Masterprüfung SS 2017 - MUSTERLÖSUNG

Fach: Ökonometrie

Prüfer: 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 beigefügt) • Tabellen der statistischen Verteilungen (sind der Klausur beigefügt) • 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: **[7 Punkte]**

In Musterstadt gibt es seit einem Monat drei neue Autohändler. Alle haben zur Eröffnung Werbeplakate aufgehängt. Der erste hat ein Plakat aufgehängt und hat seitdem 5 Autos verkauft. Der zweite hat ebenfalls ein Plakat aufgehängt und hat 4 Autos verkauft. Der dritte hat zwei Plakate aufgehängt und hat 8 Autos verkauft. Sie regressieren die Anzahl verkaufter Autos auf die Anzahl der Plakate mit folgendem Modell:

$$Verkaufte\_Autos_i = \beta_1 + \beta_2 \cdot Anzahl\_Plakate_i + \varepsilon_i$$

1.1 Berechnen Sie in Matrizenschreibweise mithilfe der KQ-Methode die geschätzten Koeffizienten b_1 und b_2. [6 Punkte]

- $(X'X)^{-1}X'y = \mathbf{b}$
- $y = \begin{pmatrix} 5 \\ 4 \\ 8 \end{pmatrix}$ und $X = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}$
- $X'X = \begin{pmatrix} 3 & 4 \\ 4 & 6 \end{pmatrix}$
- $(X'X)^{-1} = \frac{1}{3 \cdot 6 - 4 \cdot 4} \cdot \begin{pmatrix} 6 & -4 \\ -4 & 3 \end{pmatrix} = \begin{pmatrix} 3 & -2 \\ -2 & 1,5 \end{pmatrix}$
- $X'y = \begin{pmatrix} 17 \\ 25 \end{pmatrix}$
- $(X'X)^{-1}X'y = \begin{pmatrix} 3 & -2 \\ -2 & 1,5 \end{pmatrix} \cdot \begin{pmatrix} 17 \\ 25 \end{pmatrix} = \begin{pmatrix} 1 \\ 3,5 \end{pmatrix} \rightarrow b_1 = 1 \textit{ und } b_2 = 3,5$

1.2 Interpretieren Sie den geschätzten Koeffizienten b_2 inhaltlich. [1 Punkt]
Hinweis: Sollten Sie in Aufgabe 1.1 keine Lösung gefunden haben, verwenden Sie $b_2 = 2$.

- Ein Plakat hat c.p. im Mittel den Verkauf von 3,5 Autos bewirkt.

Aufgabe 2: **[20 Punkte]**

Eine Krankenkasse beauftragt Sie, die Determinanten der Häufigkeit von Arztbesuchen ihrer Versicherten zu analysieren. Ihnen stehen dazu Daten aus einer Befragung von 3670 Versicherten im Jahre 2016 mit folgenden Variablen zur Verfügung:

$ln\_dvisits_i$	Anzahl der Arztbesuche von Person i im Jahr 2016 (logarithmiert)
age_i	Alter der Person i in Jahren
$female_i$	Dummy-Variable, =1, wenn Person i weiblich, =0 sonst
$hstatus15_i$	Selbsteinschätzung des Gesundheitszustandes von Person i für das Jahr 2015 (Skala von 1-5)
$hstatus16_i$	Selbsteinschätzung des Gesundheitszustandes von Person i für das Jahr 2016 (Skala von 1-5)
$eduhigh_i$	Dummy-Variable, =1, wenn Person i eine hohe Bildung (mind. 13 Bildungsjahre) hat, =0 sonst

Sie stellen folgendes Regressionsmodell auf und schätzen es anschließend mit Stata:

$$ln\_dvisits_i = \beta_1 + \beta_2 age_i + \beta_3 female_i + \beta_4 age\_female_i + \beta_5 hstatus16_i + \beta_6 eduhigh_i + \varepsilon_i \quad \text{(Modell I)}$$

wobei $age\_female_i = age_i \cdot female_i$ gilt.

```
      Source |       SS       df       MS              Number of obs =    3670
-------------+------------------------------           F(  5,  3664) =  164.96
       Model |  397.894727     5  79.5789453           Prob > F      =  0.0000
    Residual |  1767.55976  3664  .482412599           R-squared     =  0.1837
-------------+------------------------------           Adj R-squared =  0.1826
       Total |  2165.45449  3669  .590202913           Root MSE      =  .69456

------------------------------------------------------------------------------
 ln_divisits |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0048175    .001016     4.74   0.000     .0028256    .0068094
      female |   .3289255   .0735732     4.47   0.000     .1846769     .473174
  age_female |  -.0042282    .001311    -3.23   0.001    -.0067986   -.0016578
  hstatus16 |   .3233856   .0132505    24.41   0.000     .2974064    .3493647
     eduhigh |    .068207   .0279707     2.44   0.015     .0133674    .1230467
       _cons |   1.104396   .0617708    17.88   0.000      .983287    1.225504
------------------------------------------------------------------------------
```

Runden Sie im Folgenden alle Zahlenangaben auf die dritte Nachkommastelle.

