Masterprüfung SS 2020 - 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: **[10 Punkte]**

Sie interessieren sich für die Determinanten der jährlichen Arbeitsstunden von Erwerbstätigen. Ihnen liegt hierzu eine Stichprobe aus dem Jahr 2018 mit 10.922 Befragten mit folgenden Informationen vor:

$Arbeitsstunden_i$	Arbeitsstunden von Person i im letzten Kalenderjahr
$log\_Einkommen_i$	Logarithmiertes monatliches Einkommen von Person i in €
$Kind_i$	Dummy-Variable, =1, wenn Person i mindestens ein Kind hat, =0 sonst
$Mann_i$	Dummy-Variable, =1, wenn Person i männlich, =0 sonst
$Alter_i$	Alter von Person i in Jahren, berechnet als Befragungsjahr$-$Geburtsjahr

Es wird folgendes Regressionsmodell aufgestellt und anschließend mit Stata geschätzt:

$$Arbeitsstunden_i = \beta_1 + \beta_2 log\_Einkommen_i + \beta_3 Kind_i + \beta_4 Mann_i + \beta_5 Alter_i + \beta_6 Mann\_Kind_i + \varepsilon_i$$

wobei $Mann\_Kind_i = Mann_i \cdot Kind_i$.

Arbeitsstunden	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
log_Einkommen	93.7943	20.43242	4.57	0.000	53.44684	124.3292
Kind	-129.863	23.42910	-5.43	0.000	-175.785	-85.2105
Mann	77.8549	???????	????	?????	32.18234	124.5214
Alter	0.589082	1.284000	0.46	0.647	-1.92755	1.373922
Mann_Kind	47.01218	9.234201	5.09	0.000	28.91315	65.76623
_cons	0.518024	.0729190	7.10	0.000	0.375103	13.99696

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

1.1 Interpretieren Sie den geschätzten Koeffizienten b_2 inhaltlich und statistisch. [2 Punkte]

- Eine 1% Erhöhung des Einkommens führt c.p. zu einem durchschnittlichen Anstieg der Arbeitsstunden um 0,937 Stunden.
- Der Koeffizient ist signifikant auf dem 1% Niveau. (da $p = 0,00 < 0,01$)

1.2 Berechnen Sie den Standardfehler von b_4. [3 Punkte]

- Zugehörigen t-Wert 1,960 für $\alpha = 0,05$ aus Tabelle ablesen
- Standardfehler aus Konfidenzintervallen:

$$b_4 + 1,960 * SE(b_4) = 124,521$$

$$SE(b_4) = \frac{124,521 - b_4}{1,960}$$

$$SE(b_4) = \frac{124,521 - 77,855}{1,960}$$

$$SE(b_4) = 23,300$$

- $SE(b_4) = 23,809$
- (Lösung auch über Untergrenze des KIs denkbar)

1.3 Leiten Sie den marginalen Effekt der Variable *Kind* allgemein her. Wie groß ist der marginale Effekt der Variable *Kind* für Männer und Frauen? [3 Punkte]

- Allgemein:
$$\frac{\Delta E(Arbeitsstunden_i)}{\Delta Kind_i} = \beta_3 + \beta_6 Mann_i$$
- Frauen mit Kindern arbeiten c.p. im Mittel 129,863 Stunden weniger als Frauen ohne Kindern.
- Männer mit Kindern arbeiten c.p. im Mittel 82,851 Stunden weniger als Männer ohne Kinder. (-129,863+47,012)

1.4 Ein Kommilitone schlägt vor, die Variable *Geburtsjahr* in das Modell aufzunehmen. Benennen Sie das daraus entstehende Problem und begründen Sie Ihre Antwort. [2 Punkte]

- Die Variable *Geburtsjahr* sollte nicht aufgenommen werden, da dadurch perfekte Multikollinearität vorliegen würde.
- Begründung: Da sich das Geburtsjahr aus $Geburtsjahr = 2018 - Alter - 6$ ergibt und das Modell sowohl eine Konstante, als auch das Alter bereits enthält, können die einzelnen Koeffizienten nicht mehr identifiziert werden.

