

## Masterprüfung WS 2019/20

Fach: Ökonometrie

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

### Vorbemerkungen:

**Anzahl der Aufgaben:** Die Klausur besteht aus 4 Aufgaben, die alle bearbeitet werden müssen.  
**Es wird nur der Lösungsbogen eingesammelt.**

**Bewertung:** Es können maximal 90 Punkte erworben werden. Die maximale Punktzahl ist für jede Aufgabe in Klammern angegeben. Sie entspricht der für die Aufgabe empfohlenen Bearbeitungszeit in Minuten.

**Erlaubte Hilfsmittel:**

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

**Wichtige Hinweise:**

- Sollte es vorkommen, dass die statistischen Tabellen, die dieser Klausur beiliegen, den gesuchten 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:****[21 Punkte]**

Sie interessieren sich für die Determinanten der Erwerbstätigkeit. Dazu liegt Ihnen ein Datensatz von 3223 Personen mit folgenden Variablen vor:

- $working_i$  Dummy-Variable, =1 wenn Person  $i$  arbeitet, =0 sonst  
 $educ_i$  Dauer der Ausbildung von Person  $i$  in Jahren  
 $age_i$  Alter von Person  $i$  in Jahren  
 $female_i$  Dummy-Variable, =1 wenn Person  $i$  eine Frau ist, =0 sonst

Es wird folgendes Regressionsmodell aufgestellt und anschließend mit Stata geschätzt:

$$working_i = \beta_1 + \beta_2 educ_i + \beta_3 age_i + \beta_4 female_i + \varepsilon_i$$

Source	SS	df	MS	Number of obs	=	3,223
Model	18.6327267	3	6.2109089	F(3, 3219)	=	25.40
Residual	787.11099	3,219	.244520345	Prob > F	=	0.0000
Total	805.743717	3,222	.250075642	R-squared	=	0.0231
				Adj R-squared	=	0.0222
				Root MSE	=	.49449

working	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
educ	.0202973	.0027725	7.32	0.000	.0148613 .0257334
age	.0010639	.0008778	1.21	0.226	-.0006573 .002785
female	-.078917	.0174266	-4.53	0.000	-.1130854 -.0447486
_cons	.2420014	.0510699	4.74	0.000	.1418686 .3421342

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

- 1.1 Interpretieren Sie die geschätzten Koeffizienten  $b_2$  und  $b_4$  inhaltlich und statistisch. [4 Punkte]
- 1.2 Nennen Sie die Bedingungen für das Vorliegen eines Problems ausgelassener Variablen. Diskutieren Sie, ob es sich bei dem geschätzten Koeffizienten von  $educ$  um den kausalen Effekt der Bildung auf die Erwerbstätigkeit handelt oder ob ein Problem ausgelassener Variablen vorliegt. Nennen Sie hierfür gegebenenfalls ein Beispiel. [3 Punkte]
- 1.3 Eine Möglichkeit, um Endogenitätsprobleme zu lösen, sind Instrumentenvariablen. Stellen Sie allgemein die beiden benötigten Modellgleichungen für eine IV-Schätzung (2SLS; Two Stage Least Squares) auf. Erläutern Sie kurz verbal die Vorgehensweise des Schätzers und erklären sie die verwendeten Variablen. [4 Punkte]
- 1.4 Nennen und erläutern Sie knapp die Bedingungen, die eine geeignete Instrumentenvariable erfüllen muss [2 Punkte]
- 1.5 Nennen Sie zwei Nachteile eines linearen Wahrscheinlichkeitsmodells. [2 Punkte]

Sie schätzen das Modell mit dem Logit-Schätzer und erhalten folgende Ergebnisse:

Logistic regression	Number of obs	=	3,223
	LR chi2(3)	=	75.18
	Prob > chi2	=	0.0000
Log likelihood = -2196.4086	Pseudo R2	=	0.0168

working	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
educ	.0825677	.0114229	7.23	0.000	.0601793 .1049561
age	.0043523	.0035928	1.21	0.226	-.0026894 .011394
female	-.3217212	.0713116	-4.51	0.000	-.4614894 -.1819531
_cons	-1.050164	.2096035	-5.01	0.000	-1.46098 -.639349

1.6 Berechnen Sie den marginalen Effekt der Variable  $educ_i$  für Männer im Alter von 30 Jahren mit 9 Jahren Bildung und interpretieren Sie das Ergebnis. [6 Punkte]

### Aufgabe 2:

[17 Punkte]

Unterstellen Sie das folgende einfache Modell der Geldnachfrage:

$$\log(MI_t) = \beta_1 + \beta_2 \log(GDP_t) + \beta_3 \log(CPI_t) + \varepsilon_t$$

wobei

- $MI_t$  nominaler Wert der M1-Geldmenge im Jahr  $t$  (in Mrd. US-\$)
- $GDP_t$  reales Bruttosozialprodukt im Jahr  $t$  (in Mrd. US-\$)
- $CPI_t$  Preisindex im Jahr  $t$

Das Modell wird mit US-Quartalsdaten von 1980-Q1 bis 1999-Q4 geschätzt. Aus einer KQ-Schätzung ergibt sich:

$$\sum_{t=2}^T e_t e_{t-1} = 1,432 \text{ und } \sum_{t=2}^T e_{t-1}^2 = 1,637$$

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

- 2.1 Wie würden Sie den Parameter  $\beta_2$  allgemein inhaltlich interpretieren? [1 Punkt]
- 2.2 Berechnen Sie den Autokorrelationskoeffizienten  $\hat{\rho}$  und die (approximative) Durbin-Watson Statistik  $dw$ . [3 Punkte]
- 2.3 Unterstellen Sie in diesem Aufgabenteil für die Durbin-Watson-Statistik einen Wert von 1,45 und testen Sie auf dem 5%-Signifikanz-niveau auf positive Autokorrelation erster Ordnung. Geben Sie hierzu auch die Nullhypothese, die Alternativhypothese sowie die Freiheitsgrade und die jeweiligen kritischen Werte an. Hinweis: Es liegen Quartalsdaten für 20 Jahre vor. [4 Punkte]
- 2.4 Führen Sie einen Breusch-Pagan-Test auf Heteroskedastie auf einem Signifikanzniveau von 1% durch. Geben Sie die Hypothesen und Hilfsregression an. Definieren Sie die abhängige Variable der Hilfsregression. Geben Sie zudem die Teststatistik, die Freiheitsgrade, den kritischen Wert und Ihre Testentscheidung an. Hinweis: Unterstellen Sie für die Hilfsregression ein  $R^2$  von 0,112. [6 Punkte]