2.1 Interpretieren Sie den geschätzten Koeffizienten b_6 inhaltlich und statistisch. [2 Punkte]

- Die Anzahl an Arztbesuchen im Jahr 2016 ist für Versicherte mit hoher Bildung c.p. im Mittel um ca. 6,821% höher als bei nicht hoch gebildeten Versicherten.
- Der Koeffizient ist statistisch signifikant auf dem 5% Niveau.

2.2 Leiten Sie den marginalen Effekt der Variable *age* auf die Anzahl an Arztbesuchen allgemein her. Wie groß ist der marginale Effekt für Frauen? Wie groß ist er für Männer? Interpretieren Sie die Ergebnisse. [4 Punkte]

- Marginaler Effekt allgemein:

$$\frac{\Delta E(ln\_dvisits_i)}{\Delta age_i} = \beta_2 + \beta_4 \cdot female_i$$

$$\text{Alternativ}: \frac{\Delta E(dvisits_i)/dvisits_i}{\Delta age_i} = \beta_2 + \beta_4 \cdot female_i$$

- Marginaler Effekt für Frauen:

$$b_2 + b_4 \cdot 1 = 0{,}005 - 0{,}004 = 0{,}001\,(0{,}1\%)$$

Für Frauen führt ein Anstieg des Alters um ein Jahr c.p. im Mittel zu einem Anstieg der Arztbesuche um ca. 0,1%.

- Marginaler Effekt für Männer:

$$b_2 + b_4 \cdot 0 = 0{,}005\,(0{,}5\%)$$

Für Männer führt ein Anstieg des Alters um ein Jahr c.p. im Mittel zu einem Anstieg der Arztbesuche um ca. 0,5%.

2.3 Berechnen Sie die erwartete Anzahl an Arztbesuchen eines 30-jährigen Mannes mit hoher Bildung, der seinen Gesundheitszustand 2016 mit 4 bewertet hat. [4 Punkte]

- Vorhersagewert für $ln\_dvisits_i$:

$$\widehat{ln\_dvisits_i} = b_1 + b_2 \cdot 30 + b_3 \cdot 0 + b_4 \cdot 0 + b_5 \cdot 4 + b_6 \cdot 1$$

$$= 1{,}104 + 0{,}005 \cdot 30 + 0{,}329 \cdot 0 - 0{,}004 \cdot 0 + 0{,}323 \cdot 4 + 0{,}068 \cdot 1 = 2{,}614$$

- Vorhersagewert für $dvisits_i$:

$$\widehat{dvisits}_i = exp(ln\_\widehat{dvisits}_i + 0,5 \cdot \hat{\sigma}^2) = exp(2,614 + 0,5 \cdot 0,695^2) = 17,383$$

2.4 Sie vermuten, dass sich gleichzeitig mit dem Effekt von *age* auch der Effekt von *hstatus*16 zwischen Männern und Frauen unterscheidet und führen einen entsprechenden F-Test auf Geschlechterunterschiede in beiden Effekten auf dem 1%-Signifikanzniveau durch. Sie erhalten folgenden Wert der Teststatistik: $F = 6,10$.

Stellen Sie das Modell, das dem F-Test unterliegt, nachvollziehbar dar. Geben Sie anschließend die Null- und Alternativhypothese, den kritischen Wert und das Testergebnis für den durchgeführten F-Test an. [4 Punkte]

- Modell:
$$ln\_dvisits_i = \beta_1 + \beta_2 age_i + \beta_3 female_i + \beta_4 age\_female_i + \beta_5 hstatus16_i + \beta_6 eduhigh_i + \beta_7 hstatus16\_female_i + \varepsilon_i$$
- Hypothesen: $H_0: \beta_4 = \beta_7 = 0$; H_1: Mind. einer der Koeffizienten ungleich Null.
- Kritischer Wert: $F_{2;3663;1\%} = 4,63$.
- Testergebnis: Da $F_{empirisch} = 6,10 > 4,63 = F_{kritisch}$ kann die Nullhypothese auf dem 1% Signifikanzniveau verworfen werden. Die Koeffizienten sind gemeinsam statistisch signifikant.

