Masterprüfung WS 2016/2017 - 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: **[20 Punkte]**

Sie interessieren sich für die Determinanten des täglichen Wasserkonsums von Haushalten in Nürnberg. Dazu liegt ein Datensatz mit Informationen von 1960 bis 2015 vor:

$water_t$	Wasserkonsum pro Haushalt in Kubikmeter (m^3) im Jahr t
$price_t$	Wasserpreis pro Kubikmeter (m^3) in € im Jahr t
$temp_t$	Durchschnittliche Lufttemperatur im Jahr t
$lotsize_t$	Durchschnittliche Gartenfläche in Quadratmeter (m^2) im Jahr t

Es wird folgendes Regressionsmodell aufgestellt und anschließend mit Stata geschätzt:

$$water_t = \beta_1 + \beta_2 ln(price_t) + \beta_3 temp_t + \beta_4 lotsize_t + \beta_5 lotsize_t{}^2 + \varepsilon_t$$

Source	SS	df	MS
Model	688.158228	4	172.039557
Residual	465.556058	51	9.12855015
Total	1153.71429	55	20.9766234

Number of obs	=	56
F(4, 51)	=	18.85
Prob > F	=	0.0000
R-squared	=	0.5965
Adj R-squared	=	0.5648
Root MSE	=	3.0213

Water	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ln_price	-.6156393	.0897841	-6.86	0.000	-.7917348	-.4395438
temp	1.070538	.0873033	12.26	0.000	.8993637	1.241713
lotsize	.3059412	.6343241	4.82	0.000	.1815699	.4303125
lotsize2	-.0016405	.0002589	-6.34	0.000	-.0021481	-.0011328
_cons	.6382662	.1193410	5.35	0.000	.4042772	.8722551

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

1.1 Wie hoch ist der marginale Effekt einer Erhöhung der durchschnittlichen Gartenfläche auf den Wasserkonsum pro Haushalt für eine durchschnittliche Gartenfläche von 50 Quadratmetern? Interpretieren Sie das Ergebnis. [3 Punkte]

Marginaler Effekt von *lotsize*:

$$\frac{\partial \widehat{water_t}}{\partial lotsize_t} = b_4 + b_5 \cdot 2 \cdot lotsize_t$$

Bei $lotsize_t = 50$:

$$\left.\frac{\partial \widehat{water_t}}{\partial lotsize_t}\right|_{lotsize_t=50} = b_4 + b_5 \cdot 2 \cdot 50 = 0,306 + 2 \cdot (-0,002) \cdot 50 = 0,106.$$

Interpretation: Eine Erhöhung der durchschnittlichen Gartenfläche um einen Quadratmeter erhöht c.p. im Mittel den jährlichen Wasserkonsum pro Haushalt um 0,106 m^3 bei einer durchschnittlichen Gartenfläche von 50 Quadratmetern.

1.2 Erläutern Sie verbal den Begriff der Heteroskedastie. Benennen Sie den Inhalt der Gauß-Markow-Annahme des linearen Modells, die bei Heteroskedastie verletzt ist. Welche Folgen hat Heteroskedastie für die geschätzten Parameter und deren Standardfehler? [2,5 Punkte]

- Unter Heteroskedastie ist die Varianz der Störterme nicht konstant: $Var(\varepsilon_i) = \sigma_i^2 \neq \sigma^2$.
- Dies verletzt die Annahme A3: $Var(\varepsilon_i) = \sigma^2$.
- Heteroskedastie beeinflusst die Erwartungstreue des KQ-Schätzers nicht. Die Koeffizienten sind weiterhin unverzerrt geschätzt. (Solange A1 und A2 gelten.)
- Die Standardfehler sind allerdings falsch berechnet. (Die Schätzung ist ineffizient.)