2.5 Zeigen Sie für ein lineares Modell  $y_t = x_t\beta + \varepsilon_t$  mit Autokorrelation zweiter Ordnung, wie eine Aufnahme der verzögerten abhängigen Variablen das Autokorrelationsproblem beseitigen kann. [3 Punkte]

**Aufgabe 3:**

**[22 Punkte]**

Im Rahmen ihrer Masterarbeit untersuchen Sie die Determinanten des Stundenlohns. Hierfür haben Sie bereits folgende Schätzgleichung aufgestellt:

$$\ln(wage_i) = \beta_1 + \beta_2 workexp_i + \beta_3 workexp2_i + \beta_4 yedu_i + \beta_5 female_i + \beta_6 married_i + \beta_7 (female \cdot married)_i + \varepsilon_i$$

wobei

- $\ln(wage_i)$  logarithmierter Stundenlohn
- $workexp_i$  Arbeitsmarkterfahrung in Jahren
- $workexp2_i$  quadrierte Arbeitsmarkterfahrung
- $yedu_i$  Anzahl an Bildungsjahren
- $female_i$  =1, wenn Individuum i eine Frau ist, 0 sonst
- $married_i$  =1, wenn das Individuum i verheiratet ist, 0 sonst

Das Modell wird mit Daten des SOEP geschätzt. Aus einer KQ-Schätzung ergibt sich:

Source	SS	df	MS	Number of obs	=	2,486
Model	118.897588	6	19.8162646	F(6, 2479)	=	133.28
Residual	368.584976	2,479	.148682927	Prob > F	=	0.0000
				R-squared	=	0.2439
				Adj R-squared	=	0.2421
Total	487.482563	2,485	.196170046	Root MSE	=	.38559

lnwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
workexp	.0296772	.0029668	??			
workexp2	-.000527	.0000686	-7.68	0.000	-.0006615	-.0003925
yedu	.0623873	.0030798	20.26	0.000	.056348	.0684266
female	-.1263009	.0258158	-4.89	0.000	-.1769236	-.0756783
married	.0970396	.0220943	4.39	0.000	.0537143	.1403649
female*married	-.1542244	.0339478	??			
_cons	1.503863	.0503606	29.86	0.000	1.405109	1.602616

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

- 3.1 Interpretieren Sie den Koeffizienten von *female* statistisch und inhaltlich. [2 Punkte]
- 3.2 Testen Sie auf dem 5%-Niveau die Hypothese, dass der Effekt von *yedu* höchstens 7% beträgt. Geben Sie das Testverfahren, die Hypothesen, die Teststatistik und die Testentscheidung an. [5 Punkte]

- 3.3 Bestimmen Sie den marginalen Effekt von *married* auf den logarithmierten Stundenlohn allgemein, für verheiratete Männer und für verheiratete Frauen. Überprüfen Sie, ob sich der Zusammenhang zwischen Stundenlohn und Ehestatus bei verheirateten Frauen und verheirateten Männern statistisch signifikant auf dem 5%-Niveau unterscheidet. Geben Sie die Null- und Alternativhypothese sowie die Teststatistik und Testentscheidung an. [6 Punkte]
- 3.4 Berechnen und interpretieren Sie das 90%-Konfidenzintervall für den Koeffizienten von *workexper*. Ist der geschätzte Koeffizient signifikant von Null verschieden? [5 Punkte]
- 3.5 Ein Kommilitone merkt an, dass Ihr Modell eventuell fehlspezifiziert sein könnte. Mit welchem Test können Sie auf dem 1%-Niveau überprüfen, ob nichtlineare Funktionen des geschätzten Modells dazu beitragen,  $y_i$  zu erklären? Geben Sie den Namen des Tests, die Hilfsregression und die Hypothesen an. Sie stellen fest, dass die Teststatistik am 1%-Niveau verworfen wird. Was bedeutet dieses Ergebnis? [4 Punkte]

#### Aufgabe 4 - MC Fragen

[30 Punkte]

Bitte geben Sie die zutreffende Antwort **auf Ihrem Multiple-Choice-Lösungsblatt** an. Zu jeder Frage gibt es genau eine richtige Antwort. Für jede korrekt beantwortete Frage erhalten Sie einen Punkt. Falsche Antworten führen nicht zu Punktabzug. Bei mehr oder weniger als einer markierten Antwort auf eine Frage gilt diese als nicht beantwortet. **Angaben auf dem Aufgabenblatt werden nicht gewertet.**

1.	Im linearen Wahrscheinlichkeitsmodell
a	liegt zwangsläufig Autokorrelation vor.
b	können vorhergesagte Werte größer als 1 auftreten.
c	kann nur das Vorzeichen der Koeffizienten interpretiert werden.
d	können die Schätzkoeffizienten nie als marginale Effekte interpretiert werden.

2.	Man spricht von einem überidentifizierten Modell, wenn
a	exakt soviele Instrumente wie endogene Variablen vorliegen.
b	ein Schätzmodell den Durbin-Wu-Hausman Test besteht.
c	mehr Instrumente als endogene Variablen vorliegen.
d	keine Heteroskedastie vorliegt.

3.	Aufgrund eines Problems ausgelassener Variablen in einer KQ-Schätzung kann Folgendes auftreten:
a	Homoskedastie.
b	perfekte Multikollinearität.
c	verzerrt geschätzte Koeffizienten.
d	eine effiziente Schätzung.