2.5 Ein Kommilitone behauptet, dass Ihre Variable *hstatus*16 endogen ist. Nennen Sie einen möglichen Grund für das Vorliegen der Endogenität und erläutern Sie diesen. [1,5 Punkte]

- Mögliche Gründe (1 Nennung wird bepunktet):
  1. Umgekehrte Kausalität (reverse causality bias):
     Die Selbsteinschätzung des Gesundheitszustands für das Jahr 2016 hängt von der Anzahl an Arztbesuchen im Jahr 2016 ab.
  2. Ausgelassene erklärende Variable (omitted variable bias):
     Faktoren, die mit *hstatus*16 sowie *ln_dvisits* korrelieren, sind nicht im Modell enthalten.

2.6 Ihr Kommilitone rät Ihnen, Ihr Endogenitätsproblem durch Instrumentierung von *hstatus*16 zu lösen. Er schlägt vor, die Selbsteinschätzung des Gesundheitsstatus aus dem Vorjahr (*hstatus*15) als Instrument in einer two-stage-least-squares (2SLS)-Schätzung Ihres Modells I zu verwenden.

2.6.1 Stellen Sie die für die 2SLS-Schätzung benötigten Modellgleichungen auf. Nennen Sie einen Fall, in dem *hstatus*15 kein geeignetes Instrument darstellt. [3 Punkte]

1.Stufe:

$$hstatus16_i = \theta_1 + \theta_2 age_i + \theta_3 female_i + \theta_4 age\_female_i + \theta_5 eduhigh_i + \theta_6 hstatus15_i + \upsilon_i$$

2.Stufe:

$$ln(dvisits_i) = \beta_1 + \beta_2 age_i + \beta_3 female_i + \beta_4 age\_female_i + \beta_5 \widehat{hstatus16}_i + \beta_6 eduhigh_i + u_i$$

- *hstatus*15 korreliert nicht mit *hstatus*16.
  Alternativ:
- *hstatus*15 korreliert mit ε.

2.6.2 Wie können Sie testen, ob es sich bei *hstatus*15 um ein schwaches Instrument handelt? Geben Sie die Nullhypothese des Tests an. [1,5 Punkte]

- F-Test auf Signifikanz von *hstatus*15 auf der ersten Stufe der 2SLS-Schätzung.
- Nullhypothese: $H_0 : \theta_6 = 0$

Aufgabe 3: **[20 Punkte]**

Sie interessieren sich für die Determinanten der am Saisonende erreichten Punktzahl des 1. FC Nürnberg. Dazu liegt ein Datensatz mit Informationen von 1996 bis 2017 vor:

$points_t$	Am Saisonende erreichte Punktzahl des 1. FC Nürnberg im Jahr t
$value_t$	Marktwert aller Spieler des 1. FC Nürnberg in Millionen € im Jahr t
$possession_t$	Durchschnittlicher Ballbesitz in Prozent (1-100) pro Spiel im Jahr t
$coach_t$	Dummy-Variable, =1, wenn es im Jahr t zu einem Trainerwechsel beim 1. FC Nürnberg kam, =0 sonst
$bundesliga_t$	Dummy-Variable, =1, wenn der 1. FC Nürnberg im Jahr t in der 1.Bundesliga gespielt hat, =0 sonst

Es wird folgendes Regressionsmodell aufgestellt und anschließend mit Stata geschätzt:

$$points_t = \beta_1 + \beta_2 value_t + \beta_3 possession_t + \beta_4 coach_t + \beta_5 bundesliga_t + \varepsilon_t$$

Source	SS	df	MS			
Model	688.158228	4	172.039557		Number of obs =	22
Residual	465.556058	17	27.3856504		F(4, 17) =	18.85
					Prob > F =	0.0000
Total	1153.71429	21	54.9387757		R-squared =	0.5965
					Adj R-squared =	0.5648
					Root MSE =	5.2331

points	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
value	.3790518	.0721279	5.26	0.000	.2376325	.5204711
possession	2.487062	.4710640	5.28	0.000	1.562692	3.411432
coach	-8.744706	2.404095	-3.64	0.000	-13.46227	-4.027140
bundesliga	-20.37705	7.634737	-2.67	0.008	-35.34628	-5.407819
_cons	-74.64235	20.84982	-3.58	0.000	-116.3419	-32.94271

Runden Sie im Folgenden alle Zahlenangaben auf die dritte Nachkommastelle.

3.1 Interpretieren Sie den geschätzten Koeffizienten b_3 inhaltlich und statistisch. [2 Punkte]

- Eine Erhöhung im durchschnittlichen Ballbesitz pro Spiel um 1 Prozentpunkt führt c.p. im Mittel zu einer Erhöhung der am Saisonende erreichten Punktzahl um 2,487 Punkte.
- Der Koeffizient ist statistisch signifikant auf dem 1% Niveau (da $p = 0{,}00 < 0{,}01$).