1.3 Der folgende Stata-Output enthält die Ergebnisse eines Breusch-Pagan-Tests auf Heteroskedastie. Führen Sie den Test auf einem Signifikanzniveau von 5% durch. Geben Sie die Hypothesen und Hilfsregression an. Definieren Sie die abhängige Variable der Hilfsregression. Erläutern Sie zudem die Teststatistik und bestimmen Sie Freiheitsgrade, kritischen Wert und Testentscheidung. [6 Punkte]

```
Breusch-Pagan test for heteroskedasticity
        Ho: ???
        Variables: ???

        chi2(?)      =     10.84
        Prob > chi2  =     ????
```

- 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: $e_t^2 = \alpha_1 + \alpha_2 ln(price_t) + \alpha_3 temp_t + \alpha_4 lotsize_t + \alpha_5 lotsize_t^2 + \nu_t$
- Abhängige Variable: $e_t^2 = (y_t - x_t'b)^2$, sind die quadrierten Residuen aus der ursprünglichen Schätzung.
- Teststatistik: $\chi^2_{emp} = T \cdot R^2$, wobei T die Anzahl der Beobachtungen bezeichnet; R^2 ist das Bestimmtheitsmaß aus der Hilfsregression.
- Freiheitsgrade: J=4.
- Kritischer Wert: $\chi^2_{J;\alpha} = \chi^2_{4;0,05} = 9,49$.
- Da $\chi^2_{empirisch} = 10,84 > 9,49 = \chi^2_{kritisch}$ wird die Nullhypothese auf dem 5%-Niveau verworfen. Der Test weist auf Heteroskedastie in dem vorliegenden Modell hin.

1.4 Es sei $Var(\varepsilon_t) = \frac{\sigma^2}{temp_t^3}$. Zeigen Sie formal die aus einer GLS-Transformation resultierende Schätzgleichung, welche zu konstanter Störtermvarianz führt. Leiten Sie zudem die Varianz des Störterms im transformierten Modell her. [2,5 Punkte]

- GLS-Transformation: $water_t \cdot temp_t^{\frac{3}{2}} = \beta_1 \cdot temp_t^{\frac{3}{2}} + \beta_2 ln(price_t) \cdot temp_t^{\frac{3}{2}} + \beta_3 temp_t \cdot temp_t^{\frac{3}{2}} + \beta_4 lotsize_t \cdot temp_t^{\frac{3}{2}} + \beta_5 lotsize_t{}^2 \cdot temp_t^{\frac{3}{2}} + \varepsilon_t \cdot temp_t^{\frac{3}{2}}$
- $Var(\varepsilon_t \cdot temp_t^{\frac{3}{2}}) = temp_t^3 \cdot Var(\varepsilon_t) = temp_t^3 \cdot \frac{\sigma^2}{temp_t^3} = \sigma^2$

1.5 Führen Sie einen Durbin-Watson-Test auf positive Autokorrelation 1. Ordnung auf einem Signifikanzniveau von 5% durch. Geben Sie die Hypothesen, die Teststatistik, die untere und obere Grenze für den kritischen Wert und die Testentscheidung an. *Hinweis:* $\sum_{t=2}^{T} (e_t - e_{t-1})^2 = 600$. [6 Punkte]

- Hypothesen: $H_0 : \rho \leq 0$; $H_1 : \rho > 0$ (positive Autokorrelation 1.Ordnung)
- $SSR = 465,556$
- Teststatistik: $dw = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} = \frac{600}{465,556} = 1,289$
- Untere und obere Grenze für $T = 56$ und $K = 5$ nicht tabelliert.
  Nächstmögliche Werte: $d_{L;T=55;K=5} = 1,41$ und $d_{U;T=55;K=5} = 1,72$.
- Da $dw = 1,289 < d_L = 1,41$, wird die Nullhypothese auf dem 1%-Niveau verworfen. Der Test weist auf positive Autokorrelation 1. Ordnung in dem vorliegenden Modell hin.

Aufgabe 2: **[20 Punkte]**

Sie interessieren sich für die Determinanten von Löhnen in Deutschland. Hierfür sind Ihnen Daten über 10.366 Angestellte gegeben:

$lnwage_i$	Logarithmierter Bruttostundenlohn von Person i
$educ_i$	Bildungsjahre von Person i
age_i	Alter von Person i
$female_i$	Dummy-Variable, =1 wenn Person i eine Frau ist, =0 Mann
$bayern_i$	Dummy-Variable, =1 wenn Person i Einwohner von Bayern ist, =0 sonst