4.	Sind Polynome der erklärenden Variablen $x$ im Modell enthalten, so
a	ergibt sich der marginale Effekt der Variablen $x$ als Ableitung des auf $y$ bedingten Erwartungswerts von $x$ .
b	kann mittels eines t-Tests überprüft werden, ob die Polynome der erklärenden Variablen gemeinsam einen signifikanten Erklärungsgehalt liefern.
c	kann das Modell nicht geschätzt werden.
d	können lineare Zusammenhänge zwischen $x$ und $y$ damit nicht abgebildet werden.

5.	Sie schätzen das Modell $\ln\_wage_i = \beta_1 + \beta_2 male_i + \beta_3 \ln\_firmsize_i + \varepsilon_i$ mittels einer KQ-Schätzung, wobei $\ln\_wage$ der Logarithmus des Stundenlohns (in Euro) ist, $male$ eine Dummyvariable für Männer ist und $\ln\_firmsize$ der Logarithmus der Mitarbeiterzahl des Betriebs ist. Der geschätzte Koeffizient zu $\beta_2$ ist 0,09. Welche Interpretation ist richtig?
a	Männer verdienen c.p. im Mittel etwa 9% mehr pro Stunde als Frauen.
b	Männer verdienen c.p. im Mittel etwa 0,09 Cent mehr pro Stunde als Frauen.
c	Männer verdienen c.p. im Mittel etwa 0,09% mehr pro Stunde als Frauen.
d	Ohne Kenntnis von $b_3$ kann keine Aussage bezüglich der Interpretation von $b_2$ getroffen werden.

6.	In einer IV-Schätzung wird auf der ersten Stufe
a	die endogene Variable auf die abhängige Variable der zweiten Stufe regressiert.
b	die endogene Variable auf das Instrument und alle Kontrollvariablen regressiert.
c	die endogene Variable auf das Instrument und die abhängige Variable der zweiten Stufe regressiert.
d	eine Logit-Regression durchgeführt.

7.	Instrumentvariablenschätzungen können dazu genutzt werden, um
a	Messfehler zu korrigieren.
b	Heteroskedastie zu korrigieren.
c	zu testen, ob Autokorrelation vorliegt.
d	perfekte Multikollinearität zu korrigieren.

8.	Sie schätzen das Modell $Lebenszufriedenheit_i = \beta_1 + \beta_2 Lohn_i + \epsilon_i$ . Sie vermuten, dass ein Problem ausgelassener Variablen vorliegt, da die Variable <i>Bildung</i> nicht berücksichtigt wurde. Welche Aussage ist korrekt?
a	Das Auslassen von <i>Bildung</i> in der Schätzung führt zu Homoskedastie.
b	Wenn $Cov(Bildung, Lohn) > 0$ und $Cov(Bildung, Lebenszufriedenheit) = 0$ , dann ist $\beta_2$ unterschätzt.
c	Wenn $Cov(Bildung, Lohn) > 0$ und $Cov(Bildung, Lebenszufriedenheit) < 0$ , dann ist $\beta_2$ unverzerrt geschätzt.
d	Wenn $Cov(Bildung, Lohn) > 0$ und $Cov(Bildung, Lebenszufriedenheit) > 0$ , dann ist $\beta_2$ überschätzt.

9.	Welche Aussage bezüglich marginaler Effekte ist richtig?
a	Die Schätzkoeffizienten von Logit-Regressionen können nicht als marginale Effekte interpretiert werden.
b	Marginale Effekte in Probit-Regressionen können nur bezüglich der Signifikanz und des Vorzeichens interpretiert werden.
c	Im linearen Wahrscheinlichkeitsmodell entspricht der durchschnittliche marginale Effekt nicht dem marginalen Effekt am Mittel der Daten.
d	Marginale Effekte können nur für binäre Regressoren berechnet werden.

10.	Ein konsistenter Schätzer
a	minimiert die Summe der quadrierten Residuen.
b	ist auch ein unverzerrter Schätzer.
c	konvergiert bei steigender Stichprobengröße gegen den zu schätzenden Populationsparameter.
d	gehört zur Klasse der Best Linear Unbiased Estimators (BLUE).

11.	Für ein lineares Modell mit 40 beobachteten Perioden, einer Konstante und 2 unabhängigen Variablen wird ein Durbin-Watson Test auf positive Autokorrelation auf Signifikanzniveau von 5% durchgeführt. Die Annahmen des DW-Tests treffen zu. Der empirische Wert $dw_{emp}$ beträgt 1,9. Folgende Aussage kann basierend auf den Test getroffen werden:
a	Es kann keine Aussage bezüglich der Autokorrelation getroffen werden.
b	Es besteht eine Evidenz auf positive Autokorrelation.
c	$H_0$ kann abgelehnt werden.
d	Keine der oberen Aussagen trifft zu.

12.	Das Informationskriterium AIC
a	wird ausschließlich beim Vergleich genesteter Modelle genutzt.
b	„bestraft“ für zusätzliche Regressoren weniger als das BIC.
c	wird ausschließlich bei Analysen von Querschnittsdaten verwendet.
d	nimmt Werte zwischen 0 und 1 an.

