

Bachelorprüfung im Wintersemester 2012/2013

Fach: Empirische Wirtschaftsforschung II

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

Name, Vorname	
Matrikelnr.	
Studiengang	
Semester	
Datum	
Raum	
Unterschrift	

Vorbemerkungen:

Anzahl der Aufgaben: Die Klausur besteht aus 6 Aufgaben, die alle bearbeitet werden müssen.

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

Erlaubte Hilfsmittel:

- Tabellen der statistischen Verteilung und Liste der Annahmen (sind der Klausur beigelegt)
- 1 DIN-A4-Seite mit Notizen
- Taschenrechner
- Fremdwörterbuch

Wichtige Hinweise:

- Sollte es vorkommen, dass die statistischen Tabellen, die dieser Klausur beiliegen, den exakten 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:**[9 Punkte]**

- a) Was versteht man unter Heteroskedastie? (1 Punkt)
- b) Welche beiden ungünstigen Auswirkungen hat Heteroskedastie auf KQ-Schätzergebnisse? (2 Punkte)
- c) Nennen Sie zwei Verfahren, mit Heteroskedastie umzugehen. Welche der in b) angesprochenen Probleme werden durch die beiden Verfahren jeweils gelöst, welche nicht? (6 Punkte)

Aufgabe 2:**[18 Punkte]**

Sie interessieren sich für die Determinanten der Lohnhöhe. Sie verfügen über Paneldaten von 500 Vollzeit arbeitenden Personen aus den Jahren 2002, 2003 und 2004 mit insgesamt 1500 Beobachtungen. Ihr Datensatz enthält folgende Variablen:

lohn = *Bruttostundenlohn in €*

alter = *Alter in Jahren*

mann = *1 wenn männlich, 0 wenn weiblich*

ausb = *Schul- und Hochschulausbildung in Jahren*

single = *misst den Familienstand; 1, wenn Person alleinstehend ist, sonst 0.*

Sie möchten untersuchen, ob der Familienstand mit dem Lohn zusammenhängt und schätzen folgendes Modell mit KQ:

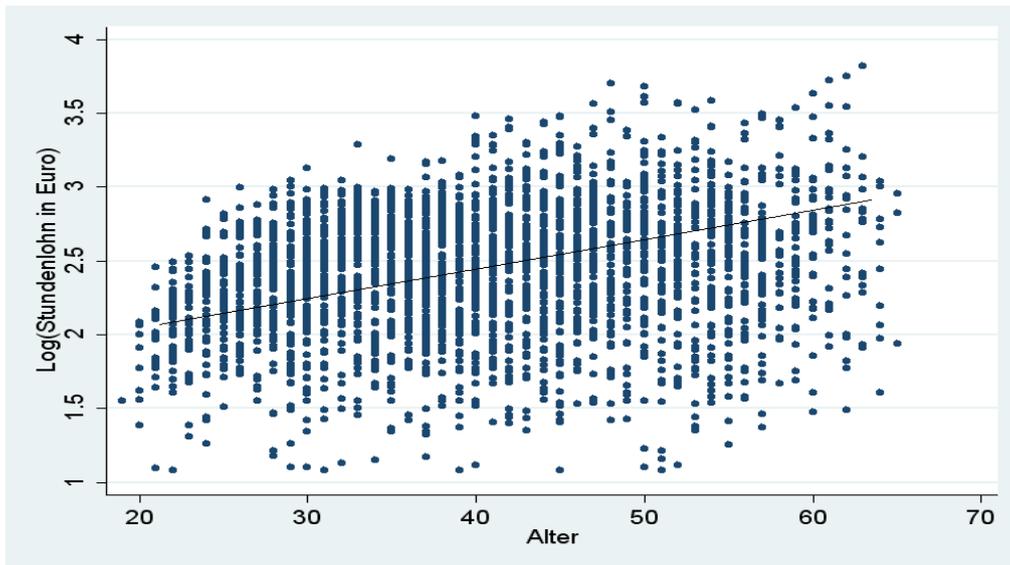
$$\text{lohn}_{it} = \beta_0 + \beta_1 \text{alter}_{it} + \beta_2 \text{single}_{it} + \beta_3 \text{mann}_{it} + \beta_4 \text{ausb}_{it} + u_{it}$$

mit i als Personenindikator und t als Zeitindikator. $u_{it} = a_i + \varepsilon_{it}$ beinhaltet unbeobachtete Faktoren.