Es wird folgendes Regressionsmodell aufgestellt und anschließend mit Stata geschätzt:

$$lnwage_i = \beta_1 + \beta_2 educ_i + \beta_3 age_i + \beta_4 age_i^2 + \beta_5 female_i + \varepsilon_i$$

```
      Source |       SS           df       MS      Number of obs   =    10,366
-------------+----------------------------------   F(4, 10361)     =    398.81
       Model |  14.5892618         4  3.64731546   Prob > F        =    0.0000
    Residual |  94.7556373    10,361  .009145414   R-squared       =    ??????
-------------+----------------------------------   Adj R-squared   =    0.1331
       Total |  109.344899    10,365  .010549436   Root MSE        =    .09563

------------------------------------------------------------------------------
      lnwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0175752   .0007643    22.99   0.000     .0162198    .0192069
         age |   .0202595   .0010111    20.04   0.000     ????????    ????????
        age2 |  -.0002052   .0000113   -18.15   0.000    -.0002273    -.000183
      female |  -.0392574    .001881   -20.87   0.000    -.0429446   -.0355703
       _cons |   7.720102   .0236457   326.49   0.000     7.673752    7.766452
------------------------------------------------------------------------------
```

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

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

- Frauen verdienen c.p. im Mittel etwa $b_5 \cdot 100\%$= 3,9% weniger als Männer.
- Der Koeffizient ist statistisch signifikant auf dem 1% Niveau.

2.2 Berechnen Sie das 95%-Konfidenzintervall des Koeffizienten von *age*. Wie lässt sich das Konfidenzintervall interpretieren? [2,5 Punkte]

- t-Wert in Tabelle ablesen: 1,962 (df=1000, $\frac{\alpha}{2} = 0{,}975$).
- Obere Grenze: $0{,}020 + 1{,}962 \cdot 0{,}001 = 0{,}022$.
- Untere Grenze: $0{,}020 - 1{,}962 \cdot 0{,}001 = 0{,}018$.
- (Das 95%-Konfidenzintervall des Koeffizienten von *age* lautet: [0,016;0,019])
- Interpretation: Mit wiederholten Stichproben liegt das wahre β_{age} in 95% 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) $= 1 - \frac{RSS}{TSS} = \frac{MSS}{TSS} = \frac{14{,}589}{109{,}345} = 0{,}133$.
- Das vorliegende Modell erklärt 13,3% der Variation des logarithmierten Lohns.

2.4 Im Bundesland Bayern wurde eine Reform durchgeführt, bei der die Schulzeit für die Schülerschaft um ein Jahr verlängert wurde. Handelt es sich bei der Variable *bayern* um eine geeignete Instrumentenvariable für Bildung? Erklären Sie kurz die Bedingungen, die hierfür erfüllt sein müssen. [2,5 Punkte]

- Relevanz: Das Instrument muss mit der endogenen Variable korreliert sein.
- Exogenität: Das Instrument darf nicht mit unbeobachteten Determinanten der abhängigen Variable und der endogenen Variable korreliert sein.

=> Im vorliegenden Fall handelt es sich vermutlich um ein geeignetes Instrument. (Es ist zwangsläufig relevant, da es mit der Bildung korreliert ist und vermutlich exogen)

- Alternativerklärung (ebenfalls bepunktet): Es handelt sich um kein geeignetes Instrument, da die Exogenitätsannahme etwa verletzt wäre, wenn gebildete Eltern in Grenznähe ihre Kinder bewusst in eine Schule in Bayern schicken, um von dem zusätzlichen Jahr an Bildung zu profitieren.

2.5 Sie entscheiden sich dafür, die Dummyvariable für das Bundesland Bayern als Instrument in einer Two Stage Least Squares (2SLS)-Schätzung zu verwenden. Erläutern Sie kurz verbal die Vorgehensweise des 2SLS-Schätzers und stellen Sie die beiden benötigten Modellgleichungen auf. [4 Punkte]

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

1.Stufe:

$$educ_i = \theta_1 + \theta_2 bayern_i + \theta_3 age_i + \theta_4 {age_i}^2 + \theta_5 female_i + \upsilon_i$$

2.Stufe:

$$lnwage_i = \beta_1 + \beta_2 \widehat{educ_i} + \beta_3 age_i + \beta_4 {age_i}^2 + \beta_5 female_i + u_i$$

2.6 Sie schätzen einen F-Test für die Signifikanz der Instrumentenvariable auf der ersten Stufe und erhalten einen F-Wert von 26,78. Handelt es sich in dem vorliegenden Fall um ein schwaches Instrument? Erläutern Sie in diesem Kontext den Begriff schwacher Instrumente. [2 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. Stellen Sie die Gleichung der benötigten Hilfsregression für das vorliegende Modell auf. [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).
- Hilfsregression:
  ($\hat{\upsilon}_i$: Residuen der Schätzung auf der ersten Stufe)

$$lnwage_i = \beta_1 + \beta_2 educ_i + \beta_3 age_i + \beta_4 age_i^2 + \beta_5 female_i + \gamma\hat{\upsilon}_i + \varepsilon_i$$