3.2 Begründen Sie, warum der geschätzte Koeffizient b_4 eventuell nicht kausal interpretiert werden kann. Erläutern Sie den Mechanismus anhand eines Beispiels. [2 Punkte]

- Antwort: Der geschätzte Koeffizient b_4 misst nicht den kausalen Effekt eines Trainerwechsels auf die am Saisonende erreichte Punktzahl.
- Begründung: Der Koeffizient ist bspw. wegen umgekehrter Kausalität verzerrt geschätzt (reverse causality bias). Der Effekt von schlechten Ergebnissen (wenig erreichten Punkten) des 1.FC Nürnberg auf den Trainerwechsel wird hier mitgeschätzt.
- (Andere Antworten möglich.)

3.3 Der folgende Stata-Output enthält die Ergebnisse eines White-Tests auf Heteroskedastie. Führen Sie den Test auf einem Signifikanzniveau von 10% durch. Geben Sie die Hypothesen und Hilfsregression an. Definieren Sie die abhängige Variable der Hilfsregression. Bestimmen Sie zudem die Teststatistik, Freiheitsgrade, kritischen Wert und Testentscheidung. [7 Punkte]

```
White's test for Ho: ???
against Ha: ???

chi2(?)      =    12,84
Prob > chi2  =    ??????
```

- Hypothesen: $H_0 : V(\varepsilon_t) = \sigma^2$ für alle t (Homoskedastie); $H_1 : V(\varepsilon_t) = \sigma_t^2 \neq \sigma^2$ (Heteroskedastie)
- Hilfsregression: $e_t^2 = \alpha_1 + \alpha_2 value_t + \alpha_3 value_t^2 + \alpha_4 possession_t + \alpha_5 possession_t^2 + \alpha_6 coach_t + \alpha_7 bundesliga_t + \alpha_8(value_t \cdot possession_t) + \alpha_9(value_t \cdot coach_t) + \alpha_{10}(value_t \cdot bundesliga_t) + +\alpha_{11}(possession_t \cdot coach_t) + \alpha_{12}(possession_t \cdot bundesliga_t) + \alpha_{13}(coach_t \cdot bundesliga_t) + \nu_t$
- wobei $e_t^2 = (y_t - x_t'b)^2$, die quadrierten Residuen aus der ursprünglichen Schätzung sind.
- Teststatistik: $\chi^2_{emp} = T \cdot R^2 = 12,84$, wobei T die Anzahl der Beobachtungen bezeichnet; R^2 ist das Bestimmtheitsmaß aus der Hilfsregression.
- Freiheitsgrade: J=12
- Kritischer Wert: $\chi^2_J = \chi^2_{12} = 18,55$
- Da $\chi^2_{empirisch} = 12,84 < 18,55 = \chi^2_{kritisch}$ wird die Nullhypothese auf dem 10%-Niveau nicht verworfen. Der Test weist auf Homoskedastie in dem vorliegenden Modell hin.

3.4 Warum könnte in dem vorliegenden Modell positive Autokorrelation im Fehlerterm vorliegen? Welche Folgen hat unkorrigierte Autokorrelation für die geschätzten Parameter und deren Standardfehler? [2,5 Punkte]

- Positive Autokorrelation im Fehlerterm könnte vorliegen, wenn Werte der abhängigen Variable über die Zeit positiv korrelieren, d.h. einer erfolgreichen Saison mit einer hohen Punktzahl folgt eine weitere erfolgreiche Saison, und diese Korrelation durch die Regressoren nicht aufgefangen wird.
- (Andere Antworten möglich.)
- Autokorrelation führt nicht zu Verzerrung der geschätzten Parameter, diese sind unverzerrt geschätzt (solange A1 und A2 gelten).
- Die Standardfehler sind allerdings falsch berechnet. (Die Schätzung ist ineffizient.)

3.5 Führen Sie einen Breusch-Godfrey-Test auf Autokorrelation 1. Ordnung auf einem Signifikanzniveau von 5% durch. Geben Sie die Hypothesen, die Hilfsregression, die Teststatistik, die Freiheitsgrade, den kritischen Wert und die Testentscheidung an. Definieren Sie dabei formal Autokorrelation 1. Ordnung. *Hinweis: Das Bestimmtheitsmaß aus der Hilfsregression lautet:* $R^2 = 0,2$. [6,5 Punkte]