Aufgabe 2: **[17,5 Punkte]**

Sie interessieren sich für die Determinanten von Löhnen in Deutschland. Hierfür sind Ihnen Daten über 40.477 Angestellte gegeben:

$lnwage_i$	Logarithmierter Bruttostundenlohn von Person i
$educyrs_i$	Bildungsjahre von Person i
age_i	Alter von Person i
$migrant_i$	Dummy-Variable, =1 wenn Person i einen Migrationshintergrund hat, =0 kein Migrationshintergrund
$educpar_i$	Höchste Bildungsjahre eines Elternteils von Person i

Es wird folgendes Regressionsmodell aufgestellt und anschließend mit Stata geschätzt:

$$lnwage_i = \beta_1 + \beta_2 educyrs_i + \beta_3 age_i + \beta_4 age_i^2 + \beta_5 migrant_i + \varepsilon_i$$

```
      Source |       SS           df       MS      Number of obs   =    40,477
-------------+----------------------------------   F(4, 40472)     =   4543.56
       Model |  5411.72332         4  1352.93083   Prob > F        =    0.0000
    Residual |  12051.3014    40,472  .297768862   R-squared       =    ??????
-------------+----------------------------------   Adj R-squared   =    ??????
       Total |  17463.0247    40,476  .431441464   Root MSE        =    .54568

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     educyrs |   .1025072   .0010551    97.16   0.000     ????????????????????
         age |   .0200607   .0002485    80.72   0.000     .0195736    .0205478
        age2 |  -9.72e-06   1.21e-07   -80.46   0.000    -9.96e-06   -9.49e-06
     migrant |   .0507535   .0082948     6.12   0.000     .0344955    .0670115
       _cons |   .3909978   .0160494    24.36   0.000     .3595405    .4224551
------------------------------------------------------------------------------
```

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

2.1 Interpretieren Sie den geschätzten Koeffizienten b_5 inhaltlich und statistisch. [2 Punkte]

- Migranten verdienen c.p. im Mittel etwa $b_5 \cdot 100\%$=5,075% mehr als Personen ohne Migrationshintergrund.
- Der Koeffizient ist statistisch signifikant auf dem 1% Niveau.

2.2 Berechnen Sie das 99%-Konfidenzintervall des Koeffizienten von *educyrs*. Wie lässt sich das Konfidenzintervall interpretieren? [2,5 Punkte]

- t-Wert in Tabelle ablesen: 2,576 (df=∞, $\frac{\alpha}{2} = 0,995$).
- Obere Grenze: $0,103 + 2,576 \cdot 0,001 = 0,106$.
- Untere Grenze: $0,103 - 2,576 \cdot 0,001 = 0,100$.
- (Das 99%-Konfidenzintervall des Koeffizienten von *educyrs* lautet: [0,100;0,106])
- Interpretation: Mit wiederholten Stichproben liegt das wahre β_{age} in 99% der Fälle im auf diese Weise berechneten Konfidenzintervall.

2.3 Berechnen Sie das R^2 der Lohnregression und interpretieren Sie dieses. [2 Punkt]

- $R^2 = 1 - \frac{\sum_{i=1}^N e_i^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2}$ (aus Formelsammlung) $= 1 - \frac{RSS}{TSS} = \frac{MSS}{TSS} = \frac{5411,723}{17463,025} = 0,310$.
- Das vorliegende Modell erklärt 31,0% der Variation des logarithmierten Lohns.

2.4 Sie wollen den kausalen Effekt der Bildung auf den Lohn schätzen und wollen hierfür die Variable *eduycrs* durch die Bildung der Eltern instrumentieren. Diskutieren Sie, unter welchen Bedingungen es sich allgemein, sowie im vorliegenden Fall bei der Variable *educpar* um eine geeignete Instrumentvariable für Bildung handelt. *Hinweis:* Sie müssen nicht entscheiden, ob es sich um ein geeignetes Instrument handelt, oder nicht. [3 Punkte]

- Relevanz: Das Instrument muss mit der endogenen Variable korreliert sein, die Bildung der Eltern korreliert vermutlich hoch mit der Bildung der Kinder .
- Exogenität: Das Instrument darf nicht mit unbeobachteten Determinanten der abhängigen Variable korreliert sein. Im vorliegenden Fall heißt das, dass Bildung der Eltern nur durch ihren Effekt auf die Bildung der Kinder mit dem Lohn korreliert.
- (Alternativerklärungen denkbar)

2.5 Sie entscheiden sich dafür, die Bildung der Eltern als Instrument in einer Two Stage Least Squares (2SLS)-Schätzung zu verwenden. Erläutern Sie kurz verbal die Vorgehensweise des 2SLS-Schätzers. *Hinweis:* Das Aufstellen von Modellgleichungen ist nicht nötig. [2 Punkte]

- Der 2SLS-Schätzer ist ein zweistufiger KQ-Schätzer.
- 1. Stufe: Die endogene Variable *educ* wird auf das Instrument *educpar* und alle exogenen Regressoren regressiert.
- 2.Stufe: Die Lohnregression wird geschätzt, wobei für *educ* die vorhergesagten Werte aus Stufe 1 eingesetzt werden.