- a) Was ist der Unterschied zwischen einem balanced und einem unbalanced Panel? Welche der beiden Alternativen trifft auf Ihren Datensatz zu? Erläutern Sie Ihre Antwort kurz. (3 Punkte)
- b) Könnte der Koeffizient von *single* mit KQ inkonsistent geschätzt werden, falls die Panelstruktur ignoriert wird? Begründen Sie Ihre Antwort. (2 Punkte)
- c) Gehen Sie davon aus, dass a_i nicht mit den erklärenden Variablen korreliert ist. Erklären Sie, warum bei der Schätzung der gegebenen Spezifikation mit KQ bzgl. der Standardfehler ein Problem auftritt. Nennen Sie zwei alternative Verfahren, um das Problem zu beheben. (4 Punkte)
- d) Schreiben Sie für das gegebene Beispiel die Schätzgleichung für einen Within-Schätzer sowie für einen First-Difference-Schätzer auf. Unter welchen Umständen können Sie mit Hilfe dieser Schätzansätze β_2 konsistent schätzen? (5 Punkte)
- e) Gehen Sie nun davon aus, dass u_{it} einem Random Walk folgt, d.h. $u_{it} = u_{it-1} + \varepsilon_{it}$ mit ε_{it} i.i.d. Welches Schätzverfahren würden Sie vorziehen, Within- oder First-Difference-Schätzer? Erläutern Sie Ihre Antwort ausführlich und begründen Sie Ihre Entscheidung. (4 Punkte)

Aufgabe 3:**[11 Punkte]**

Die nachfolgende Grafik stellt den Zusammenhang zwischen logarithmiertem Stundenlohn $\log(\text{lohn})$ und Lebensalter *alter* für die Grundgesamtheit dar:



Sie verfügen über eine Stichprobe aus dieser Grundgesamtheit und möchten damit folgendes Modell mit KQ schätzen:

$$\log(\text{lohn}_i) = \beta_0 + \beta_1 \text{alter}_i + u_i$$

Gehen Sie nachfolgend davon aus, dass der funktionale Zusammenhang zwischen abhängiger und erklärender Variable im Modell korrekt spezifiziert ist und $\beta_1 > 0$.

- Ist bei einer KQ-Schätzung dieses Zusammenhangs $E(\hat{\beta}_1)$ größer als, kleiner als oder gleich β_1 , wenn bei der Messung des Lohns ein zufälliger Messfehler auftritt? Begründen Sie. (2 Punkte)
- Ist bei einer KQ-Schätzung dieses Zusammenhangs $E(\hat{\beta}_1)$ größer als, kleiner als oder gleich β_1 , wenn bei der Messung des Alters ein zufälliger Messfehler auftritt? Begründen Sie. (2 Punkte)
- In Ihrer Stichprobe fehlen alle Beobachtungen mit $\log(\text{lohn}_i) > 3$.
 - Liegt endogene oder exogene Stichprobenselektion vor? Begründen Sie kurz. (2 Punkte)
 - Ist bei einer KQ-Schätzung mit dieser Stichprobe $E(\hat{\beta}_1)$ größer als, kleiner als oder gleich β_1 ? Begründen Sie. (2 Punkte)
- In Ihrer Stichprobe fehlen alle Beobachtungen für Personen die älter als 50 Jahre sind. Liegt endogene oder exogene Stichprobenselektion vor? Begründen Sie kurz. Welche Bedingung muss gelten, damit die Schätzung trotz Selektion unverzerrt ist? (3 Punkte)

Aufgabe 4:

[7 Punkte]

- Was ist mit "serieller Korrelation im Störterm einer linearen Regression" gemeint? (1 Punkt)
- Welche Auswirkungen hat ein autoregressiver Prozess im Störterm des Modells $y_t = \beta_0 + \beta_1 x_t + e_t$ für eine Schätzung mit KQ, wenn die Annahmen TS.1' bis TS.3' erfüllt sind? (2 Punkte)
- Unterstellen Sie nun, dass der Störterm des Regressionsmodells $y_t = \beta_0 + \beta_1 x_t + e_t$ dem Prozess $e_t = \rho e_{t-1} + v_t$ folgt und v_t die Gauss-Markov Annahmen erfüllt. Geben Sie die transformierte Schätzgleichung für das Cochrane-Orcutt Verfahren an. Erläutern Sie kurz das Ziel des Verfahrens und warum dieses erreicht wird. (3 Punkte)
- Welchen Nachteil hat das Cochrane-Orcutt Verfahren im Vergleich zum Prais-Winsten Verfahren? (1 Punkt)