- Autokorrelation 1.Ordnung: $\varepsilon_t = \rho\varepsilon_{t-1} + \nu_t$
- Hypothesen: $H_0 : \rho = 0$; $H_1 : \rho \neq 0$ (Autokorrelation 1.Ordnung)
- Hilfsregression: $e_t = \alpha_1 + \alpha_2 e_{t-1} + \alpha_3 value_t + \alpha_4 possession_t + \alpha_5 coach_t + \alpha_6 bundesliga_t + \nu_t$
- Teststatistik: $\chi^2_{emp} = (T-1) \cdot R^2 = (22-1) \cdot 0,2 = 4,2$, wobei T die Anzahl der beobachteten Perioden bezeichnet; R^2 ist das Bestimmtheitsmaß aus der Hilfsregression.
- Freiheitsgrade: J=1
- Kritischer Wert: $\chi^2_J = \chi^2_1 = 3,842$
- Da $\chi^2_{empirisch} = 4,2 > 3,842 = \chi^2_{kritisch}$, wird die Nullhypothese auf dem 5%-Niveau verworfen. Der Test weist auf Autokorrelation 1. Ordnung in dem vorliegenden Modell hin.

Aufgabe 4: **[13 Punkte]**

Sie interessieren sich für die Determinanten des Rauchverhaltens von Amerikanern. Hierfür steht Ihnen ein Datensatz mit 807 Befragten zur Verfügung. Er enthält folgende Variablen:

$smoker_i$	Dummy-Variable, =1, wenn Person i raucht, =0 sonst
$educ_i$	Bildung von Person i in Jahren
age_i	Alter von Person i in Jahren
inc_i	Jahreseinkommen von Person i in Dollar

Ihnen sind folgende deskriptive Statistiken zu den Daten gegeben:

```
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      smoker |       807    .3841388    .4866926          0          1
        educ |       807    12.47088    3.057161          6         18
         age |       807    41.23792    17.02729         17         88
      income |       807    19970.58    9498.139        500      32700
```

Sie führen mit Stata eine Logit-Schätzung des folgenden Modells durch:

$$P(smoker = 1|X) = F(\beta_1 + \beta_2 educ_i + \beta_3 age_i + \beta_4 age_i^2)$$

```
Logistic regression                               Number of obs   =        807
                                                  LR chi2(3)      =      47.57
                                                  Prob > chi2     =     0.0000
Log likelihood = -513.71831                       Pseudo R2       =     0.0443

------------------------------------------------------------------------------
      smoker |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |  -.1350465   .0268896    -5.02   0.000    -.1877492   -.0823438
         age |   .1060429   .0269834     3.93   0.000     .0531563    .1589295
       age^2 |  -.0013844   .0003073    -4.51   0.000    -.0019867   -.0007822
       _cons |  -.4479964   .5725251    -0.78   0.434    -1.570125    .6741321
------------------------------------------------------------------------------
```

Runden Sie im Folgenden alle Zahlenangaben auf die dritte Nachkommastelle.

4.1 Wie hoch ist die durchschnittliche vorhergesagte Wahrscheinlichkeit zu rauchen? [1 Punkt]

- 38,41% (Genau der Durchschnitt der abhängigen Variable)

4.2 Berechnen und interpretieren Sie den marginalen Effekt eines zusätzlichen Bildungsjahres am Mittelwert der Daten. [6 Punkte]

- Marginale Effekte im Logit-Modell allgemein (aus der Formelsammlung):

$$\frac{\partial P(y_i = 1|\bar{\mathbf{x}})}{\partial x_{ik}} = \frac{exp(\mathbf{x}'_{ik}\beta)}{(1 + exp(\mathbf{x}'_{ik}\beta))^2} \cdot \beta_k$$

- Bestimmung des marginalen Effekts im Logit-Modell für $educ_i$:

$$\mathbf{x}'_{ik}b\Big|_{\mathbf{x}=\bar{\mathbf{x}}} = -0,448 - 0,135 \cdot 12,471 + 0,106 \cdot 41,238 - 0,001 \cdot 1700,573 = 0,539$$

$$\frac{\partial P(smoker = 1|\bar{\mathbf{x}})}{\partial educ_i} = \frac{exp(0,539)}{(1 + exp(0,539))^2} \cdot (-0,135) = -0,031$$

- Die Wahrscheinlichkeit zu rauchen sinkt am Mittelwert der Daten c.p. im Mittel mit jedem zusätzlichen Bildungsjahr um 3,1 Prozentpunkte.