2.6 Sie führen einen F-Test für die Signifikanz der Instrumentvariable auf der ersten Stufe durch und erhalten einen F-Wert von 20,99. Handelt es sich in dem vorliegenden Fall um ein schwaches Instrument? Erläutern Sie in diesem Kontext den Begriff schwacher Instrumente. [3 Punkte]

- Schwache Instrumente liegen vor, wenn der Zusammenhang zwischen Instrumentenvariable und endogener Variable nur relativ klein ist.
- Man spricht bei einem F-Wert kleiner 10 von einem schwachen Instrument.

=> Bei dem vorliegenden Instrument handelt es sich nicht um ein schwaches Instrument, da der F-Wert des Instruments größer als 10 ist.

2.7 Erläutern Sie verbal die Vorgehensweise und die Schlusslogik des Durbin-Wu-Hausman-Tests. [3 Punkte]

- Schritt 1: Man bestimmt die Residuen auf der ersten Stufe der 2SLS-Schätzung.
- Schritt 2: Die Residuen der ersten Stufe werden als unabhängige Variablen in die Lohnregression aufgenommen.
- Schritt 3: Ist der Koeffizient für die Residuen statistisch signifikant, so kann davon ausgegangen werden, dass die abhängige Variable endogen ist (bei einem gültigen Instrument).

Aufgabe 3: **[20,5 Punkte]**

Für die Aktie der Lange Gasse AG wird ein CAPM-Modell geschätzt.

$$LG\_Rendite_t = \beta_1 + \beta_2 Dax\_Rendite_t + \varepsilon_t$$

Der Datensatz enthält 173 monatliche Beobachtungen mit folgenden Variablen:

LG_Rendite	Monatliche Rendite der Lange Gasse AG (in Prozent; [1–100])
Dax_Rendite	Durchschnittliche monatliche DAX Rendite (in Prozent; [1–100])

In der Tabelle sind Ergebnisse der KQ-Schätzung dargestellt:

```
Source   |       SS           df          MS   Number of obs    =       173
---------+---------------------------------   F(1, 171)        =    561.77
Model    |  3.49613217         1  3.49613217   Prob > F         =    0.0000
Residual |  1.06421465       171  .006223477   R-squared        =    0.7666
---------+---------------------------------   Adj R-squared    =    0.7653
Total    |  4.56034682       172  .026513644   Root MSE         =    .07889

---------------------------------------------------------------------------
lgag     |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------+-----------------------------------------------------------------
dax      |    .0807463   .0034068    23.70   0.000     .0740215    .0874711
_cons    |    1.537372   .0104992   146.43   0.000
---------------------------------------------------------------------------
```

$\sum_{t=2}^{T} e_t e_{t-1} = 1,035$ und $\sum_{t=2}^{T} e_{t-1}^2 = 1,085$

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

3.1 Interpretieren Sie den geschätzten Koeffizienten b_2 inhaltlich und statistisch. [2 Punkte]

- Steigt die durchschnittliche monatliche DAX-Rendite um einen Prozentpunkt, so steigt die monatliche LG AG Rendite c.p. im Mittel um etwa $b_2 = 0.08$ Prozentpunkte.
- Der Koeffizient ist statistisch signifikant auf dem 1% Niveau.

3.2 Testen Sie auf dem Signifikanzniveau von $\alpha = 5\%$ die einseitige Hypothese, dass die Konstante größer als $1,5$ ist. [3,5 Punkte]

- Hypothesen: $H_0 : \beta_1 \leq 1,5$; $H_1 : \beta_3 > 1,5$
- Teststatistik: $t_{empirisch} = \frac{1,53-1,5}{0,01} = 3$
- Kritischer Wert: $t_{kritisch} = t_{\alpha;N-K-1} = t_{0,05;171} \approx t_{0,05;\infty} = 1,645$
- Testentscheidung: Auf dem Signifikanzniveau von 5% kann die Nullhypothese verworfen werden, da $t_{empirisch} = 3 > 1,645 = t_{kritisch}$. Der Effekt ist somit größer als 1,5.

3.3 Berechnen Sie den Autokorrelationskoeffizienten $\hat{\rho}$ und die (approximative) Durbin-Watson Statistik dw. [3 Punkte]

- $\hat{\rho} = \frac{\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=2}^{T} e_{t-1}^2} = \frac{1,035}{1,085} = 0,954$.
- $dw \approx 2(1-\hat{\rho}) = 2 - 2 \cdot 0,954 = 0,092$.

3.4 Nennen Sie zwei Bedingungen, die für die Gültigkeit des asymptotischen Durbin-Watson-Tests zutreffen müssen. [2 Punkte]

- Die Regressoren sind nichtstochastisch, d.h. A2 gilt und es sind keine verzögerten endogenen Variablen im Modell.
- X enthält die Regressionskonstante.