13.	Folgendes lineare Modell $\ln(h\_wage) = \beta_1 + \beta_2 yedu_i + \beta_3 yedu_i^2 + \beta_4 fem_i + \beta_5 yedu_i \cdot fem_i$ wird mit dem KQ-Verfahren geschätzt. Dabei ist <i>h_wage</i> der Stundenlohn, <i>yedu</i> Gesamtanzahl der Bildungsjahre und <i>fem</i> eine Dummy-Variablen für Frauen. Der marginale Effekt einer Erhöhung der Bildungsjahre um ein Jahr beträgt:
a	$\beta_2 + \beta_3$
b	$\beta_2 + 2\beta_3 \cdot yedu_i + \beta_5 \cdot fem_i$
c	$\beta_2$
d	$\beta_2 + \beta_3 + \beta_5$

14.	Welche Aussage bezüglich marginaler Effekte ist falsch?
a	Marginale Effekte für binäre Regressoren können durch den Unterschied in den vorhergesagten Wahrscheinlichkeiten bestimmt werden, wobei alle anderen erklärenden Variablen feste Werte zugewiesen bekommen.
b	Die Schätzkoeffizienten von Logit-Regressionen können als marginale Effekte interpretiert werden.
c	Die Schätzkoeffizienten von Probit-Regressionen können nur bezüglich der Signifikanz und des Vorzeichens interpretiert werden.
d	Der marginale Effekt am Mittel der Daten aus Logit-Regressionen variiert nicht zwischen Individuen.

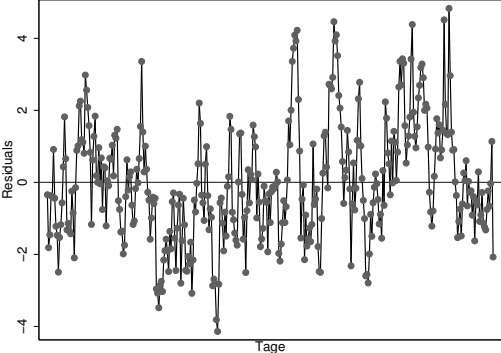
15.	Eine Logit-Schätzung resultiert in einer Log-Likelihood von -1335,12. Um 6 Parameter auf gemeinsame Signifikanz zu testen, wird ein Likelihood-Ratio Test am 5% Signifikanzniveau durchgeführt. Der Log-Likelihoodwert des restringierten Modells ist -1340,67. Welche der Aussagen trifft zu?
a	$H_0$ kann abgelehnt werden.
b	Mindestens einer der getesteten Koeffizienten hat signifikanten Einfluss auf die Wahrscheinlichkeit.
c	$H_0$ kann nicht abgelehnt werden.
d	Der Likelihood-Ratio Test ist ungeeignet, da es sich hier um ein genestetes Modell handelt.

16.	Der KQ-Schätzer $b$ lässt sich berechnen, wenn
a	die Matrix $X'X$ singulär ist.
b	nicht invertierbar ist.
c	die Matrix $X'X$ vollen Spaltenrang aufweist.
d	perfekte Multikollinearität besteht.

17.	Der F-Test für die Signifikanz einer Instrumentenvariable auf der ersten Stufe einer IV(2SLS)-Schätzung liefert einen F-Wert von 7,32. Welche Aussage ist richtig?
a	Sowohl die Exogenitäts- als auch die Relevanzbedingung sind erfüllt.
b	Es kann keine Aussage bezüglich der Relevanz des Instruments getroffen werden.
c	Das Instrument ist exogen.
d	Es handelt sich um ein schwaches Instrument.

18.	Die Varianz-Kovarianzmatrix der Störterme eines Regressionsmodells sei $\sigma^2\Omega$ , wobei $\Omega = \begin{bmatrix} 1 & 0,4 & 1,2 \\ 0,4 & 1 & 0 \\ 1,2 & 0 & 1 \end{bmatrix}$
a	Homoskedastie und positive Autokorrelation.
b	Heteroskedastie und positive Autokorrelation.
c	Homoskedastie und keine Autokorrelation.
d	Heteroskedastie und keine Autokorrelation.

19.	Die Maximum-Likelihood-Methode zur Schätzung eines linearen Modells
a	ergibt unverzerrte Varianzschätzer.
b	benötigt keine Annahme über die Verteilung der Störterme.
c	kann zur Schätzung des linearen Regressionsmodells verwendet werden.
d	benötigt Annahmen bezüglich der Verteilung der Regressoren.

20.	<p>Die Abbildung zeigt das Muster der Residuen aus der KQ-Regression des Modells</p> $\ln(\text{unfall}_t) = \beta_1 + \beta_2 \text{temp}_t + \beta_3 \text{tempsq}_t + \beta_4 \text{wochenende}_t + \varepsilon_t$ <p>Die Anzahl der Verkehrsunfälle (Variable <i>unfall</i>) wird mit tagesgenauen Zeitreihendaten analysiert. Erklärende Variablen sind Temperatur (Variable <i>temp</i>, in Grad Celsius gemessen) und eine Indikatorvariable für das Wochenende (Variable <i>wochenende</i>= 1, falls Wochentag Samstag oder Sonntag;= 0 sonst). Die Temperatur wird zusätzlich in quadrierter Form in das Modell aufgenommen (Variable <i>tempsq</i>).</p>  <p>Der Plot der Residuen deutet ...</p>
a	auf nicht korrelierte Fehler ( $Cov(\varepsilon_t, \varepsilon_{t-1}) = 0$ ), da zeitlich aufeinanderfolgende Residuen $e_t$ und $e_{t-1}$ nicht korreliert sind.
b	auf positive Autokorrelation der Fehler hin.
c	auf heteroskedastische Fehler hin.
d	auf homoskedastische Fehler hin.

21.	In einem Regressionsmodell wird der Zusammenhang zwischen $x_i$ und $y_i$ geschätzt. Wenn unter Hinzunahme der Variable $z_i$ eine positive Verzerrung des Koeffizienten $x$ entsteht, dann muss Folgendes gelten:
a	der Einfluss der Variable $x_i$ wird unterschätzt.
b	Der Koeffizient der Variable $x_i$ ist negativ und die Kovarianz von $x_i$ und $z_i$ ist negativ.
c	Der Koeffizient der Variable $x_i$ ist positiv und die Kovarianz von $x_i$ und $z_i$ ist negativ.
d	Der Koeffizient der Variable $x_i$ ist negativ und die Kovarianz von $x_i$ und $z_i$ ist positiv.