Aufgabe 5:**[16 Punkte]**

Eine Justizvollzugsanstalt (JVA) wurde im Jahr 2007 fertiggestellt. Nach der Fertigstellung behaupten die im Umkreis von 10km wohnenden Hausbesitzer, dass der Preis Ihrer Häuser aufgrund des Baus der JVA gesunken sei und fordern eine Entschädigung. Ihnen liegen Daten über regionale Hauspreise aus dem Jahr 2004 (zu diesem Zeitpunkt wussten die Bewohner nicht, dass die JVA gebaut werden soll) und aus dem Jahr 2010 vor. Sie schätzen folgendes Modell mit KQ:

$$\ln(\text{preis}) = \beta_0 + \beta_1 d2010 + \beta_2 \text{nah} + \beta_3 (d2010 * \text{nah}) + u.$$

preis = Preis des Hauses in Euro

d2010 = 1, wenn die Beobachtung aus dem Jahr 2010 ist, sonst = 0

nah = 1, wenn die Beobachtung zu einem Haus im Umkreis von 10km um die JVA gehört, sonst = 0

Sie erhalten folgendes Regressionsergebnis (Standardfehler in Klammern):

$$\beta_0 = 11,51 (2,22), \beta_1 = 0,01 (0,05), \beta_2 = -0,08 (0,02), \beta_3 = -0,04 (0,03)$$

- Wie heißt das hier verwendete KQ-Schätzverfahren und welche zentrale Annahme muss neben den Gauss-Markov Annahmen gelten, damit Sie den Effekt der JVA identifizieren können? (2 Punkte)
- Erläutern Sie für das gegebene Beispiel kurz die Begriffe "Treatmentgruppe" und "Kontrollgruppe". (2 Punkte)
- Interpretieren Sie die geschätzten Parameter für β_1 und β_2 inhaltlich und statistisch. (6 Punkte)
- Hat der Bau der JVA die Hauspreise im 10km Radius um die Anlage negativ beeinflusst? Begründen Sie Ihre Antwort. (3 Punkte)
- Unter welchen Umständen beeinflusst die Berücksichtigung weiterer Kontrollvariablen in der Schätzgleichung den geschätzten Wert für β_3 ? (3 Punkte)

Aufgabe 6:**[29 Punkte]**

Welche Antwort ist richtig? Kreuzen Sie nur **eine Antwort** pro Aufgabe an. Falls mehrere Aussagen korrekt sind, kreuzen Sie **nur** die entsprechende **Antwortkombination** an. Für jede richtige Antwort gibt es 1 Punkt. Für falsche Antworten werden keine Punkte abgezogen.

1.	Paneldatenmodelle	
a	<input type="checkbox"/>	können nur mit Regressionskonstante geschätzt werden.
b	<input type="checkbox"/>	können auch auf Datenstrukturen ohne zeitliche Dimension angewendet werden, z.B. Clusterstichproben.
c	<input type="checkbox"/>	sind nur für Balanced Panels konsistent.
d	<input type="checkbox"/>	a und b.

2.	Der Fixed Effects Schätzer	
a	<input type="checkbox"/>	ist BLUE, falls Paneldaten verwendet werden.
b	<input type="checkbox"/>	ist inkonsistent, wenn ein Regressor mit dem zeitinvarianten Teil des Störterms korreliert.
c	<input type="checkbox"/>	ist inkonsistent, wenn ein Regressor mit dem zeitvarianten Teil des Störterms korreliert.
d	<input type="checkbox"/>	Keine der genannten Antworten.