3.5 Sie vermuten, dass die Störgröße positiv autokorreliert ist. Unterstellen Sie in diesem Aufgabenteil für die Durbin-Watson-Statistik einen Wert von 0.08 und überprüfen Sie diese Hypothese auf dem 5%-Signifikanzniveau. Geben Sie hierzu die Nullhypothese, die Alternativhypothese, die Freiheitsgrade und die jeweiligen kritischen Werte an. [4 Punkte]

- Nullhypothese: $H_0 : \rho \leq 0$
- Alternativhypothese: $H_1 : \rho > 0$
- Freiheitsgrade: T=173 und K=2. T=173 nicht tabuliert, nächstkleinster Wert: 150
- Kritische Werte: bei T=150 und K=2: $d_L = 1,72; d_U = 1,75$
- Die Durbin-Watson Statistik von 0.08 liegt unter der unteren Grenze von $d_L = 1,72$; die Nullhypothese, dass keine positive Autokorrelation erster Ordnung vorliegt, kann auf dem 5%-Signifikanzniveau somit verworfen werden.

3.6 Führen Sie einen Breusch-Pagan-Test auf Heteroskedastie auf dem Signifikanzniveau von 5% 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 0,134. [6 Punkte]

- Hypothesen: $H_0 : V(\varepsilon_t) = \sigma^2$ für alle t (Homoskedastie); $H_1 : Var(\varepsilon_t) = \sigma_t^2 \neq \sigma^2$ für mindestens ein t (Heteroskedastie).
- Hilfsregression: $e_t^2 = \alpha_1 + \alpha_2 Dax\_Rendite_t + \nu_t$
- Abhängige Variable: $e_t^2 = (y_t - x_t'b)^2$, sind die quadrierten Residuen aus der ursprünglichen Schätzung.
- Teststatistik: $\chi^2_{emp} = T \cdot R^2 = 173 \cdot 0,134 = 22,914$ (wobei T die Anzahl der Beobachtungen bezeichnet; R^2 ist das Bestimmtheitsmaß aus der Hilfsregression).
- Freiheitsgrade: J=1 (Anzahl der Steigungsparameter - ohne Konstante).
- Kritischer Wert: $\chi^2_{J;\alpha} = \chi^2_{1;0,05} = 3,842$.
- Da $\chi^2_{empirisch} = 22,914 > 3,842 = \chi^2_{kritisch}$, wird die Nullhypothese auf dem 5%-Niveau verworfen. Der Test weist auf Heteroskedastie in dem vorliegenden Modell hin.

Aufgabe 4: **[12 Punkte]**

Der Datensatz beinhaltet Informationen über 2.725 US-amerikanische Männer, die vor 1986 mindestens einmal von der Polizei festgenommen wurden. Folgende Variablen sind gegeben:

$arr86$	=1, wenn Person 1986 festgenommen wurde, =0 sonst
$pcnv$	Anteil an vorherigen Festnahmen, die zu Verurteilungen bzw. Inhaftierungen führten (Wahrscheinlichkeit der Verurteilung)
$tottime$	Gesamtanzahl Monate, die vor 1986 im Gefängnis verbracht wurden
$qemp86$	Anzahl Quartale (0 bis 4), in denen die Person 1986 legal beschäftigt gewesen war
$black$	=1, wenn die Person Afroamerikaner ist, =0 sonst

Ihnen sind folgende deskriptive Statistiken zu den Daten gegeben:

Variable	Obs	Mean	Std. Dev.	Min	Max
arr86	2,725	.2770642	.4476306	0	1
pcnv	2,725	.3577872	.395192	0	1
tottime	2,725	.8387523	4.607019	0	63.4
qemp86	2,725	2.309028	1.610428	0	4
black	2,725	.1611009	.3676915	0	1

Folgendes lineares Wahrscheinlichkeitsmodell wird mit der KQ-Methode geschätzt, um die Wahrscheinlichkeit der Festnahme im Jahr 1986 zu beschreiben:

$$arr86_i = \beta_0 + \beta_1 pcnv_i + \beta_2 tottime_i + \beta_3 qemp86_i + \beta_4 black_i + \varepsilon_i$$

```
Source   |          SS        df          MS   Number of obs   =     2,725
---------+----------------------------------   F(4, 2720)      =     37.48
Model    |  28.5123599         4  7.12808997   Prob > F        =    0.0000
Residual |  517.304154     2,720  .190185351   R-squared       =    0.0522
---------+----------------------------------   Adj R-squared   =    0.0508
Total    |  545.816514     2,724   .20037317   Root MSE        =     .4361

---------------------------------------------------------------------------
arr86    |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------+-----------------------------------------------------------------
pcnv     |  -.1582026   .0212061     -7.46   0.000    -.1997843    -.116621
tottime  |  -.0014002   .0018371     -0.76   0.446    -.0050025     .002202
qemp86   |  -.0316549   .0052856     -5.99   0.000    -.0420191   -.0212908
black    |   .1434694   .0231842      6.19   0.000     .0980089    .1889299
_cons    |   .3848206   .0177212     21.72   0.000     .3500723     .419569
---------------------------------------------------------------------------
```

