

## Masterprüfung WS 2021/22

Fach: Ökonometrie

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

### Vorbemerkungen:

- Anzahl der Aufgaben:** Die Klausur besteht aus 5 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 beigelegt sind, 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:**

[17 Punkte]

Sie interessieren sich für die Determinanten der Erwerbstätigkeit von verheirateten Frauen. Hierzu untersuchen Sie Daten von 753 verheirateten Frauen. Der Datensatz enthält folgende Informationen:

Variable	Mittelwert	Std. Abw.	Min.	Max.	Beschreibung
<i>employed</i>	0.56	0.50	0	1	=1 falls erwerbstätig, =0 sonst
<i>children</i>	1.59	1.46	0	8	Anzahl der Kinder
<i>educ</i>	12.29	2.28	5	17	Anzahl an Schuljahren
<i>huswage</i>	7.48	4.23	0.41	40.51	Stundenlohn des Ehemanns in Euro
<i>age</i>	42.54	8.07	30	60	Alter in Jahren
<i>west</i>	0.84	0.48	0	1	=1 falls Person in West-Deutschland lebt, =0 sonst
<i>uni</i>	0.15	0.36	0	1	=1 falls Person einen Uniabschluss hat, =0 sonst

Sie schätzen das folgende Modell:

$$employed_i = \beta_0 + \beta_1 \cdot children_i + \beta_2 \cdot educ_i + \beta_3 \cdot huswage_i + \epsilon_i \quad (1)$$

Source	SS	df	MS	Number of obs	=	753
				F(3, 749)	=	15.25
Model	10.6345609	3	3.55485365	Prob > F	=	0.0000
Residual	174.093195	749	.232434172	R-squared	=	0.0576
				Adj R-squared	=	0.0538
Total	184.727756	752	.245648611	Root MSE	=	.48211

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
<i>employed</i>						
<i>children</i>	-.026378	.0120398	-2.19	0.029	-.0500137	-.0027422
<i>educ</i>	.0488221	.0080439	6.07	0.000	.0330307	.0646134
<i>huswage</i>	-.0157761	.0043357	-3.64	0.000	-.0242876	-.0072645
<i>_cons</i>	.1285296	.0986089	1.30	0.193	-.0650531	.3221124

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

1.1 Interpretieren Sie  $b_2$  inhaltlich. (1 Punkt)

1.2 Ihr Kommilitone untersucht ebenfalls die Determinanten der Erwerbstätigkeit verheirateter Frauen. Er schätzt folgendes Modell (Modell 2) und erhält ein  $R^2 = 0.1022$ . Interpretieren Sie das  $R^2$  der Schätzung von Modell 2. Ist das  $R^2$  der beiden Schätzungen (Modell 1 und Modell 2) geeignet, zu bestimmen, welches Modell verwendet werden sollte? (2 Punkte)

$$employed_i = \delta_1 + \delta_2 \cdot age_i + \delta_3 \cdot uni_i + \nu_i \quad (2)$$

1.3 Nennen Sie zwei Nachteile des linearen Wahrscheinlichkeitsmodells. (2 Punkte)

Sie schätzen nun das Modell (1) erneut und erhalten den folgenden Output:

Logit regression		Number of obs	=	753
		LR chi2(3)	=	45.11
		Prob > chi2	=	0.0000
Log likelihood = -492.31983		Pseudo R2	=	0.0438

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
employed					
children	-.0701107	.0317867	-2.21	0.027	-.1324115    -.0078098
educ	.1325045	.0223629	5.93	0.000	.0886739    .1763351
huswage	-.0432866	.0121638	-3.56	0.000	-.0671272    -.0194461
_cons	-1.011702	.2676351	-3.78	0.000	-1.536257    -.487147

- 1.4 Nennen Sie die in der Schätzung getroffene Annahme bezüglich der Verteilung der Störgröße sowie eine alternative Annahme der Verteilung der Störgröße. (2 Punkte)
- 1.5 Interpretieren Sie  $b_2$  statistisch. (1 Punkt)
- 1.6 Berechnen und interpretieren Sie den marginalen Effekt von  $children_i$  auf die Wahrscheinlichkeit der Erwerbstätigkeit am Mittel der Daten. (5 Punkte)
- 1.7 Sie nehmen zusätzlich das Alter in Jahren  $age_i$  und eine Dummy-Variable für Wohnort in Westdeutschland  $west_i$  in das Modell (1) auf.

$$employed_i = \beta_0 + \beta_1 \cdot children_i + \beta_2 \cdot educ_i + \beta_3 \cdot huswage_i + \beta_4 \cdot age_i + \beta_5 \cdot west_i + \epsilon_i \quad (3)$$

Testen Sie die Hypothese, dass die beiden Variablen ( $age_i$  und  $west_i$ ) gemeinsam einen signifikanten Beitrag zum Erklärungsgehalt des Modells leisten. Führen Sie hierzu einen Likelihood-Ratio Test am 1%-Signifikanzniveau durch. Geben Sie die Hypothesen, Teststatistik, kritischen Wert und Testentscheidung an.

