Tabelle 4: Regressionsergebnisse

Source	SS	df	MS			
Model	63580.4361	5	12716.0872	Number of obs =	40	
Residual	19799.3588	34	582.334082	F(5, 34) =	21.84	
Total	83379.7949	39	2137.94346	Prob > F =	0.0000	
				R-squared =	0.7625	
				Adj R-squared =	0.7276	
				Root MSE =	24.132	

sales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Tue	71.05403	12.10972	5.87	0.000	46.44411	95.66395
Wed	19.98817	12.13436	1.65	0.109	-4.671818	44.64816
Thu	59.53434	11.81593	5.04	0.000	35.52148	83.54721
Fri	-42.49019	12.49265	-3.40	0.002	-67.87831	-17.10206
temp	-.5841342	.2576219	-2.27	0.030	-1.107685	-.0605834
_cons	148.3571	13.40325	11.07	0.000	121.1185	175.5958

- 2.1 Interpretieren Sie den Koeffizienten der Variable Fri inhaltlich und statistisch. (1.5 Punkte)
- 2.2 Erklären Sie kurz verbal, was inhaltlich unter positiver Autokorrelation zu verstehen ist. *Hinweis:* Es ist keine formale Darstellung notwendig. (1 Punkt)
- 2.3 Erläutern Sie allgemein die kritischen Werte und die möglichen Testentscheidungen beim Durbin-Watson-Test auf positive Autokorrelation. (4 Punkte)
- 2.4 Führen Sie einen Durbin-Watson-Test auf positive Autokorrelation durch. Geben Sie Null- und Alternativhypothesen, Entscheidungsregel mit kritischen Werten und Testergebnis an. *Hinweis:* Der empirische Wert der Teststatistik beträgt $d_w = 1.60$. (4 Punkte)
- 2.5 Wie unterscheiden sich der Moving-Average-Prozess $\varepsilon_t = v_t + v_{t-1}$ und der autoregressive Prozess erster Ordnung $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$ in Bezug auf Beobachtungen, die mehr als eine Periode auseinanderliegen? *Hinweis:* Für v_t gelten die Gauß-Markov-Annahmen. (2 Punkte)
- 2.6 Nennen Sie zwei mögliche Ursachen von Autokorrelation. (1 Punkt)

Aufgabe 3 (16 Punkte)

Der Zusammenhang zwischen Job-Zufriedenheit und gewünschter Arbeitszeit wird mit einem binären Logit-Modell und einer unabhängig gezogenen Stichprobe von $N = 732$ Individuen 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
jobsat	0.07	0.25	0	1	=1, falls Person mit ihrem Job zufrieden ist, =0 sonst.
age	40.87	6.15	21	57	Alter in Jahren
mismatch	0.68	0.47	0	1	=1, falls gewünschte von tatsächlicher Arbeitszeit abweicht, =0 sonst.

Tabelle 6: Regressionsergebnisse für binäres Logit-Modell

Logistic regression	Number of obs	=	732
	LR chi2(2)	=	?
	Prob > chi2	=	0.0005
Log likelihood = -172.23716	Pseudo R2	=	0.0421

jobsat	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
age	-.0441807	.0241626	-1.83	0.067	-.0915385 .0031772
mismatch	-1.024564	.2998061	-3.42	0.001	-1.612173 -.4369546
_cons	-.2725406	.9748076	-0.28	0.780	-2.183128 1.638047

3.1 Erläutern Sie den folgenden Ausdruck, der der Schätzung des Modells zugrunde liegt. Erklären Sie dabei die Komponenten A, B und C. (3 Punkte)

$$L(\beta) = \underbrace{\prod_{i=1}^N}_{A} \underbrace{P(y_i = 1|\mathbf{x})^{y_i} P(y_i = 0|\mathbf{x})^{1-y_i}}_C$$

3.2 Bestimmen und interpretieren Sie den marginalen Effekt des Alters. Unterstellen Sie Stichprobenmittelwerte der erklärenden Variablen. (4 Punkte)