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

4.1 Erläutern Sie, für welche Personen die Konstante die vorhergesagte Wahrscheinlichkeit, 1986 festgenommen zu werden, angibt. [2 Punkte]

> Die durchschnittliche erwartete Wahrscheinlichkeit der Festnahme im Jahr 1986 für eine Person, die nie verurteilt wurde, ($pcnv = 0$), zuvor nie inhaftiert wurde ($tottime = 0$), nicht Afroamerikaner ist ($black = 0$) und während des ganzen Jahres 1986 nicht beschäftigt war ($qemp86 = 0$).

4.2 Nennen Sie zwei Nachteile des linearen Wahrscheinlichkeitsmodells. [2 Punkte]

> - Die vorhergesagten Werte für die abhängige Variable können Werte außerhalb [0,1] annehmen.
> - Die Varianz der Störterme ist nicht konstant, es liegt Heteroskedastie vor. Alternativ: Die Schätzung ist nicht effizient.
> - Die Störterme sind nicht normalverteilt, daher sind t- und F-Tests nicht exakt gültig.

4.3 Interpretieren Sie den geschätzten Koeffizienten b_4 inhaltlich und statistisch. [2 Punkte]

> - Afroamerikaner hatten in 1986 c.p. im Mittel 14,3 %-Punkte höhere Wahrscheinlichkeit festgenommen zu werden.
> - Dieser Effekt ist auf dem 1%-Niveau signifikant von 0 verschieden.

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

```
Logistic regression                        Number of obs     =      2,725
LR chi2(4)        =     144.67
Prob > chi2       =     0.0000
Log likelihood    = -1535.8484             Pseudo R2         =     0.0450

--------------------------------------------------------------------------
arr86   |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+------------------------------------------------------------
pcnv    |  -.8958444   .1210598    -7.40   0.000    -1.133117   -.6585716
tottime |  -.0062353   .0094339    -0.66   0.509    -.0247254    .0122549
qemp86  |  -.1654769   .0275192    -6.01   0.000    -.2194137   -.1115402
black   |   .6615984    .111675     5.92   0.000     .4427195    .8804773
_cons   |  -.4183695   .0885738    -4.72   0.000     -.591971   -.2447679
--------------------------------------------------------------------------
```

4.4 Berechnen Sie die Stärke des Effektes der Variable $black_i$ auf die Festnahmewahrscheinlichkeit am Mittel der Daten. Zeigen Sie Ihren Rechenweg. *Hinweis:* $b_0 + b_1\overline{pcnv} + b_2\overline{tottime} + b_3\overline{qemp86} = -1,125$ [6 Punkte]

- Bei dichotomen (0/1) erklärenden Variablen wird statt des marginalen Effekts der Unterschied in den vorhergesagten Wahrscheinlichkeiten bestimmt.
- Im Logit Modell gilt bei $P = \Lambda(\cdot) = \mathtt{exp}(x'\beta)/(1+\mathtt{exp}(x'\beta))$

$$P(y_i|black_i = 1, x_i = \bar{x}) - \mathtt{P}(y_i|black_i = 0, x_i = \bar{x}) = \Lambda(b_4 \cdot 1 + b_0 + b_1\overline{pcnv} + b_2\overline{tottime} + b_3\overline{qemp86})$$

$$-\Lambda(b_4 \cdot 0 + b_0 + b_1\overline{pcnv} + b_2\overline{tottime} + b_3\overline{qemp86})$$

$$b_4 \cdot 1 + b_0 + b_1\overline{pcnv} + b_2\overline{tottime} + b_3\overline{qemp86} = 0,662 \cdot 1 - 1,125 = -0,463$$

$$b_4 \cdot 0 + b_0 + b_1\overline{pcnv} + b_2\overline{tottime} + b_3\overline{qemp86} = 0,662 \cdot 0 + -1,125 = -1,125$$

$$\Lambda(-0,463) - \Lambda(-1,125) = \frac{\mathtt{exp}(-0,463)}{1+\mathtt{exp}(-0,463)} - \frac{\mathtt{exp}(-1,125)}{1+\mathtt{exp}(-1,125)} = 0,386 - 0,245 = 0,141$$

- Die Wahrscheinlichkeit verurteilt zu werden ist c.p. im Mittel für afroamerikanische Männer um 14,1 Prozentpunkte höher als für nicht-afroamerikanische Männer.