*Hinweis: Der Log-Likelihoodwert des unrestringierten Modells ist  $\ln L(\tilde{\beta}) = -473,183$ .* (4 Punkte)

## Aufgabe 2:

[16,5 Punkte]

Der Konsum  $C$  zum Zeitpunkt  $t$  hängt vom Volkseinkommen  $Y_t$  und vom Konsum der Vorperiode  $C_{t-1}$  ab, d.h.

$$C_t = \beta_1 Y_t + \beta_2 C_{t-1} + \epsilon_t \quad (4)$$

Genutzt werden die Quartalsdaten vom ersten Quartal 1999 bis zum zweiten Quartal 2020,  $t = 1, \dots, 86$ , mit  $C$  und  $Y$  in Milliarden Euro. Folgende Tabelle zeigt KQ-Schätzergebnisse für dieses multiple Regressionsmodell :

Kovariable	$b_j$	se( $b_j$ )
$Y_t$	0,0800	0,0144
$C_{t-1}$	0,9239	0,0159

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

- 2.1 Führen sie einen Signifikanztest für den Koeffizienten  $b_1$  am  $\alpha = 0,01$  Signifikanzniveau durch. (3 Punkte)
- 2.2 Prognostizieren Sie den Konsum für das dritte Quartal 2020 bei einem Volkseinkommen von 6,4 Milliarden Euro und einem verzögerten Konsum von 5,7 Milliarden Euro. (1,5 Punkte)

2.3 Erläutern Sie zwei Gründe, warum Sie den Durbin-Watson Test hier nicht anwenden können. (2 Punkte)

2.4 Sie regressieren die Residuen ( $e$ ) Ihrer vorherigen Regression auf die um eine Periode verzögerten Residuen ( $L1.e$ ), um einen Breusch-Godfrey-Test auf Autokorrelation 1. Ordnung am 1%-Signifikanzniveau durchzuführen. Sie erhalten folgenden Output:

Source	SS	df	MS	Number of obs	=	85
Model	25.7279993	1	25.7279993	F(1, 83)	=	1328.00
Residual	3.42911017	84	0.04082274	Prob > F	=	0.0000
				R-squared	=	0.8824
				Adj R-squared	=	0.8817
Total	29.1571094	85	0.3430248	Root MSE	=	0.2020

e	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
L1.	.9385411	.0257546	36.44	0.000	.8877155 .9893666
_cons	.0005874	.0104035	0.06	0.955	-.0199435 .0211183

Geben Sie Null- und Alternativhypothese, Teststatistik, Freiheitsgrade, Entscheidungsregel, Schlusslogik und Testergebnis an. (5 Punkte)

2.5 Erläutern Sie kurz, wie der Breusch-Godfrey-Test vorgeht, wenn Sie zusätzlich auf Autokorrelation zweiter Ordnung testen. Geben Sie für diesen Fall die Regressionsgleichung der Hilfsregression an. (2 Punkte)

2.6 Nehmen Sie an, dass Autokorrelation erster Ordnung der Form  $\epsilon_t = \rho\epsilon_{t-1} + \nu_t$  vorliegt. Erläutern Sie, inwiefern für das gegebene Modell in Gleichung (4) das Problem der Endogenität besteht. Welche Konsequenz hätte das für den KQ-Schätzer? (3 Punkte)

### Aufgabe 3:

[20,5 Punkte]

Sie interessieren sich für die Determinanten der Prüfungsleistungen von Schüler:innen im Fach Mathematik. Sie nutzen einen Datensatz mit 545 Schüler:innen aus der 8. Klasse. Es liegen folgende Variablen vor:

- $points_i$  Punktzahl von Schüler:in  $i$  in der Mathematikprüfung, zwischen 0 und 100
- $size_i$  Größe der Schulklasse von Schüler:in  $i$ , gemessen in Anzahl von Schüler:innen
- $size2_i$   $size_i$  quadriert
- $female_i$  Dummy-Variable, =1 wenn Schüler:in  $i$  weiblich ist, =0 sonst
- $piano_i$  Dummy-Variable, =1 wenn Schüler:in  $i$  regelmäßig Klavier spielt, =0 sonst

Sie stellen folgendes Regressionsmodell auf und schätzen dieses anschliessend mit STATA:

$$points_i = \beta_1 + \beta_2 size_i + \beta_3 size2_i + \beta_4 female_i + \beta_5 piano_i + \epsilon_i$$

points	Coef.	Std. Err.	t
size	.86264521	.1255323	6.87
size2	-.0802538	.0045643	-17.58
female	3.1657789	.9344655	3.39
piano	2.2553111	.8544266	2.64
_cons	.26441218	.0552238	4.79

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

- 3.1 Interpretieren Sie  $b_5$  inhaltlich. (1 Punkt)
- 3.2 Berechnen und interpretieren Sie das 95%-Konfidenzintervall für den Koeffizienten von *female*. Ist der geschätzte Koeffizient signifikant von Null verschieden? Begründen Sie Ihre Antwort. (5 Punkte)
- 3.3 Berechnen und interpretieren Sie den geschätzten marginalen Effekt der Klassengröße auf die Prüfungsleistung für eine Klassengröße von 15 Schüler:innen. Ab welcher Klassengröße wird der marginale Effekt negativ? (4 Punkte)
- 3.4 Stellen Sie ein Modell auf, mit dem Sie testen können, ob sich der marginale Effekt des Klavierspielens zwischen Schülern und Schülerinnen unterscheidet. Für welche Parameterwerte ist der marginale Effekt für Schüler grösser als für Schülerinnen? (2,5 Punkte)
- 3.5 Erklären Sie kurz allgemein, wann bei einer Schätzung das Problem „omitted variable bias“ vorliegt. Erklären Sie anschließend mit Hilfe eines Beispiels, warum dieses Problem bei der Schätzung des Effektes des Klavierspielens auf die Prüfungsleistung vorliegen könnte. (3 Punkte)
- 3.6 Ihr Kommilitone schlägt vor, die Information, ob ein Elternteil Klavier spielt als Instrumentvariable für  $piano_i$  zu nutzen. Erklären Sie kurz allgemein, welche zwei Bedingungen eine gültige Instrumentvariable erfüllen muss. Diskutieren Sie anschliessend kurz, ob diese beiden Bedingungen in diesem Fall erfüllt sind. (5 Punkte)

#### Aufgabe 4:

[6 Punkte]

Betrachten Sie einen alternativen Schätzer für den Steigungsparameter  $\beta_1$  für das simple lineare Modell

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

gegeben durch

$$\tilde{b}_1 = \frac{y_2 - y_1}{x_2 - x_1}$$

wobei  $(x_1, y_1)$  die Werte der ersten und  $(x_2, y_2)$  die der zweiten Beobachtung in den Daten sind. Nehmen Sie an, dass  $x_i$  deterministisch sind und  $E[\epsilon_i] = 0$  mit  $i = 1, 2$ . Zeigen Sie, dass der Schätzer  $\tilde{b}_1$  erwartungstreu ist.

*Hinweis: Für die erste und die zweite Beobachtung gilt:  $y_1 = \beta_0 + \beta_1 x_1 + \epsilon_1$  und  $y_2 = \beta_0 + \beta_1 x_2 + \epsilon_2$ .*

**Aufgabe 5:**

[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.	Die Varianz-Kovarianzmatrix der Störterme eines Regressionsmodells sei $\sigma^2\Psi$ , wobei $\Psi = \begin{pmatrix} 5 & 1 & 0 \\ 1 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix}$ . Hier liegt vor
a	Homoskedastie und positive Autokorrelation.
b	Heteroskedastie und positive Autokorrelation.
c	Homoskedastie und keine Autokorrelation.
d	Heteroskedastie und keine Autokorrelation.

2.	Welches der folgenden Regressionsmodelle wird den im Streudiagramm dargestellten Daten am besten gerecht?
a	$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ .
b	$y_i = \beta_0 + \beta_1 x_i^2 + \epsilon_i$ .
c	$y_i = \beta_0 + \beta_1 x_i^3 + \epsilon_i$ .
d	$y_i = \beta_0 + \beta_1 x_i^5 + \epsilon_i$ .