4.3 Eine zweite Logit-Schätzung mit dem gleichen Datensatz, bei der zusätzlich zu den bisherigen Variablen noch das Einkommen berücksichtigt wird, erzielt einen Log-Likelihoodwert von $-510,272$. Prüfen Sie, ob das Einkommen einen signifikanten Effekt auf die Wahrscheinlichkeit zu rauchen hat. Führen Sie dazu einen Likelihood-Ratio-Test am 1%-Signifikanzniveau durch. Geben Sie zuerst das restringierte Modell und das unrestringierte Modell an. Benennen Sie dann Hypothesen und kritischen Wert. Treffen Sie eine Testentscheidung. [6 Punkte]

- Restringiertes Modell: $P(smoker = 1|X) = F(\beta_1 + \beta_2 educ_i + \beta_3 age_i + \beta_4 age_i^2)$
- Unrestringiertes Modell: $P(smoker = 1|X) = F(\beta_1 + \beta_2 educ_i + \beta_3 age_i + \beta_4 age_i^2 + \beta_5 inc_i)$
- Nullhypothese: $H_0 : \beta_5 = 0$
- Alternativhypothese: $H_1 : \beta_5 \neq 0$
- Kritischer Wert: $\chi^2_1 = 6,635$
- Testentscheidung: $LR_{empirisch} = -2 \cdot (-513.718 - (-510.272)) = -2 \cdot (-3,446) = 6,892 > 6,635$ Die H_0 kann auf dem 1%-Signifikanzniveau verworfen werden.

Aufgabe 5 - 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 Modell $y_t = \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + \varepsilon_i$ führt Autokorrelation 1. Ordnung in ε
a	zu unobserved heterogeneity (unbeobachtete Heterogenität).
b	zu Endogenität von y_{t-1}. **X**
c	zum attenuation bias.
d	zu einer Verzerrung von β_2.

2.	Ein RESET-Test überprüft
a	mit Hilfe eines t-Tests, ob Fehlspezifikation des Regressionsmodells vorliegt.
b	, ob durch Änderung der Modellspezifikation das angepasste Bestimmtheitsmaß $\overline{R}^2$ signifikant steigt.
c	auf Basis eines F-Tests, ob ein Problem ausgelassener Variablen vorliegt. **X**
d	ob sich die marginalen Effekte im Modell für Teilgruppen unterscheiden.

3.	Das Modell $Lohn_i = \beta_1 + \beta_2 uni_i + \beta_3 west_i + \beta_4 uni_i \cdot west_i + \varepsilon_i$ wird mittels KQ geschätzt ($uni_i = 1$, falls Uni-Abschluss , $west_i = 1$ falls Westdeutscher). Welche Aussage trifft für die Schätzung des Modells zu?
a	Der marginale Effekt von uni_i für Ostdeutsche lautet: $b_2 + b_4$.
b	Der Lohnunterschied zwischen west- und ostdeutschen Uni-Absolventen lautet: $b_3 + b_4$. **X**
c	Der vorhergesagte Lohn für Westdeutsche mit Uni-Abschluss lautet: $b_1 + b_3 + b_4$.
d	Der Lohnunterschied zwischen westdeutschen Uni-Absolventen und westdeutschen Nicht-Absolventen lautet: b_2.

4.	Sie schätzen das Modell $y_i = \beta_1{}^* + \beta_2{}^* x_i + \varepsilon_i{}^*$ und Ihr Kommilitone das Modell $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$. Sei $cov(x,z) < 0$. Dann gilt:
a	$E[b_2{}^*] < \beta_2$, wenn $\beta_3 < 0$.
b	$E[b_2] < E[b_2{}^*]$, wenn $\beta_3 < 0$. **X**
c	$E[b_2{}^*] - \beta_2 > 0$, wenn $\beta_3 > 0$.
d	$E[b_2] \neq E[b_2{}^*]$, wenn $\beta_3 = 0$.

5.	Eine KQ-Schätzung des Modells $y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + \varepsilon_i$ ergibt b_1 = 5, b_2 = -0,5 und b_3=0,025. Bei welchem Wert von $\hat{y}$ ist der marginale Effekt von x gleich null?
a	5
b	2,5 **X**
c	10
d	0,25

6.	Für das Modell $y_i = \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i} + u_i$ testen Sie $H_0 : \beta_2 = -2$ vs. $H_1 : \beta_2 \neq -2$. Auf einem Signifikanzniveau von 5% führt eine t-Teststatistik von $-2,16$
a	bei $n = 16$ nicht zur Ablehnung von H_0. **X**
b	bei $n = 15$ zur Ablehnung von H_0.
c	bei $n = 17$ nicht zur Ablehnung von H_0.
d	bei $n = 14$ zur Ablehnung von H_0.

7.	Die Schätzung des Modells $zufriedenheit_i = \beta_1 + \beta_2 freizeit_i + \varepsilon_i$ mit KQ ergibt bei 150 Beobachtungen $b_2 = 2,5$, $se(b_2) = 0,5$. Welche Aussage trifft für die Schätzung von β_2 zu?
a	Die untere Grenze des 95 % Konfidenzintervalls für β_2 beträgt $3,48$.
b	Der kritische Wert für einen t-Test auf Signifikanz mit $\alpha = 0,01$ beträgt $2,326$.
c	Der p-Wert für den beidseitigen t-Test der H_0: $\beta_2 = 1,25$ ist größer als $0,01$. **X**
d	Die Nullhypothese $\beta_2 \leq 0$ eines t-Tests wird nicht verworfen, wenn der kritische Wert $1,282$ beträgt.