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.	Für den Durbin-Watson Test gilt nicht:
a	er ist nur bei einer Schätzung mit Konstante gültig.
b	er ist sowohl asymptotisch als auch bei kleinen Stichproben gültig, wenn die notwendigen Annahmen zutreffen.
c	er eignet sich als Test auf Konsistenz. **X**
d	er eignet sich nur für den Test auf Autokorrelation erster Ordnung.

2.	Die obere und untere Grenze für den kritischen Wert eines Durbin-Watson Tests auf Autokorrelation für ein lineares Modell mit 65 beobachteten Perioden, einer Konstante und 6 unabhängigen Variablen beträgt auf einem Signifikanzniveau von 5%
a	$d_L = 1,44$ und $d_W = 1,77$. **X**
b	$d_L = 1,56$ und $d_W = 1,72$.
c	$d_L = 1,54$ und $d_W = 1,71$.
d	$d_L = 1,51$ und $d_W = 1,74$.

3.	Autokorrelation im Störterm kann behoben werden durch
a	eine Vergrößerung der Stichprobe.
b	die Aufnahme von irrelevanten abhängigen Variablen.
c	eine Umskalierung der abhängigen Variable.
d	die Anwendung eines Prais-Winsten Schätzers. **X**

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

5.	Welche Aussage ist richtig?
a	Für binäre Regressoren können generell marginale Effekte nicht berechnet werden.
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. **X**
d	Der marginale Effekt am Mittel der Daten aus Logit-Regressionen variiert zwischen Individuen.

6.	Sie schätzen das Modell $Glück_i = \beta_1 + \beta_2 Lohn_i + \varepsilon_i$. Sie vermuten, dass ein Problem ausgelassener Variablen vorliegt, da die Variable *Bildung* nicht berücksichtigt wurde. Welche Aussage ist korrekt?
a	Das Auslassen von *Bildung* in der Schätzung führt zu Homoskedastie.
b	Wenn $Cov(Bildung, Lohn) > 0$ und $Cov(Bildung, Glück) = 0$, dann ist β_2 aus diesem Grund nicht unterschätzt. **X**
c	Wenn $Cov(Bildung, Lohn) > 0$ und $Cov(Bildung, Glück) < 0$, dann ist β_2 unverzerrt geschätzt.
d	Wenn $Cov(Bildung, Lohn) < 0$ und $Cov(Bildung, Glück) > 0$, dann ist β_2 aus diesem Grund überschätzt.

7.	Sie schätzen das Modell $ln\_wage_i = \beta_1 + \beta_2 male_i + \beta_3 firmsize_i + \varepsilon_i$ mittels einer KQ-Schätzung, wobei *ln_wage* der Logarithmus des Stundenlohns (in Euro) ist, *male* eine Dummyvariable für Männer ist und *firmsize* die Mitarbeiterzahl des Betriebs ist. Der geschätzte Koeffizient zu β_3 ist $-0,02$. Welche Interpretation ist richtig?
a	Steigt die Firmengröße um ein Prozent, so fällt der Stundenlohn c.p. im Mittel um etwa 20%.
b	Steigt die Firmengröße um eine Person, so fällt der Stundenlohn c.p. im Mittel um etwa 0,02%.
c	Steigt die Firmengröße um ein Prozent, so fällt der Stundenlohn c.p. im Mittel um etwa 2%.
d	Steigt die Firmengröße um eine Person, so fällt der Stundenlohn c.p. im Mittel um etwa 2%. **X**

8.	Folgende Regressorentabelle ist gegeben: $X = \begin{bmatrix} 0,2 & 0,4 & 0,8 \\ 0,3 & 0,6 & 0 \\ 0,4 & 0,8 & 1 \end{bmatrix}$. Welche Aussage trifft zu?
a	KQ- Schätzer hat keine Lösung. **X**
b	$X'X$ ist invertierbar.
c	$X'X$ ist symmetrisch.
d	$X'X$ ist nicht singulär.

9.	Für das Model $y = \beta_0 + \beta_1 x + \beta_1 x^2 + \beta_1 x^3 + u$ gilt:
a	Mittels eines t-Tests kann überprüft werden, ob die Polynome der erklärenden Variablen gemeinsam einen signifikanten Erklärungsgehalt liefern.
b	Das Modell kann nicht geschätzt werden.
c	Der marginale Effekt der Variable x ergibt sich als Ableitung des auf x bedingten Erwartungswerts von y. **X**
d	Lineare Zusammenhänge zwischen x und y können damit nicht abgebildet werden.

10.	Für das Modell $income_i = \beta_1 + \beta_2 educ_i + \beta_3 exper_i + \beta_4 exper_i^2 + u_i$ ergibt eine KQ-Schätzung $b_1 = 100$, $b_2 = 50$, $b_3 = 20$ und $b_4 = -0,5$. Bei welchem Wert von *exper* ist der marginale Effekt von Berufserfahrung null?
a	20. **X**
b	18.
c	20.
d	40.