3.	Wald-, Likelihood-Ratio- und Lagrange-Multiplier-Tests
a	prüfen jeweils, ob die Bedingungen erster Ordnung der Likelihoodfunktion erfüllt sind.
b	werden jeweils nach einer unrestringierten Schätzung angewendet.
c	haben asymptotisch die gleiche Verteilung.
d	können nur bei mit KQ-geschätzten linearen Modellen angewendet werden.

4.	Das Bestimmtheitsmaß im multiplen Regressionsmodell und endlicher Stichprobe
a	ist immer kleiner als das korrigierte Bestimmtheitsmaß $R^2$ .
b	ist immer größer als das korrigierte Bestimmtheitsmaß $R^2$ .
c	ist definiert als der Anteil der Residuenvariation an der Gesamtvariation.
d	ist definiert als der Anteil der erklärten Variation an der Residuenvariation.

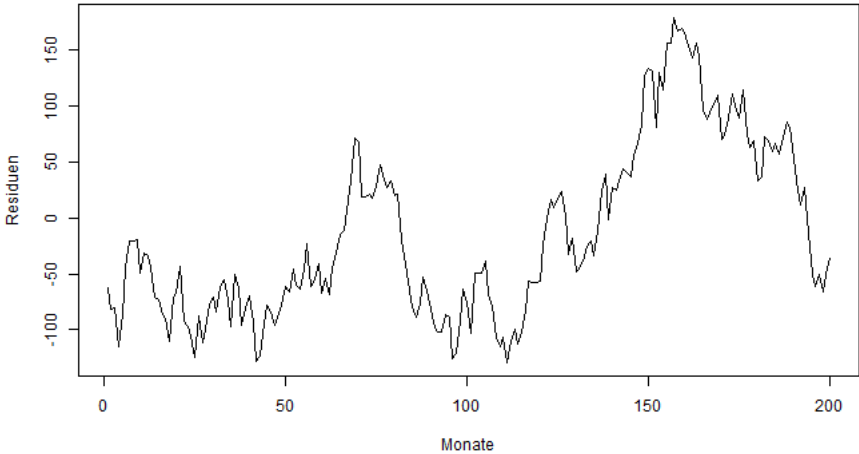
5.	Bei Vorliegen von perfekter Multikollinearität gilt für die Kreuzproduktmatrix $\mathbf{X}'\mathbf{X}$ , dass sie
a	vollen Rang hat.
b	singulär ist.
c	nicht quadratisch ist.
d	idempotent ist.

6.	Autokorrelation im Störterm kann behoben werden durch
a	eine Vergrößerung der Stichprobe.
b	die Anwendung eines Prais-Winsten Schätzers.
c	die Aufnahme von irrelevanten erklärenden Variablen.
d	eine Umskalierung der abhängigen Variable in eine kleinere Einheit.

7.	Der FGLS-Schätzer beim Prais-Winsten-Verfahren
a	ist konsistent, aber nicht unverzerrt.
b	ist nicht effizienter als der Cochrane-Orcutt Schätzer.
c	gehört zur Klasse der BLUE Schätzer (best-linear-unbiased estimator).
d	benutzt die erste Beobachtung nicht.

8.	Der Durbin-Watson Test
a	ist bei positiver Autokorrelation nicht durchführbar.
b	ist auch bei Schätzung ohne Konstante gültig.
c	eignet sich zum Testen auf Heteroskedastie.
d	ist auch in kleinen Stichproben gültig.

9.	Sie schätzen folgendes Modell: $y_i = \beta_1 + \beta_2 x + \nu_i$ , obwohl das wahre Modell $y_i = \gamma_1 + \gamma_2 x_i + \gamma_3 z_i + \epsilon_i$ lautet mit $\gamma_2 > 0$ und $\gamma_3 < 0$ . Sei $cov(x, z) < 0$ . Dann ist
a	$E[\hat{\beta}_2] = \beta_2$ .
b	$E[\hat{\beta}_2] > \beta_2$ .
c	$E[\hat{\beta}_2] < \beta_2$ .
d	$E[\hat{\beta}_2] = \beta_2 + \beta_3$ .