3.3 Berechnen und interpretieren Sie die Wahrscheinlichkeit, dass eine Person mit ihrem Job zufrieden ist, wenn die gewünschte und tatsächliche Arbeitszeit übereinstimmen. Unterstellen Sie ein Alter von 30 Jahren. *Hinweis:* Es gilt, dass $P(y_i = 1|\mathbf{x}_i) = F(\mathbf{x}'_i\beta) = \frac{\exp\{\mathbf{x}'_i\beta\}}{1+\exp\{\mathbf{x}'_i\beta\}} = \frac{1}{1+\exp\{-\mathbf{x}'_i\beta\}}$. (3 Punkte)

3.4 Beurteilen Sie die Signifikanz des Modells anhand eines Likelihood-Ratio-Tests am 1% Niveau. Geben Sie Null- und Alternativhypothesen, Entscheidungsregel mit kritischem Wert und Testergebnis an. *Hinweis:* Ein Modell, das nur mit einer Konstanten geschätzt wurde, liefert einen Wert der Log-Likelihood-Funktion von -179.82. (4 Punkte)

3.5 Benennen und erläutern Sie knapp die Vorgehensweise bei einem asymptotisch äquivalenten Test, mit dem man die Signifikanz des Modells (d.h. die Hypothese aus Aufgabe 3.4) ebenfalls überprüfen kann. *Hinweis:* Sie müssen den Test nicht durchführen. (2 Punkte)

Aufgabe 4 (19 Punkte)

4.1 Wie lautet die formale Definition der Unverzerrtheit? (1 Punkt)

4.2 Leiten Sie die Unverzerrtheit des KQ-Schätzers in Matrixschreibweise her. Machen Sie kenntlich, an welcher Stelle der Herleitung welche Annahme (vgl. Formelsammlung) verwendet wird. (3 Punkte)

4.3 Was ändert sich bei der Lösung von Aufgabe 4.2, wenn \mathbf{X} deterministisch ist. (2 Punkte)

4.4 Sei $Var[\boldsymbol{\epsilon}] = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ eine 2×2 Matrix. Welche Einträge a_{ij} in $Var[\boldsymbol{\epsilon}]$ nehmen gleiche Werte an, falls

- (a) die Gauß-Markov-Annahmen gelten,
- (b) Heteroskedastizität vorliegt,
- (c) Heteroskedastizität und Autokorrelation vorliegen,
- (d) die ϵ_i durch einen AR(1)-Prozess erzeugt wurden?

(4 Punkte)

4.5 Ihnen liegen folgende Daten vor:

i	y_i	x_{i1}	x_{i2}
1	2	1	-1
2	3	1	0
3	4	1	1

Sie schätzen das Modell $y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$. Berechnen Sie \mathbf{b} mittels KQ-Verfahren. *Hinweis:* $(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$. (4 Punkte)

4.6 Berechnen Sie $\widehat{Var}[\mathbf{b}]$ unter Angabe des Rechenwegs. Unterstellen Sie hierbei, dass Annahmen A1 bis A4 erfüllt sind. Erläutern Sie kurz Ihr Ergebnis. *Hinweis:* Verwenden Sie $\mathbf{b} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$, falls Sie die vorherige Teilaufgabe nicht gelöst haben. (5 Punkte)

Aufgabe 5: Wahr-Falsch Fragen (30 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.

Ein omitted variable bias liegt vor, wenn eine ausgelassene erklärende Variable mit der abhängigen Variablen, nicht jedoch mit einer anderen erklärenden Variable korreliert ist.	
Die Informationsmatrix wird für die Schätzung der Varianz-Kovarianz-Matrix des ML-Schätzers verwendet.	
Zur Durchführung eines Lagrange-Multiplier-Tests ist die zweimalige Schätzung des Modells (d.h. mit bzw. ohne Restriktion) notwendig.	
Die Parameter eines FGLS-Modells müssen wie die Parameter des Original-Modells interpretiert werden.	
Mit steigender Zahl von Freiheitsgraden konvergiert die t-Verteilung zur F-Verteilung.	
Quadratische Matrizen mit vollem Spaltenrang sind invertierbar.	
Die Normalverteilung ist eine einparametrische Verteilungsfunktion.	
Das Produkt aus einem Spaltenvektor und einem Zeilenvektor ist ein Skalar.	
Bei negativer Autokorrelation ist der Durbin-Watson Test nicht durchführbar.	
Ein Parameterschätzer ist effizient, wenn er gegen seinen wahren Wert konvergiert.	
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.	
Der verallgemeinerte Kleinstquadrateschätzer kann als gewichteter KQ Schätzer interpretiert werden.	