11.	Man spricht von einem überidentifizierten Modell, wenn
a	exakt so viele Instrumente wie endogene Variablen vorliegen.
b	ein Schätzmodell den Durbin-Wu-Hausman Test besteht.
c	weniger Instrumente als endogene Variablen vorliegen.
d	keine der Antworten ist korrekt. **X**

12.	In der ersten Stufe einer 2SLS-Schätzung
a	liegt zwangsläufig Endogenität vor.
b	wird die endogene Variable auf das Instrument, aber nicht die exogenen Variablen regressiert.
c	wird die endogene Variable auf das Instrument und die exogenen Variablen regressiert. **X**
d	wird die abhängige Variable der zweiten Stufe auf das Instrument regressiert.

13.	Im Logit-Modell
a	liegt zwangsläufig Endogenität vor.
b	entsprechen die geschätzten Koeffizienten den marginalen Effekten.
c	kann die Schätzgüte mit einem Likelihood-Ratio-Test bewertet werden. **X**
d	sind bei gleicher Modellspezifikation die marginalen Effekte exakt identisch zum Probit-Modell.

14.	Was ergibt sich aus dem Produkt der Matrizen $A = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}$ und $B = \begin{pmatrix} 4 & 5 \\ 7 & 8 \end{pmatrix}$?
a	$AB = \begin{pmatrix} 4 & 5 \\ 28 & 32 \end{pmatrix}$.**X**
b	$AB = \begin{pmatrix} 4 & 5 \\ 11 & 13 \end{pmatrix}$.
c	$AB = \begin{pmatrix} 1 & 9 \\ 4 & 15 \end{pmatrix}$.
d	Die Matrix AB ist nicht definiert.

15.	Die Varianz-Kovarianzmatrix der Störterme eines Regressionsmodells sei $\sigma^2\Omega$, wobei $\Omega = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{bmatrix}$. Hier gilt:
a	Homoskedastie und positive Autokorrelation.
b	Heteroskedastie und positive Autokorrelation.
c	Homoskedastie und keine Autokorrelation.
d	Heteroskedastie und keine Autokorrelation. **X**

16.	Ein nicht-systematischer Messfehler in x führt bei einer KQ-Schätzung des Modells $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$
a	zum Problem des ability bias.
b	zu einer Überschätzung von β_2.
c	zu endogenen Störtermen.
d	zu einer Verzerrung von β_1. **X**

17.	Bei GLS (generalized least squares) Schätzern
a	muss die Varianz-Kovarianz Matrix geschätzt werden.
b	sind nach der GLS-Transformation die Gauß-Markov Annahmen verletzt.
c	ist die Varianz-Kovarianz Matrix des Störterms bekannt. **X**
d	sind t- und F-Tests nicht gültig.

18.	Sei $V(\boldsymbol{\epsilon}) = \sigma^2\boldsymbol{\Psi}$ die Varianz-Kovarianz-Matrix der Störterme. Liegt Heteroskedastie, aber keine Autokorrelation vor, ist die Matrix $\boldsymbol{\Psi}$
a	keine symmetrische Matrix.
b	keine quadratische Matrix.
c	eine Diagonalmatrix. **X**
d	eine Einheitsmatrix.

19.	Was ist eine Eigenschaft des KQ-Schätzers bei binärer abhängiger Variable?
a	Der Fehlerterm ist heteroskedastisch. **X**
b	Es werden immer erwartete Wahrscheinlichkeiten außerhalb des Intervalls $[0, 1]$ errechnet.
c	Er ist verzerrt.
d	Das R^2 hat keine sinnvolle Interpretation.

20.	Der Prais-Winsten-Schätzer
a	wird zur Korrektur von Heteroskedastie angewendet.
b	korrigiert für Autokorrelation höherer als 1. Ordnung.
c	nimmt die Beobachtung $t = 1$ aus.
d	ist eine Erweiterung des Cochrane-Orcutt-Verfahrens. **X**

21.	Unkorrigierte Autokorrelation im linearen Regressionsmodel führt zu
a	kleinstmöglicher Varianz des KQ-Schätzers.
b	Verzerrung des KQ-Schätzers.
c	falschen Werten der Konfidenzintervalle der Steigungsparameter. **X**
d	richtigen Standardfehlern des KQ-Schätzers.

22.	Für die Schätzgleichung $\hat{y}_i = 3,5 - 2 \cdot x_{i1} + 1 \cdot x_{i2}$ und die Beobachtung $(y_i, x_{i1}, x_{i2}) = (-1, 1, 1)$ beträgt das Residuum
a	-3,5. **X**
b	-2,5.
c	-0,5.
d	0,5.