2.8 Sie führen nun den Durbin-Wu-Hausman Test in Stata durch und erhalten folgenden Output:

```
Tests of endogeneity
Ho: ???

Durbin (score) chi2(1)          =  12.6232  (p = 0.0004)
Wu-Hausman F(1,10378)           =  12.6313  (p = 0.0004)
```

Wie lautet Nullhypothese des Durbin-Wu-Hausman-Tests? Zu welchem Ergebnis kommt er hinsichtlich der Exogenität der Bildungsvariable? [2 Punkte]

- H0: $educ_i$ ist exogen.
- (Alternativ: γ = 0).
- H0 wird abgelehnt -> es wird davon ausgegangen, dass die Bildung nicht exogen ist.

Aufgabe 3: **[16 Punkte]**

Sie interessieren sich für den Zusammenhang zwischen dem Abstimmungsverhalten britischer Bürger beim Brexit-Referendum und persönlichen und sozio-ökonomischen Charakteristika. Ihnen liegen Daten aus der Befragung 1.000 aktiver Wähler vor, mit Informationen zu

$vote_i$	Dummy-Variable, =1 wenn Person i für Brexit gestimmt hat, =0 sonst
age_i	Alter von Person i
$unempl_i$	Dummy-Variable, =1 wenn Person i arbeitslos ist, =0 sonst
$ln\_townsize_i$	logarithmierte Einwohnerzahl des Wohnortes von Person i
$english_i$	Dummy-Variable, =0 wenn Person i Einwohner Englands ist, =0 sonst

Die dazugehörigen deskriptiven Statistiken sind in der unteren Tabelle dargestellt:

Variable	Obs	Mean	Std.Dev.	Min.	Max.
vote	1000	.516	.499994	0	1
age	1000	55.369	11.52778	18	92
unempl	1000	.159	.3658591	0	1
ln_townsize	1000	12.82776	.9746702	6.898	13.815
english	1000	.134	.3408228	0	1

Mithilfe der Maximum-Likelihood-Methode schätzen Sie folgendes Regressionsmodell:

$$P(vote_i = 1|\mathbf{x_i}) = \Lambda(\beta_1 + \beta_2 age_i + \beta_3 unempl_i + \beta_4 ln\_townsize_i + \beta_5 english_i)$$

```
Logistic regression                               Number of obs   =       1000
                                                  LR chi2(4)      =     449.16
                                                  Prob > chi2     =     0.0000
Log likelihood = -468.05698                       Pseudo R2       =     0.3242

------------------------------------------------------------------------------
vote        |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------+-----------------------------------------------------------------
age         |   .1043212   .0087577    11.91   0.000     .0871564    .1214859
unempl      |   3.162236   .4126543     7.66   0.000     2.353448    3.971024
ln_townsize |  -.0453689   .0814242    -0.56   0.577    -.2049574    .1142196
english     |   2.286329   .3145702     7.27   0.000     1.669783    2.902875
_cons       |  -5.663914   1.170498    -4.84   0.000    -7.958047   -3.369781
------------------------------------------------------------------------------
```

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

3.1 Welche zwei Probleme ergeben sich allgemein bei der Schätzung eines linearen Wahrscheinlichkeitsmodells? [2 Punkte]

- Die Störterme sind heteroskedastisch und nicht normalverteilt.
- $\hat{y}_i$ ist oft größer als 1 oder kleiner als 0.

3.2 Welche Annahme bezüglich der Verteilung der Störgröße wurde in obiger Schätzung getroffen? Nennen Sie eine alternative Annahme. [1 Punkt]

- Oben getroffene Annahme: standardlogistisch verteilte Störgröße (Logit-Modell).
- Alternative Annahme: standardnormalverteilte Störgröße (Probit-Modell).