3.	Der Fixed Effects Schätzer	
a	<input type="checkbox"/>	ist stets effizienter als der Random Effects Schätzer.
b	<input type="checkbox"/>	erfordert serielle Korrelation in den zeitkonstanten Variablen.
c	<input type="checkbox"/>	liefert umso präzisere Schätzungen je geringer die Within-Variation der Regressoren ist.
d	<input type="checkbox"/>	Keine der genannten Antworten.
4.	Kovarianzstationarität eines Prozesses x_t impliziert, dass	
a	<input type="checkbox"/>	x_t strikt stationär ist.
b	<input type="checkbox"/>	die Varianz von x_t für alle Zeitpunkte gleich ist.
c	<input type="checkbox"/>	die Kovarianz zwischen Ausprägungen gleich weit entfernter Zeitpunkte konstant ist.
d	<input type="checkbox"/>	b und c.
5.	Folgt der Störterm in einem Zeitreihenmodell einem Random Walk	
a	<input type="checkbox"/>	sollte das Modell in ersten Differenzen geschätzt werden.
b	<input type="checkbox"/>	liegt Autokorrelation vor.
c	<input type="checkbox"/>	ist der KQ Schätzer BLUE.
d	<input type="checkbox"/>	a und b.
6.	Eine zur Lösung des Problems ausgelassener Variablen verwendete Proxy-Variable sollte	
a	<input type="checkbox"/>	mit der ausgelassenen Variable möglichst hoch korreliert sein.
b	<input type="checkbox"/>	möglichst hoch mit den übrigen erklärenden Variablen des Modells korreliert sein.
c	<input type="checkbox"/>	mit allen Regressoren des ursprünglichen Modells möglichst hoch korreliert sein.
d	<input type="checkbox"/>	b und c.
7.	Heteroskedastie	
a	<input type="checkbox"/>	kann bei i.i.d. Störtermen auftreten.
b	<input type="checkbox"/>	kann zusammen mit Autokorrelation auftauchen.
c	<input type="checkbox"/>	führt zu verzerrten Punktschätzern.
d	<input type="checkbox"/>	b und c.
8.	Der RESET Test	
a	<input type="checkbox"/>	basiert auf einer Regression der abhängigen Variable auf ein Polynom der vorhergesagten abhängigen Variable.
b	<input type="checkbox"/>	testet auf Fehlspezifikation der funktionalen Form.
c	<input type="checkbox"/>	gibt keine Hinweise auf die Ursache einer Fehlspezifikation.
d	<input type="checkbox"/>	Alle genannten Antworten.
9.	$E(u_i X) = 0$ impliziert	
a	<input type="checkbox"/>	$E(u_i) = 0$.
b	<input type="checkbox"/>	$E(u_i u_j) = 0$ für $i \neq j$.
c	<input type="checkbox"/>	$Var(u_i) > 0$.
d	<input type="checkbox"/>	Alle genannten Antworten.
10.	Instrumentvariablen	
a	<input type="checkbox"/>	sollen möglichst hoch mit dem endogenen Regressor korreliert sein.
b	<input type="checkbox"/>	sollen möglichst hoch mit dem Störterm korreliert sein.
c	<input type="checkbox"/>	dürfen nicht mit den exogenen Regressoren korreliert sein.
d	<input type="checkbox"/>	a und c.

11.	Bei Fehlspezifikation der funktionalen Form	
a	<input type="checkbox"/>	müssen Panelschätzverfahren eingesetzt werden.
b	<input type="checkbox"/>	sollte der RESET Test die Alternativhypothese verwerfen.
c	<input type="checkbox"/>	sollte der Breusch-Pagan Test die Nullhypothese verwerfen.
d	<input type="checkbox"/>	Keine der genannten Antworten.

12.	$Cov(x_t, x_{t+1}) = Cov(x_t, x_{t-1})$ gilt stets, sobald	
a	<input type="checkbox"/>	strikte Stationarität vorliegt.
b	<input type="checkbox"/>	der KQ-Schätzer konsistent ist.
c	<input type="checkbox"/>	y_t seriell korreliert ist.
d	<input type="checkbox"/>	a und c.

13.	Die Annahme TS.5: $corr(u_t, u_s X) = 0$ mit $t \neq s$,	
a	<input type="checkbox"/>	ist im Querschnittsfall nicht erforderlich, wenn die Annahme der Zufallsstichprobe gilt.
b	<input type="checkbox"/>	bedeutet, dass die Störterme verschiedener Zeitpunkte unkorreliert sind.
c	<input type="checkbox"/>	ist notwendig für die Gültigkeit der Inferenzverfahren nach einer KQ-Schätzung.
d	<input type="checkbox"/>	Alle der genannten Antworten.