8.	Im linearen Modell $y_i = \beta_1 + \beta_2 x_{i1} + \beta_3 x_{i1}^2 + \beta_4 x_{i2} + \beta_5 x_{i2}^2 + \varepsilon_i$ mit $b_2 < 0$, $b_3 > 0$, $b_4 < 0$ und $b_5 < 0$ ist
a	der marginale Effekt von x_2 für große x_2-Werte negativ und betragsmäßig zunehmend. **X**
b	der marginale Effekt von x_1 für kleine x_1-Werte negativ und betragsmäßig zunehmend.
c	der marginale Effekt von x_2 für kleine x_2-Werte negativ und betragsmäßig abnehmend.
d	der marginale Effekt von x_1 für große x_1-Werte positiv und betragsmäßig abnehmend.

9.	Welche Aussage bezüglich des Durbin-Wu-Hausman Test ist richtig?
a	Die Nullhypothese besagt, dass die instrumentierte Variable endogen ist.
b	Bei Ablehnung der Nullhypothese liegt Endogenität vor. **X**
c	In der Hilfsregression wird die Instrumentenvariable auf die geschätzten Residuen aus der 1. Stufe der 2SLS-Schätzung regressiert.
d	Die Anzahl der Zählerfreiheitsgrade des kritischen Wertes entspricht der Anzahl der Instrumente.

10.	Welche Aussage bezüglich des Bestimmtheitsmaßes R^2 im multiplen Regressionsmodell bei N Beobachtungen ist richtig?
a	Im Modell ohne Konstante gilt $R^2 = 0$.
b	Wenn $\sum_{i=1}^{n} e_i = 0$ gilt, gilt $R^2 = 1$.
c	Wenn $R^2 = 0,49$ gilt, gilt für das angepasste Bestimmtheitsmaß $\overline{R}^2 = 1 - 0,51 \cdot \frac{N-1}{N-K}$. **X**
d	Im Modell mit Konstante und $K > 1$ gilt $0 \leq R^2 < 0,5$.

11.	Ein Konfidenzintervall
a	wird mit c.p. steigender Präzision schmaler. **X**
b	wird mit c.p. steigender Stichprobengröße schmaler.
c	wird mit c.p. sinkendem Signifikanzniveau α schmaler.
d	hat stets negative Intervallgrenzen, wenn $b_j < 0$.

12.	Sei $V(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{\Psi}$ die Varianz-Kovarianz-Matrix der Störterme. Liegt Heteroskedastie und Autokorrelation vor, ist die Matrix $\boldsymbol{\Psi}$
a	eine untere Dreiecksmatrix.
b	eine obere Dreiecksmatrix.
c	eine symmetrische Matrix. **X**
d	eine Diagonalmatrix.

13.	Der Breusch-Pagan Test
a	kann die falsche H_0 mit höherer Wahrscheinlichkeit verwerfen als der White-Test. **X**
b	ist allgemeiner als der White-Test.
c	prüft in der Hilfsregression, ob e^2 durch die ursprünglichen Regressoren und deren Quadrate erklärt werden kann.
d	hat eine höhere Anzahl an Freiheitsgraden J als der White-Test.

14.	Unkorrigierte Heteroskedastie im linearen Regressionsmodel führt zu
a	Verzerrung des KQ-Schätzers.
b	Inkonsistenz des KQ-Schätzers.
c	Effizienz des KQ-Schätzers.
d	fehlerhaft geschätzten Konfidenzintervallen. **X**

15.	Es sei $V(\varepsilon_i) = \sigma^2 \cdot Einkommen_i^4$. Zu konstanter Störtermvarianz führt die Tranformation
a	$\varepsilon_i / Einkommen_i^{16}$.
b	$\varepsilon_i / Einkommen_i^{8}$.
c	$\varepsilon_i \cdot Einkommen_i^{4}$.
d	$\varepsilon_i / Einkommen_i^{2}$. **X**

16.	Der Cochrane-Orcutt Schätzer
a	korrigiert nicht für negative Autokorrelation im Störterm.
b	korrigiert für Autokorrelation höherer als 1. Ordnung.
c	nutzt die Beobachtung für $t = 1$ nicht. **X**
d	liefert immer die gleichen Standardfehler wie der Prais-Winston Schätzer.