22.	Der Zusammenhang zwischen Typ I ( $\alpha$ ) und Typ II ( $\beta$ ) Fehlern ist folgendermaßen:
a	Je größer $\alpha$ , umso größer $\beta$ .
b	$\alpha$ und $\beta$ hängen nicht zusammen.
c	Je kleiner $\alpha$ , umso kleiner $\beta$ .
d	Je größer $\alpha$ , umso kleiner $\beta$ .

23.	Welche Annahme wird nicht für das Gauß-Markov-Theorem benötigt?
a	$\{\varepsilon_1, \dots, \varepsilon_N\}$ und $\{x_1, \dots, x_N\}$ sind unabhängig
b	$E\{\varepsilon_i\} = 0$
c	$Cov\{\varepsilon_i, \varepsilon_j\} = 0$ für $i, j = 1, \dots, N$ $i \neq j$
d	$Var\{\varepsilon_i\} = \sigma^2 \Psi$

24.	Beim Vergleich der Schätzgüte von nicht genesteten linearen Modellen verwendet man:
a	AIC und BIC.
b	$R^2$ .
c	Likelihood-Ratio Test.
d	keine der oben genannten Möglichkeiten.

25.	Bei Vorliegen von Autokorrelation
a	ist das Modell unter Umständen fehlspezifiziert.
b	sind die Elemente der Hauptdiagonale der Varianz der Störterme nicht identisch.
c	ist die Varianz-Kovarianz-Matrix orthogonal.
d	sind die Schätzkoeffizienten verzerrt geschätzt.



26.	Für eine gültige Instrumentvariable müssen zwei Bedingungen gelten:
a	Endogenitäts- und Relevanzbedingung.
b	Effizienz und Linearität.
c	Relevanz- und Exogenitätsbedingung.
d	Unverzerrtheits- und Konsistenzbedingung.

27.	Die Hilfsregression im Breusch-Pagan-Test regressiert:
a	die unabhängige Variable auf die abhängigen Variablen.
b	die quadrierten Residuen auf die unabhängigen Variablen des Modells.
c	die Residuen auf die verzögerten Residuen der vorherigen Zeitpunkte.
d	die abhängige Variable auf die Quadrate der unabhängigen Variablen.

28.	In Bezug auf das McFadden $R^2$ ist folgende Aussage richtig (wobei $L_1$ =Likelihood im unrestringierten Modell und $L_0$ = Likelihood im restringierten Modell):
a	passt das Modell perfekt, so gilt: $L_1 = 1$ und $\log L_1 = 0$ .
b	die Log-Werte liegen zwischen 0 und 1.
c	im besten Fall ist $L_1 = 0$ .
d	passt das Modell perfekt, so gilt: $\log L_1 = \log L_0$ .

29.	Welche Aussage bezüglich statistischer Tests ist falsch?
a	Der F-Test kann auf gemeinsame Signifikanz von Koeffizienten testen.
b	Der Chow-Test ist ein Test auf Strukturbruch.
c	Wald-, Likelihood-Ratio- und Lagrange-Multiplier-Test sind nicht geeignet, um nicht-lineare Restriktionen zu testen.
d	Der Breusch-Godfrey-Test testet auf Autokorrelation erster Ordnung.

30.	Die Aufnahme einer irrelevanten Variable $z$ in die KQ-Schätzung des Modells $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$
a	erhöht die Stichprobenvarianz von $y$ .
b	erhöht die Varianz des Koeffizientenschätzers $b_2$ , wenn $cov(x, z) \neq 0$ .
c	führt zu inkonsistenten Koeffizientenschätzern, wenn die Stichprobe klein ist.
d	führt schon bei leichter Multikollinearität zu verzerrten Koeffizientenschätzern.

## Tabelle 2: Perzentile der $t$ -Verteilung

Zelleneintrag:  $x$ , sodass  $\text{Prob}[t_n \leq x] = P$ , mit  $n$  Freiheitsgraden

P n	0.75	0.9	0.95	0.975	0.99	0.995
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.817	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707
7	0.711	1.415	1.895	2.365	2.998	3.500
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.262	2.821	3.250
10	0.700	1.372	1.813	2.228	2.764	3.169
11	0.697	1.363	1.796	2.201	2.718	3.106
12	0.696	1.356	1.782	2.179	2.681	3.055
13	0.694	1.350	1.771	2.160	2.650	3.012
14	0.692	1.345	1.761	2.145	2.624	2.977
15	0.691	1.341	1.753	2.131	2.602	2.947
16	0.690	1.337	1.746	2.120	2.583	2.921
17	0.689	1.333	1.740	2.110	2.567	2.898
18	0.688	1.330	1.734	2.101	2.552	2.878
19	0.688	1.328	1.729	2.093	2.539	2.861
20	0.687	1.325	1.725	2.086	2.528	2.845
21	0.686	1.323	1.721	2.080	2.518	2.831
22	0.686	1.321	1.717	2.074	2.508	2.819
23	0.685	1.320	1.714	2.069	2.500	2.807
24	0.685	1.318	1.711	2.064	2.492	2.797
25	0.684	1.316	1.708	2.060	2.485	2.787
26	0.684	1.315	1.706	2.056	2.479	2.779
27	0.684	1.314	1.703	2.052	2.473	2.771
28	0.683	1.313	1.701	2.048	2.467	2.763
29	0.683	1.311	1.699	2.045	2.462	2.756
30	0.683	1.310	1.697	2.042	2.457	2.750
35	0.682	1.306	1.690	2.030	2.438	2.724
40	0.681	1.303	1.684	2.021	2.423	2.705
45	0.680	1.301	1.679	2.014	2.412	2.690
50	0.679	1.299	1.676	2.009	2.403	2.678
60	0.679	1.296	1.671	2.000	2.390	2.660
70	0.678	1.294	1.667	1.994	2.381	2.648
80	0.678	1.292	1.664	1.990	2.374	2.639
90	0.677	1.291	1.662	1.987	2.368	2.632
100	0.677	1.290	1.660	1.984	2.364	2.626
$\infty$	0.674	1.282	1.645	1.960	2.326	2.576