14.	Ist die zeitkonstante, unbeobachtete Heterogenität a_i unkorreliert mit der erklärenden Variablen x_{it} , d.h. $Cov(a_i, x_{it}) = 0$, dann	
a	<input type="checkbox"/>	ist der KQ-Schätzer effizient.
b	<input type="checkbox"/>	ist der Random Effects Schätzer effizienter als der Fixed Effects Schätzer.
c	<input type="checkbox"/>	ist der KQ Schätzer inkonsistent.
d	<input type="checkbox"/>	ist der Random Effects Schätzer inkonsistent.

15.	White-Test und Breusch-Pagan Test unterscheiden sich im multiplen Regressionsmodell	
a	<input type="checkbox"/>	in der Annahme über die funktionalen Form der Heteroskedastie.
b	<input type="checkbox"/>	in der abhängigen Variable in der Hilfsregression.
c	<input type="checkbox"/>	in der Anzahl der Regressoren in der Hilfsregression.
d	<input type="checkbox"/>	a und c.

16.	Endogene Stichprobenselektion	
a	<input type="checkbox"/>	bedeutet systematische Stichprobenselektion auf Basis exogener erklärender Variablen.
b	<input type="checkbox"/>	erhöht die Stichprobengröße.
c	<input type="checkbox"/>	kann zu inkonsistenten Parameterschätzern führen.
d	<input type="checkbox"/>	erhöht die Effizienz der Schätzung.

17.	Attenuation bias	
a	<input type="checkbox"/>	ist eine Folge von Heteroskedastie.
b	<input type="checkbox"/>	kann bei Messfehlern in der erklärenden Variable auftreten.
c	<input type="checkbox"/>	ist eine Folge von Autokorrelation.
d	<input type="checkbox"/>	ist eine Folge von Multikollinearität.

18.	Im linearen Wahrscheinlichkeitsmodell	
a	<input type="checkbox"/>	ist die erklärende Variable eine Dummy Variable.
b	<input type="checkbox"/>	kann Heteroskedastie im Störterm mit dem Prais-Winsten Verfahren korrigiert werden.
c	<input type="checkbox"/>	tritt zwangsläufig Heteroskedastie auf.
d	<input type="checkbox"/>	können bei einer FGLS Schätzung zur Behebung von Heteroskedastie nie alle Beobachtungen verwendet werden.

19.	Der Between Schätzer für Paneldaten mit Störterm $u_{it} = a_i + e_{it}$	
a	<input type="checkbox"/>	ist verzerrt, wenn die erklärenden Variablen und a_i korreliert sind.
b	<input type="checkbox"/>	und der LSDV Schätzer liefern identische Schätzergebnisse.
c	<input type="checkbox"/>	ist effizienter als eine gepoolte KQ Schätzung.
d	<input type="checkbox"/>	a und b.
20.	Ist MLR.1 bis MLR.5 (siehe Anhang) erfüllt, so ist	
a	<input type="checkbox"/>	KQ effizient.
b	<input type="checkbox"/>	FGLS effizient.
c	<input type="checkbox"/>	Fixed Effects Schätzung effizient.
d	<input type="checkbox"/>	Alle genannten Antworten.
21.	Das LAD Verfahren	
a	<input type="checkbox"/>	schätzt Effekte am Median der auf X bedingten Verteilung von y.
b	<input type="checkbox"/>	ist robuster gegenüber einzelnen Ausreißerbeobachtungen als der KQ Schätzer.
c	<input type="checkbox"/>	minimiert die Summe der quadrierten Absolutwerte der Residuen.
d	<input type="checkbox"/>	a und b.
22.	Im Fall seriell korrelierter Störterme	
a	<input type="checkbox"/>	kann $\hat{\beta}$ fälschlicherweise als insignifikant ausgewiesen werden.
b	<input type="checkbox"/>	kann $\hat{\beta}$ fälschlicherweise als signifikant ausgewiesen werden.
c	<input type="checkbox"/>	ist $\hat{\beta}$ verzerrt.
d	<input type="checkbox"/>	a und b.
23.	Der Prozess $u_t = e_t - 0,99e_{t-1}$ mit $e_t \sim i. i. d. (0, \sigma^2)$	
a	<input type="checkbox"/>	ist schwach abhängig.
b	<input type="checkbox"/>	ist kovarianzstationär.
c	<input type="checkbox"/>	ist asymptotisch unkorreliert.
d	<input type="checkbox"/>	Alle genannten Antworten.
24.	Gepoolte Querschnittsdaten können	
a	<input type="checkbox"/>	die Anwendung eines Within Schätzers ermöglichen.
b	<input type="checkbox"/>	die Anwendung des Difference-in-Difference-Schätzers erlauben.
c	<input type="checkbox"/>	die Anwendung eines First Difference Schätzers ermöglichen.
d	<input type="checkbox"/>	bei der Anwendung von KQ zu heteroskedastischen Störtermen führen.
25.	Periodendummies in Fixed Effects Schätzungen	
a	<input type="checkbox"/>	erlauben die Schätzung periodenspezifischer Achsenabschnitte.
b	<input type="checkbox"/>	haben keine Within-Variation.
c	<input type="checkbox"/>	führen nur bei einem Balanced Panel zu Autokorrelation im Störterm.
d	<input type="checkbox"/>	erlauben die Schätzung individuenspezifischer Achsenabschnitte.
26.	Der Vorteil des Durbin-Watson Tests besteht in	
a	<input type="checkbox"/>	seiner Fähigkeit auf Autokorrelation höherer Ordnung testen zu.
b	<input type="checkbox"/>	seiner exakten Gültigkeit auch in kleinen Stichproben.
c	<input type="checkbox"/>	seiner exakten Gültigkeit auch bei endogenen Regressoren.
d	<input type="checkbox"/>	b und c.