17.	Für die Schätzgleichung $\hat{y}_i = 1 - 2 \cdot x_{i2} + 3 \cdot x_{i3}$ und die Beobachtung $(y_i, x_{i2}, x_{i3}) = (-1, -1, -1)$ beträgt das Residuum
a	3.
b	1.
c	-1. **X**
d	3.

18.	Die obere und untere Grenze für den kritischen Wert eines Durbin-Watson Tests auf positive Autokorrelation für ein lineares Modell mit 100 beobachteten Perioden, einer Konstante und 2 unabhängigen Variablen beträgt auf einem Signifikanzniveau von 5%
a	$d_L = 1,63$ und $d_W = 1,72$. **X**
b	$d_L = 1,65$ und $d_W = 1,69$.
c	$d_L = 1,62$ und $d_W = 1,71$.
d	$d_L = 1,61$ und $d_W = 1,74$.

19.	Wenn keine Autokorrelation 1. Ordnung in den Residuen vorliegt, ist der Wert der Durbin-Watson Teststatistik nahe
a	-2.
b	0
c	2. **X**
d	4.

20.	Wenn unter den erklärenden Variablen eines linearen Regressionsmodels perfekte Multikollinearität vorliegt, gilt für die Matrix $X'X$, dass sie
a	nicht den vollen Rang hat. **X**
b	regulär ist.
c	invertierbar ist.
d	eine Determinante ungleich Null hat.

21.	Wird ein Modell einmal mit Probit und einmal mit Logit geschätzt
a	sind die marginalen Effekte identisch.
b	sind die geschätzten Koeffizienten identisch.
c	sind beide Schätzungen effizient.
d	ist mindestens eine der beiden Schätzungen nicht konsistent. **X**

22.	Wenn eine KQ-Schätzung nicht konsistent ist, dann
a	sind die Koeffizienten immer verzerrt.
b	enthalten die Koeffizienten keine validen Informationen über Korrelationen zwischen den Variablen.
c	ist der Erwartungswert des Residuums trotzdem 0. **X**
d	ist das angepasste R^2 kleiner als 0.

23.	Die Matrix $X'X$
a	ist immer invertierbar.
b	ist immer quadratisch. **X**
c	hat immer einen niedrigeren Rang als die Matrix X.
d	hat immer einen höheren Rang als die Matrix X.

24.	Ein Unterschied zwischen Wald-Test und Likelihood Ratio-Test ist, dass der Wald-Test
a	weniger Freiheitsgrade hat.
b	mehr Restriktionen auf einmal testen kann.
c	präziser ist.
d	nur eine Schätzung erfordert. **X**

25.	Der marginale Effekt in einem Probit-Modell mit nur einer unabhängigen Variable
a	ist am größten am Mittelwert der unabhängigen Variable.
b	ist am größten bei Beobachtungen mit $x = 0$.
c	kann immer als Semi-Elastizität interpretiert werden.
d	ist betragsmäßig immer kleiner als der Koeffizient. **X**

26.	Wenn das Pseudo-R^2 einer Logit-Schätzung 0,8 beträgt, die Log-Likelihood des unrestringierten Modells -100 und die des restringierten Modells -120, wie hoch ist dann die Anzahl der Beobachtungen?
a	5
b	10 **X**
c	40
d	100

27.	Bei unabhängigen Beobachtungen entspricht die Likelihoodfunktion
a	dem Produkt der individuellen Wahrscheinlichkeitsdichten der beobachteten abhängigen Variablen. **X**
b	der Summe der quadrierten individuellen Wahrscheinlichkeitsdichten der beobachteten abhängigen Variablen.
c	der Summe der individuellen Wahrscheinlichkeitsdichten der beobachteten abhängigen Variablen.
d	dem Logarithmus der individuellen Wahrscheinlichkeitsdichten der beobachteten abhängigen Variablen.

28.	Ein Messfehler in x führt bei einer KQ-Schätzung des Modells $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$
a	zum Problem des attenuation bias. **X**
b	zur Überschätzung von β_2.
c	zu endogenen Störtermen.
d	zu einer Unterschätzung der Störtermvarianz.

29.	Was ist eine Eigenschaft des KQ-Schätzers in einem linearen Modell mit binärer abhängiger Variable?
a	Der Fehlerterm ist homoskedastisch.
b	Bei großer Stichprobengröße ist er effizienter als Probit.
c	Bei gewöhnlich berechneten Standardfehlern sind t-Tests auf Signifikanz der Kovariate ungültig. **X**
d	Die Summe der errechneten Wahrscheinlichkeiten ist gleich $\frac{N}{2} \cdot \bar{y}$.

30.	Die Verletzung welcher Annahme schließt die kausale Interpretation einer Schätzung aus? (s. Formelsammlung)
a	A2
b	A3
c	A7 **X**
d	A11