Quelle: In R generiert

**Tabelle 3: Perzentile der  $\chi^2$ -Verteilung**

Zelleneintrag: c, sodass  $\text{Prob}[\chi_n^2 \leq c] = P$ , mit n Freiheitsgraden

<b>n \ P</b>	<b>0.005</b>	<b>0.01</b>	<b>0.025</b>	<b>0.05</b>	<b>0.1</b>	<b>0.25</b>	<b>0.5</b>	<b>0.75</b>	<b>0.9</b>	<b>0.95</b>	<b>0.975</b>	<b>0.99</b>	<b>0.995</b>
<b>1</b>	0.00004	0.0002	0.001	0.004	0.016	0.102	0.455	1.323	2.706	3.842	5.024	6.635	7.879
<b>2</b>	0.01	0.02	0.05	0.10	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
<b>3</b>	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.81	9.35	11.34	12.84
<b>4</b>	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
<b>5</b>	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
<b>6</b>	0.68	0.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
<b>7</b>	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
<b>8</b>	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.09	21.96
<b>9</b>	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.67	23.59
<b>10</b>	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
<b>11</b>	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.73	26.76
<b>12</b>	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.22	28.30
<b>13</b>	3.57	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82
<b>14</b>	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.68	26.12	29.14	31.32
<b>15</b>	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
<b>16</b>	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
<b>17</b>	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
<b>18</b>	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.16
<b>19</b>	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58
<b>20</b>	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.57	40.00
<b>21</b>	8.03	8.90	10.28	11.59	13.24	16.34	20.34	24.93	29.62	32.67	35.48	38.93	41.40
<b>22</b>	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.92	36.78	40.29	42.80
<b>23</b>	9.26	10.20	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18
<b>24</b>	9.89	10.86	12.40	13.85	15.66	19.04	23.34	28.24	33.20	36.42	39.36	42.98	45.56
<b>25</b>	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.31	46.93
<b>30</b>	13.79	14.95	16.79	18.49	20.60	24.48	29.34	34.80	40.26	43.77	46.98	50.89	53.67
<b>35</b>	17.19	18.51	20.57	22.47	24.80	29.05	34.34	40.22	46.06	49.80	53.20	57.34	60.27
<b>40</b>	20.71	22.16	24.43	26.51	29.05	33.66	39.34	45.62	51.81	55.76	59.34	63.69	66.77
<b>45</b>	24.31	25.90	28.37	30.61	33.35	38.29	44.34	50.98	57.51	61.66	65.41	69.96	73.17
<b>50</b>	27.99	29.71	32.36	34.76	37.69	42.94	49.33	56.33	63.17	67.50	71.42	76.15	79.49

Quelle: In R generiert

**Tabelle 4a: 95% Perzentile der F-Verteilung**

Zelleneintrag:  $f$ , sodass  $\text{Prob}[F_{n_1, n_2} \leq f] = 0.95$

n2 \ n1	n1 = Freiheitsgrade des Zählers								
	1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97
∞	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.88

n2 \ n1	n1 = Freiheitsgrade des Zählers								
	10	12	15	20	30	40	50	60	∞
1	241.88	243.91	245.95	248.01	250.10	251.14	251.77	252.20	254.19
2	19.40	19.41	19.43	19.45	19.46	19.47	19.48	19.48	19.49
3	8.79	8.74	8.70	8.66	8.62	8.59	8.58	8.57	8.53
4	5.96	5.91	5.86	5.80	5.75	5.72	5.70	5.69	5.63
5	4.74	4.68	4.62	4.56	4.50	4.46	4.44	4.43	4.37
6	4.06	4.00	3.94	3.87	3.81	3.77	3.75	3.74	3.67
7	3.64	3.57	3.51	3.44	3.38	3.34	3.32	3.30	3.23
8	3.35	3.28	3.22	3.15	3.08	3.04	3.02	3.01	2.93
9	3.14	3.07	3.01	2.94	2.86	2.83	2.80	2.79	2.71
10	2.98	2.91	2.85	2.77	2.70	2.66	2.64	2.62	2.54
15	2.54	2.48	2.40	2.33	2.25	2.20	2.18	2.16	2.07
20	2.35	2.28	2.20	2.12	2.04	1.99	1.97	1.95	1.85
25	2.24	2.16	2.09	2.01	1.92	1.87	1.84	1.82	1.72
30	2.16	2.09	2.01	1.93	1.84	1.79	1.76	1.74	1.63
40	2.08	2.00	1.92	1.84	1.74	1.69	1.66	1.64	1.52
50	2.03	1.95	1.87	1.78	1.69	1.63	1.60	1.58	1.45
70	1.97	1.89	1.81	1.72	1.62	1.57	1.53	1.50	1.36
100	1.93	1.85	1.77	1.68	1.57	1.52	1.48	1.45	1.30
∞	1.83	1.75	1.67	1.57	1.46	1.39	1.34	1.31	1.30

Quelle: In R generiert

**Tabelle 4b: 99% Perzentile der F-Verteilung**

Zelleneintrag: f, sodass  $\text{Prob}[F_{n1,n2} \leq f] = 0.99$

n2 \ n1	n1 = Freiheitsgrade des Zählers								
	1	2	3	4	5	6	7	8	9
1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.79
70	7.01	4.92	4.07	3.60	3.29	3.07	2.91	2.78	2.67
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59
∞	6.66	4.63	3.80	3.34	3.04	2.82	2.66	2.53	2.43