27.	Folgt der Störterm einem Moving Average Prozess $u_t = e_t - 0,1e_{t-1}$ mit $e_t \sim i. i. d. (0, \sigma^2)$,	
a	<input type="checkbox"/>	kann der Prozess nicht mehr schwach abhängig sein.
b	<input type="checkbox"/>	sind u_t und u_{t-1} positiv korreliert.
c	<input type="checkbox"/>	beträgt die Durbin-Watson Teststatistik ungefähr zwei.
d	<input type="checkbox"/>	Keine der genannten Antworten.

28.	Schwache Abhängigkeit	
a	<input type="checkbox"/>	impliziert Kovarianzstationarität.
b	<input type="checkbox"/>	impliziert Stationarität.
c	<input type="checkbox"/>	gilt für $u_t = 0,01u_{t-1} + e_t$, wobei $e_t \sim i. i. d. (0, \sigma^2)$.
d	<input type="checkbox"/>	Keine der genannten Antworten.

29.	Sei $y_i = \beta x_i + u_i$. Das Auslassen der Variable z_i führt stets zu einem verzerrten $\hat{\beta}$, falls	
a	<input type="checkbox"/>	z_i mit y_i korreliert ist.
b	<input type="checkbox"/>	z_i mit x_i korreliert ist.
c	<input type="checkbox"/>	z_i mit u_i korreliert ist.
d	<input type="checkbox"/>	Keine der genannten Antworten.

Annahmen im linearen Regressionsmodell

Einfaches Modell

SLR.1 Das Bevölkerungsmodell ist linear in den Parametern

SLR.2 Die Stichprobe ist zufällig

SLR.3 Die Realisationen von x_i in der Stichprobe sind nicht alle identisch

SLR.4 $E(u|x) = 0$

SLR.5 $\text{Var}(u|x) = \sigma^2$

Multiples Modell

MLR.1 Das Bevölkerungsmodell ist linear in den Parametern

MLR.2 Die Stichprobe ist zufällig

MLR.3 Keine perfekte Kollinearität

MLR.4 $E(u|x_1, \dots, x_k) = E(u) = 0$

MLR.5 $\text{Var}(u|x_1, \dots, x_k) = \text{Var}(u) = \sigma^2$

MLR.6 $u \sim \text{Normal}(0, \sigma^2)$

Zeitreihenmodelle

TS.1 Der stochastische Prozess ist linear in den Parametern

TS.2 Keine perfekte Kollinearität

TS.3 $E(u_t | X) = 0, t=1, 2, \dots, n$

TS.4 $\text{Var}(u_t | X) = \text{Var}(u_t) = \sigma^2, t = 1, 2, \dots, n$

TS.5 $\text{corr}(u_t, u_s | X) = 0, t \neq s$

TS.6 $u_t \sim N(0, \sigma^2)$

TS.1' Der stochastische Prozess ist stationär, schwach abhängig und linear in den Parametern

TS.2' Keine perfekte Kollinearität

TS.3' $E(u_t | x_t) = 0$

TS.4' $\text{Var}(u_t | x_t) = \sigma^2$

TS.5' $E(u_s u_t | x_t, x_s) = 0$ für alle $t \neq s$.