10.	Der folgende Plot der Residuen deutet ...
 <p>The plot shows residuals on the y-axis (ranging from -100 to 150) against months on the x-axis (ranging from 0 to 200). The residuals exhibit a strong positive linear trend, starting near zero and ending near 150, which is characteristic of a positive linear trend in the error term.</p>	
a	auf nicht korrelierte Fehler hin.
b	auf positive Autokorrelation der Fehler hin.
c	auf heteroskedastische Fehler hin.
d	auf homoskedastische Fehler hin.

11.	Schätzmodelle mit quadratischen, erklärenden Variablen
a	müssen mit nicht-linearen Schätzverfahren geschätzt werden.
b	können nicht für logarithmierte abhängige Variablen berechnet werden.
c	haben stets ein höheres $R^2$ als eine Schätzung ohne quadratische Terme.
d	können mit KQ geschätzt werden.

12.	Das korrigierte Bestimmtheitsmaß
a	berücksichtigt die zur Schätzung verwendeten Freiheitsgrade.
b	kann nie negativ sein.
c	entspricht dem Produkt des logarithmierten Störterms.
d	ist größer als das unkorrigierte Bestimmtheitsmaß.

13.	Ein konsistenter Schätzer
a	ist unverzerrt.
b	nähert sich für steigende Stichprobengrößen dem unbekanntem Populationsparameter an.
c	maximiert die Varianz des Störterms.
d	beschreibt einen kausalen Effekt.

14.	Das Logarithmieren einer rechtsschiefen abhängigen Variable
a	behebt das Vorliegen von Autokorrelation im Modell.
b	nähert die Verteilung der Variable einer Normalverteilung an.
c	erfordert die Schätzung nicht linearer Modelle.
d	ist auch bei Variablen mit negativen Ausprägungen möglich.

15.	Auf der ersten Stufe einer 2SLS-Schätzung
a	wird die endogene Variable auf die abhängige Variable der zweiten Stufe regressiert.
b	wird die exogene Variable auf das Instrument regressiert.
c	wird die exogene Variable auf die endogene Variable regressiert.
d	wird die endogene Variable auf das Instrument und die exogenen Variablen regressiert.

16.	Homoskedastie im linearen Regressionsmodell
a	hat einen Einfluss auf die Unverzerrtheit der Schätzer.
b	hat einen Einfluss auf die Effizienz der Schätzer.
c	hat keinen Einfluss auf die Varianz der Schätzer.
d	hat einen Einfluss auf den Punktschätzer.

17.	Im vorliegenden Modell $y_i = \beta_0 + \beta_1 \cdot x_i + \beta_2 \cdot x_i^2 + \epsilon_i$ mit $b_1 > 0$ und $b_2 < 0$ ist
a	der marginale Effekt von $x_i$ immer positiv.
b	das Maximum von $y_i$ unterschätzt.
c	der Zusammenhang zwischen $y_i$ und $x_i$ umgekehrt u-förmig.
d	das Modell nicht linear in den Parametern.

18.	Ein Konfidenzintervall
a	wird c.p. mit steigender Stichprobengröße breiter.
b	wird c.p. mit sinkender Varianz der Schätzer kleiner.
c	wird c.p. mit steigendem Koeffizienten schmaler.
d	wird c.p. mit steigendem Signifikanzniveau $\alpha$ breiter.

19.	Die Verletzung welcher Annahme schließt die kausale Interpretation einer Schätzung aus? (siehe Formelsammlung)
a	$\epsilon_i \approx N(0, \sigma^2 I_N)$ .
b	$V(\epsilon_i) = \sigma^2$ für $i=1,2,\dots,N$ .
c	$V(\epsilon X) = \sigma^2 \text{Diag}(h_i^2) = \sigma\Psi$ .
d	$E(x_i \epsilon_i) = 0$ .

20.	Das 95%-Konfidenzintervall des Schätzers $b_1$ lautet $[-0,5;0,5]$ . Welche der Aussagen ist korrekt?
a	$b_1$ ist mit einer Wahrscheinlichkeit von 5% größer als 0.
b	Es kann keine Aussage über $b_1$ getroffen werden.
c	$b_1$ ist auf dem 5%-Niveau nicht signifikant von Null verschieden.
d	$b_1$ ist auf dem 1%-Niveau signifikant von Null verschieden.