n2 \ n1	n1 = Freiheitsgrade des Zählers								
	10	12	15	20	30	40	50	60	∞
1	6055.85	6106.32	6157.28	6208.73	6260.65	6286.78	6302.52	6313.03	6362.68
2	99.40	99.42	99.43	99.45	99.47	99.47	99.48	99.48	99.50
3	27.23	27.05	26.87	26.69	26.50	26.41	26.35	26.32	26.14
4	14.55	14.37	14.20	14.02	13.84	13.75	13.69	13.65	13.47
5	10.05	9.89	9.72	9.55	9.38	9.29	9.24	9.20	9.03
6	7.87	7.72	7.56	7.40	7.23	7.14	7.09	7.06	6.89
7	6.62	6.47	6.31	6.16	5.99	5.91	5.86	5.82	5.66
8	5.81	5.67	5.52	5.36	5.20	5.12	5.07	5.03	4.87
9	5.26	5.11	4.96	4.81	4.65	4.57	4.52	4.48	4.32
10	4.85	4.71	4.56	4.41	4.25	4.17	4.12	4.08	3.92
15	3.80	3.67	3.52	3.37	3.21	3.13	3.08	3.05	2.88
20	3.37	3.23	3.09	2.94	2.78	2.69	2.64	2.61	2.43
25	3.13	2.99	2.85	2.70	2.54	2.45	2.40	2.36	2.18
30	2.98	2.84	2.70	2.55	2.39	2.30	2.25	2.21	2.02
40	2.80	2.66	2.52	2.37	2.20	2.11	2.06	2.02	1.82
50	2.70	2.56	2.42	2.27	2.10	2.01	1.95	1.91	1.70
70	2.59	2.45	2.31	2.15	1.98	1.89	1.83	1.78	1.56
100	2.50	2.37	2.22	2.07	1.89	1.80	1.74	1.69	1.45
∞	2.34	2.20	2.06	1.90	1.72	1.61	1.54	1.50	1.16

Quelle: In R generiert

**Tabelle 5: Durbin-Watson Teststatistik**

$d_L$  und  $d_U$  am 5% Signifikanzniveau

n	k=2		k=3		k=4		k=5		k=6		k=11		k=16	
	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21				
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15	0.16	3.30		
17	1.13	1.38	1.02	1.53	0.90	1.71	0.78	1.90	0.67	2.21	0.20	3.18		
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.15	0.24	3.07		
19	1.18	1.40	1.08	1.54	0.97	1.68	0.86	1.85	0.75	2.10	0.29	2.97		
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	2.06	0.34	2.89	0.06	3.68
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	2.02	0.38	2.81	0.09	3.58
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.99	0.42	2.73	0.12	3.55
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.96	0.47	2.67	0.15	3.41
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.64	0.51	2.61	0.19	3.33
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.92	0.54	2.57	0.22	3.25
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.90	0.58	2.51	0.26	3.18
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.89	0.62	2.47	0.29	3.11
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.88	0.65	2.43	0.33	3.05
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.86	0.68	2.40	0.36	2.99
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.85	0.71	2.36	0.39	2.94
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.84	0.74	2.33	0.43	2.99
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.83	0.77	2.31	0.46	2.84
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.83	0.80	2.28	0.49	2.80
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.82	0.82	2.26	0.52	2.75
35	1.40	1.52	1.34	1.53	1.28	1.65	1.22	1.73	1.16	1.81	0.85	2.24	0.55	2.72
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.81	0.87	2.22	0.58	2.68
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80	0.89	2.20	0.60	2.65
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.80	0.91	2.18	0.63	2.61
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.80	0.93	2.16	0.65	2.59
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79	0.95	2.15	0.68	2.56
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.79	1.04	2.09	0.79	2.44
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.79	1.11	2.04	0.88	2.35
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.78	1.17	2.01	0.96	2.28
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77	1.22	1.98	1.03	2.23
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77	1.27	1.96	1.09	2.18
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77	1.30	1.95	1.14	2.15
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77	1.34	1.94	1.18	2.12
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77	1.37	1.93	1.22	2.09
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77	1.40	1.92	1.26	2.07
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78	1.42	1.91	1.29	2.06
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78	1.44	1.90	1.32	2.04
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78	1.46	1.90	1.35	2.03
150	1.72	1.75	1.72	1.76	1.71	1.76	1.69	1.77	1.68	1.79	1.61	1.86	1.54	1.52
200	1.76	1.78	1.76	1.78	1.75	1.79	1.74	1.80	1.73	1.81	1.68	1.86	1.62	1.61

Anmerkung: k ist die Anzahl der Regressoren mit Konstante.

Quelle: Johnston, J. and J. DiNardo, 1997, *Econometric Methods*, 4<sup>th</sup> ed., New York: McGraw-Hill

# Formeln Ökonometrie

## I. Mathematische Grundlagen

### i. Algebra

$$(AB)' = B' A'$$

$$(A')' = A$$

$$AA^{-1} = I \text{ und } A^{-1}A = I$$

$$(A^{-1})' = (A')^{-1}$$

$$(AB)^{-1} = B^{-1}A^{-1}$$

### Ableitung von Matrizen

Für die Matrix A und die Vektoren x und c gilt bei passender Ordnung:

$$\frac{\partial c'x}{\partial x} = c$$

$$\frac{\partial Ax}{\partial x} = A'$$

Wenn A symmetrisch ist:  $\frac{\partial x'Ax}{\partial x} = 2Ax$

### ii. Varianz, Kovarianz und Korrelationskoeffizient

#### Varianz:

$$\begin{aligned} \sigma^2 &= V\{Y\} = E\{(Y - E(Y))^2\} \\ &= E\{Y^2\} - E\{Y\}^2 \end{aligned}$$

#### Kovarianz:

$$\begin{aligned} \sigma_{YX} &= \text{cov}\{Y, X\} = E\{(Y - E(Y))(X - E(X))\} \\ &= E\{XY\} - E\{X\}E\{Y\} \end{aligned}$$

#### Korrelationskoeffizient:

$$\rho_{YX} = \frac{\text{cov}\{Y, X\}}{\sqrt{V\{X\} \cdot V\{Y\}}} = \frac{\sigma_{YX}}{\sigma_Y \sigma_X}, \quad -1 \leq \rho_{YX} \leq 1$$