23.	Gegeben ist folgendes Modell: $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i$. Die aufgestellten Hypothesen lauten $H_0 : \beta_3 \leq 0$ vs. $H_1 : \beta_3 > 0$. Auf einem Signifikanzniveau von 5% führt eine t-Teststatistik von $1,710$
a	bei $n = 24$ zur Ablehnung von H_0.
b	bei $n = 25$ zur Ablehnung von H_0. **X**
c	bei $n = 26$ nicht zur Ablehnung von H_0.
d	bei $n = 27$ nicht zur Ablehnung von H_0.

24.	Um einen RESET-Test durchzuführen, regressiert man in einer Hilfsregression
a	Residuen auf alle unabhängigen Variablen und verzögerte Residuen.
b	die abhängige Variable auf die quadrierten Residuen.
c	die abhängige Variable auf Polynome der vorhergesagten abhängigen Variable und alle unabhängigen Variablen. **X**
d	quadrierte Residuen auf alle unabhängigen Variablen, deren Quadrate und deren Interaktionen.

25.	Gegeben ist folgendes Modell: $y_i = \beta_1 + \beta_2 x_{i2} + u_i$. b_2 beträgt -4 und $SE(b_2) = 0,5$. Die aufgestellten Hypothesen lauten $H_0 : \beta_2 = 0$ vs. $H_1 : \beta_2 \neq 0$. Die dazugehörige t-Teststatistik beträgt
a	$t^{empirisch} = 15$.
b	$t^{empirisch} = 4$.
c	$t^{empirisch} = -8$. **X**
d	$t^{empirisch} = -4$.

26.	Welche der Aussagen zu R^2 ist richtig?
a	Ein $R^2 \leq 0.05$ weist auf ein Endogenitätsproblem hin.
b	R^2 sinkt bei der Aufnahme zusätzlicher Regressoren, wenn diese keinen Erklärungsgehalt haben.
c	Das angepasste R^2 kann nur Werte zwischen -1 und 0 annehmen.
d	In einem Modell nur mit der Konstanten liegt R^2 bei 0. **X**

27.	Sie schätzen das Modell $ln\_wage_i = \beta_1 + \beta_2 male_i + \beta_3 ln\_workinghours_i + \varepsilon_i$ mittels einer KQ-Schätzung, wobei *ln_wage* den logarithmierten Stundenlohn in $ beschreibt, *male* eine Dummyvariable für Männer ist und *ln_workinghours* der Logarithmus der Arbeitszeit in Stunden ist. Der geschätzte Koeffizient b_3 ist $0,93$. Welche Interpretation von b_3 ist richtig?
a	Steigt die Arbeitszeit um 1 Stunde, so steigt im Durchschnitt c.p. der Lohn um 0,93%.
b	Steigt die Arbeitszeit um 1%, so steigt im Durchschnitt c.p. der Lohn um 0,93$.
c	Steigt die Arbeitszeit um 1 Stunde, so steigt im Durchschnitt c.p. der Lohn um 0,93$.
d	Steigt die Arbeitszeit um 1%, so steigt im Durchschnitt c.p. der Lohn um 0,93%. **X**

28.	Sie schätzen das Modell $ln(Stundenlohn_i) = \beta_1 + \beta_2 frau_i + \beta_3 educ_i + \beta_4 (frau_i \cdot educ_i) + \varepsilon_i$ mittels einer KQ-Schätzung ($educ_i$ misst Bildung in Jahren, $frau_i = 1$ ob die Person eine Frau ist). Welche Aussage über die Schätzung trifft zu?
a	b_4 gibt die geschätzte Bildungsrendite für Frauen an.
b	b_3 gibt die geschätzte Bildungsrendite für Männer an. **X**
c	b_1 gibt den durchschnittlichen Stundenlohnen für Frauen mit 0 Bildungsjahren an.
d	b_2 gibt den mittleren Lohnunterschied für Frauen und Männer in der Stichprobe an.

29.	In einer 2SLS-Schätzung werden die Werte der endogenen Variable in der zweiten Stufe ersetzt durch
a	die Residuen der Schätzung auf der ersten Stufe.
b	die quadrierten Residuen der ersten Stufe.
c	die Werte der Instrumentvariable.
d	die vorhergesagten Werte der abhängigen Variable aus der ersten Stufe. **X**

30.	Im Fall einer binären abhängigen Variable
a	wird im Logit-Modell eine logistische, im Probit-Modell eine χ^2-Verteilung unterstellt.
b	sind die Ergebnisse einer Probit-Schätzung mit denen des linearen Wahrscheinlichkeitsmodells identisch.
c	müssen alle Variablen transformiert werden, damit eine Schätzung mittels KQ-Methode möglich ist.
d	entsprechen die Koeffizienten stetiger erklärender Variablen im Fall eines linearen Wahrscheinlichkeitsmodells den marginalen Effekten. **X**