21.	Sie führen einen Chow-Test auf Geschlechterunterschiede für die KQ Schätzung des Modells $Lohn_i = \beta_1 + \beta_2 \cdot Mann_i + \beta_3 \cdot Bildung_i + \epsilon_i$ durch. Die Teststatistik des Chow-Tests hat den p-Wert 0,1542. Was können Sie daraus schließen?
a	Die Bildungsrendite von Männern und Frauen unterscheidet sich nicht signifikant.
b	Der mittlere Stundenlohn von Frauen und Männern unterscheidet sich signifikant.
c	Es gibt einen statistisch signifikanten Unterschied in den geschätzten Koeffizienten für $Bildung_i$ für Männer und Frauen bei einer Irrtumswahrscheinlichkeit von $\alpha = 0,01$ .
d	Männer verdienen c.p. im Mittel mehr als Frauen.

22.	Die irrelevante Variable $z$ in der KQ-Schätzung des Modells $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \epsilon_i$
a	führt zu inkonsistenten Koeffizientenschätzern.
b	erhöht das korrigierte Bestimmtheitsmaß $R^2$ .
c	führt dazu, dass das Modell nicht geschätzt werden kann.
d	erhöht die Varianz des Koeffizientenschätzers $b_2$ , wenn $cov(x, z) \neq 0$ .



23.	Sie schätzen folgendes Modell: $y_i = \beta_1 + \beta_2 x + \beta_3 x^2 + \nu_i$ . Für welche Parameterkonstellation ergibt sich ein U-förmiger Zusammenhang zwischen x und y:
a	$\beta_2 > 0$ und $\beta_3 < 0$ .
b	$\beta_2 > 0$ und $\beta_3 > 0$ .
c	$\beta_2 < 0$ und $\beta_3 < 0$ .
d	$\beta_2 < 0$ und $\beta_3 > 0$ .

24.	Bei einem linearen Wahrscheinlichkeitsmodell...
a	können keine konsistenten marginalen Effekte bestimmt werden.
b	können die vorhergesagten Werte der abhängigen Variable größer als 1 sein.
c	können die marginalen Effekte als odds ratio interpretiert werden.
d	keine der Antworten ist korrekt.

25.	Ein Typ-2-Fehler liegt vor, wenn
a	die nicht zutreffende Nullhypothese nicht verworfen wird.
b	die nicht zutreffende Nullhypothese verworfen wird.
c	die zutreffende Nullhypothese verworfen wird.
d	die zutreffende Nullhypothese nicht verworfen wird.

26.	Der P-Wert gibt...
a	den kleinstmöglichen Wert für den geschätzten Parameter an.
b	die Wahrscheinlichkeit, dass die Nullhypothese zutrifft an.
c	die Wahrscheinlichkeit für einen Fehler zweiter Art an.
d	das kleinstmögliche Signifikanzniveau, zu dem die Nullhypothese verworfen werden kann, an.

27.	Bei einer Instrumentvariablenschätzung spricht man von einem überidentifizierten Modell, wenn...
a	mehr Instrumente als Kontrollvariablen vorliegen.
b	mehr Instrumente als exogene Variablen vorliegen.
c	mehr Instrumente als abhängige Variablen vorliegen.
d	mehr Instrumente als endogene Variablen vorliegen.

28.	Sie schätzen folgendes Modell: $\text{Log}(\text{Lohn}_i) = \beta_1 + \beta_2 \cdot \text{Bildung}_i + \epsilon_i$ . $\text{Log}(\text{Lohn}_i)$ ist der logarithmierte Monatslohn und $\text{Bildung}_i$ ist die Anzahl der Bildungsjahre von Person $i$ . Es sei $b_2 = 0.5$ und $se(b_2) = 0.9$ Welche der folgenden Aussagen trifft zu? Der geschätzte Koeffizient der Bildung ist...
a	statistisch signifikant.
b	unbekannt.
c	ökonomisch signifikant.
d	eine Elastizität.

29.	Welche der folgenden Aussagen ist nicht korrekt? Umgekehrte Kausalität...
a	führt zu inkonsistenten Parameterschätzungen.
b	führt zu endogenen erklärenden Variablen.
c	liegt dann vor, wenn die $X'X$ -Matrix nicht invertiert werden kann.
d	liegt dann vor, wenn die abhängige Variable eine der erklärenden Variablen beeinflusst.

30.	Bei Multikollinearität, im Vergleich zu einer Situation ohne Multikollinearität,...
a	sind die Standardfehler klein.
b	sind die Konfidenzintervalle breit.
c	werden die KQ-Koeffizienten inkonsistent geschätzt.
d	ist die Wahrscheinlichkeit für einen Typ-2-Fehler klein.

## 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

# 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*

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