X, Y sind *unkorreliert*, wenn  $\text{cov}\{Y, X\} = 0$

#### Rechenregeln:

Wenn a, b, c, d Skalare und X, Y Zufallsvariablen sind:

$$V\{aY + b\} = a^2 V\{Y\}$$

$$V\{aY + bX\} = a^2 V\{Y\} + b^2 V\{X\} + 2ab \text{cov}\{Y, X\}$$

## II. Annahmen im linearen Modell

**A 1**  $E\{\varepsilon_i\} = 0 \quad i = 1, 2, \dots, N$

**A 2**  $\{x_1, \dots, x_N\}$  und  $\{\varepsilon_1, \dots, \varepsilon_N\}$  sind unabhängig

**A 3**  $V\{\varepsilon_i\} = \sigma^2 \quad i = 1, 2, \dots, N$

**A 4**  $\text{cov}\{\varepsilon_i, \varepsilon_j\} = 0 \quad i, j = 1, \dots, N, i \neq j$

**A 5**  $\varepsilon \sim N(0, \sigma^2 I_N)$

**A 5'**  $\varepsilon_i \sim \text{NID}(0, \sigma^2)$

**A 6**  $\frac{1}{N} \sum_{i=1}^N x_i x_i'$  konvergiert gegen eine positiv definite nichtsinguläre Matrix  $\Sigma_{xx}$ .

**A 7**  $E\{x_i \varepsilon_i\} = 0$

**A 8**  $x_t$  und  $\varepsilon_t$  sind für gegebenes t statistisch unabhängig

**A 9**  $V\{\varepsilon | X\} = \sigma^2 \text{Diag}\{h_i^2\} = \sigma^2 \Psi$

**A 10**  $E\{\varepsilon | X\} = 0$

**A 11**  $\varepsilon_t \sim \text{IID}(0, \sigma^2)$

**A 12**  $\varepsilon_t$  ist über die Zeit unkorreliert, mit Erwartungswert 0.

## III. Das Lineare Regressionsmodell

### Lösung für $\beta$ :

$$b = \left( \sum_{i=1}^N x_i x_i' \right)^{-1} \sum_{i=1}^N x_i y_i \quad \text{wenn } \sum_{i=1}^N x_i x_i' \text{ invertierbar}$$

ist bzw.

$$b = (X'X)^{-1} X'y \quad \text{wenn } X'X \text{ invertierbar ist}$$

### Lösung für $b_2$ wenn $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ :

$$b_2 = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2}$$

### Varianz des KQ Schätzers:

$$V\{b | X\} = \sigma^2 (X'X)^{-1} = \sigma^2 \left( \sum_{i=1}^N x_i x_i' \right)^{-1}$$

*Unverzerrter Schätzer für  $\sigma^2$ :*

$$s^2 = \frac{1}{N-K} \sum_{i=1}^N e_i^2$$

#### IV. Maximum Likelihood

*Likelihood Funktion im Modell mit einer binären abhängigen Variable:*

$$L(\beta) = \prod_{i=1}^N P\{y_i = 1 | x_i; \beta\}^{y_i} P\{y_i = 0 | x_i; \beta\}^{1-y_i}$$

*Log-Likelihood Funktion im Modell mit einer binären abhängigen Variable:*

$$\log L(\beta) = \sum_{i=1}^N y_i \log F(x_i; \beta) + \sum_{i=1}^N (1-y_i) \log (1-F(x_i; \beta))$$

*Marginale Effekte im Probit und Logit Modell:*

Probit:

$$\frac{\partial \Phi(x_i; \beta)}{\partial x_{ik}} = \phi(x_i; \beta) \cdot \beta_k$$

Logit:

$$\frac{\partial \Lambda(x_i; \beta)}{\partial x_{ik}} = \frac{\exp(x_i; \beta)}{(1 + \exp(x_i; \beta))^2} \cdot \beta_k$$

#### V. Gütemaße

$$R^2 = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^N e_i^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

*Angepasstes  $R^2$*

$$\bar{R}^2 = 1 - \frac{\sum_{i=1}^N e_i^2 / (N-K)}{\sum_{i=1}^N (y_i - \bar{y})^2 / (N-1)}$$

*AIC*

$$AIC = \log \frac{1}{N} \sum_{i=1}^N e_i^2 + \frac{2K}{N}$$

*BIC*

$$BIC = \log \frac{1}{N} \sum_{i=1}^N e_i^2 + \frac{K}{N} \log N$$

*pseudo  $R^2$*

$$\text{pseudo}R^2 = 1 - \frac{1}{1 + 2(\log L_1 - \log L_0) / N}$$

*Mc Fadden  $R^2$*

$$\text{McFadden } R^2 = 1 - (\log L_1 / \log L_0)$$

#### VI. Tests

*Kritischer Wert bei einem einseitigen Test*

$$P\{t_k > t_{N-K, \alpha}\} = \alpha$$

*Konfidenzintervall zum Niveau  $1 - \alpha$*

$$b_k - t_{N-K, \frac{\alpha}{2}} \cdot \text{se}(b_k) < \beta_k < b_k + t_{N-K, \frac{\alpha}{2}} \cdot \text{se}(b_k)$$

*Teststatistik für einen F-Test auf gemeinsame Signifikanz*

$$F = \frac{(S_0 - S_1) / J}{S_1 / (N-K)} \sim F_{J, N-K}$$

$$F = \frac{(R_1^2 - R_0^2) / J}{(1 - R_1^2) / (N-K)}$$

*Teststatistik Goldfeld-Quandt-Test*

$$\lambda = \frac{S_A^2}{S_B^2} \sim F_{N_A - K, N_B - K}$$

*Teststatistik Durbin-Watson-Test*

$$dw = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2}$$

*Teststatistik Wald-Test*

$$\xi_W = (R\hat{\theta} - q)' [R\hat{V}R']^{-1} (R\hat{\theta} - q) \sim \chi_J^2$$

*Teststatistik Likelihood-Ratio-Test*

$$\xi_{LR} = -2 \left[ \log L(\hat{\theta}) - \log L(\tilde{\theta}) \right] \sim \chi_J^2$$