Maximum-Likelihood-Schätzer benötigen eine Annahme bezüglich der Verteilung der abhängigen Variable.	
Autokorrelation führt zu Inkonsistenz und Verzerrtheit des KQ-Schätzers.	
Die geschätzten Parameter in Logit- bzw. Probit-Modellen haben typischerweise die gleiche Größenordnung, da die beiden Verteilungsfunktionen sehr ähnlich sind.	
Die Likelihood-Funktion wird auch als Score-Vektor bezeichnet.	
Im linearen Modell gibt die Regressionskonstante den Mittelwert der abhängigen Variable an.	
Wenn der p-Wert größer ist als das Signifikanzniveau eines Tests, wird die Nullhypothese verworfen.	
Bei einem gegebenen Schätzverfahren und Modell können sich für unterschiedliche Stichproben verschiedene Schätzwerte ergeben.	
Im Logit-Modell mit Konstante ist die Summe der vorhergesagten Wahrscheinlichkeiten gleich dem Stichprobenumfang, d.h. $\sum_i p_i = N$.	
Die vorhergesagten Wahrscheinlichkeiten eines Probit-Modells liegen immer im Intervall $[0; 1]$.	
Der KQ-Schätzer erzeugt für Steigungsparameter aber nicht für die Regressionskonstante Zufallsvariablen.	
Der ML-Schätzer für die Varianz des Störterms im linearen Modell ist verzerrt.	
Der Score-Vektor kann zur Approximation der Varianz-Kovarianz-Matrix der ML-Schätzer verwendet werden.	
Die Störterme im GLS transformierten Modell haben einen Erwartungswert von 0.	
Die Prais-Winsten-Schätzer müssen vor der Interpretation transformiert werden.	
Die Annahme $\varepsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ schließt sowohl Heteroskedastie als auch Autokorrelation aus.	
Für Matrizen gilt allgemein: $(\mathbf{A}')' = \mathbf{A}$	
Unterschiede in Koeffizienten für verschiedene Gruppen lassen sich mittels Interaktionstermen überprüfen.	
Multikollinearitätsprobleme können durch Vergrößerung der Stichprobe reduziert werden.	
Wenn statt eines Cochrane-Orcutt-Schätzers ein Prais-Winsten-Schätzer verwendet wird, steigt die Effizienz der Schätzung.	
Wenn keine analytische Lösung für die unbekannt Parameter existiert, kann der ML-Schätzer nicht bestimmt werden.	
Der Prais-Winsten-Schätzer verwendet transformierte Daten ohne Berücksichtigung der ersten Beobachtung.	
Bei Autokorrelation sind die mit dem Cochrane-Orcutt-Schätzer ausgewiesenen p-Werte ungültig.	
Die statistische Signifikanz eines Steigungsparameters lässt sich mittels eines F-Tests testen.	
Der Durbin-Watson Test verallgemeinert den White Test.	
Der geschätzte marginale Effekt einer erklärenden Variablen kann je nach Spezifikation des Modells unterschiedlich ausfallen.	
Mithilfe eines linearen Regressionsmodells lassen sich Elastizitäten schätzen.	
Im Fall eines Moving-Average-Prozesses der Form $\varepsilon_t = v_t + v_{t-1}$ sind die Störterme homoskedastisch, wenn v_t die Gauß-Markov-Annahmen erfüllt.	
Bei perfekter Multikollinearität ist die Matrix $\mathbf{X}'\mathbf{X}$ nicht invertierbar.	