3.3 Erläutern Sie das Maximum-Likelihood-Schätzverfahren verbal. Gehen Sie dabei auf die Annahmen und die Schritte zur Bestimmung der Parameter ein. [3 Punkte]

- Das Verfahren wählt die Parameter so, dass die Wahrscheinlichkeit, die Stichprobe mit den Werten für Y_i zu beobachten, gegeben der Werte von x_i, maximiert wird.
- Es wird eine Annahme bzgl. der Verteilung der Störgröße getroffen und darauf basierend die Likelihood-Funktion gebildet. Ferner wird Unabhängigkeit der Beobachtungen angenommen.
- Die Likelihood-Funktion wird logarithmiert und durch Nullsetzen der Ableitung in Hinblick auf die Parameter maximiert.

3.4 Berechnen und interpretieren Sie den marginalen Effekt der Variable age_i am Mittel der Daten. [5 Punkte]

- Marginaler Effekt $\frac{\partial E[y|X]}{\partial age_i} = \frac{e^{x_i'\beta}}{(1+e^{x_i'\beta})^2}\beta_2$ am Mittel der Daten: $\overline{x_i}'\beta = -5{,}664 + 0{,}104 \cdot 55{,}369 + 3{,}162 \cdot 0{,}159 + 2{,}287 \cdot 0{,}134 - 0{,}045 \cdot 12{,}828 = 0{,}326$ also $\frac{\partial E[y|X]}{\partial age_i} = \frac{e^{0{,}326}}{(1+e^{0{,}326})^2} \cdot 0{,}104 = 0{,}025$
- Ceteris paribus erhöht ein weiteres Altersjahr am Mittel der Daten die Wahrscheinlichkeit, für den Brexit gestimmt zu haben, durchschnittlich um 2,5 Prozentpunkte.

3.5 McFadden's R^2 wird im Stata-Output als Pseudo-R^2 bezeichnet. Berechnen Sie die Log-Likelihood des Modells, wenn es nur mit Konstante geschätzt wird. Welches Modell beschreibt die Daten besser? [2 Punkte]

- McFadden $R^2 = 1 - \frac{L_1}{L_0}$, wobei L_1 der Wert für die Log-Likelihood des unrestringierten und L_0 des restringierten Modells ist. Also ist $L_0 = \frac{L_1}{1-R^2}$.
- $L_0 = \frac{-468{,}057}{1-0{,}324} = -690{,}913$.
- Das unrestringierte Modell beschreibt die Daten besser als das restringierte Modell.

3.6 Wie hoch ist die vorhergesagte Wahrscheinlichkeit dafür, dass ein 60-jähriger, arbeitender Engländer aus einer Stadt mit 2.000 Einwohnern für den Brexit stimmt? [3 Punkte]

- $(x'\beta|age_i = 60, unempl_i = 0, english_i = 1, townsize_i = 2000) = -5{,}663 + 60 \cdot 0{,}104 + 2{,}286 - 7{,}601 \cdot 0{,}045 = 2{,}521$.
- $Prob(vote_1 = 1|age_i = 60, unempl_i = 0, english_i = 1, townsize_i = 2000) = \frac{1}{1+e^{-2{,}521}} = 0{,}926 \approx 93\%$.

Aufgabe 4: [4 Punkte]

4.1 Vorbereitete Studierende bestehen immer die Ökonometrie-Klausur, wohingegen nicht vorbereitete Studierende nicht bestehen. Gegeben ist folgende Likelihood-Funktion: $L(\theta) = \binom{n}{k} \cdot \theta^k \cdot (1-\theta)^{n-k}$ wobei n die Stichprobengröße, k die Anzahl an bestandenen Prüfungen und θ die Wahrscheinlichkeit, dass ein Studierender vorbereitet ist, bezeichnet. Es werden vier Prüfungen geschrieben, davon werden drei bestanden. Ermitteln Sie mithilfe der Maximum-Likelihood-Methode $\hat{\theta}$. Hinweis: $\binom{n}{k} = \binom{4}{3} = 4$. [4 Punkte]

- $L(\theta) = \binom{n}{k} \cdot \theta^k \cdot (1-\theta)^{n-k} = \binom{4}{3} \cdot \theta^3 \cdot (1-\theta)^{4-3} = 4 \cdot \theta^3 \cdot (1-\theta)^1$
- $logL = log(4 \cdot \theta^3 \cdot (1-\theta)^1) = log4 + 3log(\theta) + log(1-\theta)$
- $\frac{\partial logL}{\partial \theta} = \frac{3}{\theta} - \frac{1}{1-\theta} = 0$
- $\hat{\theta} = \frac{3}{4}$
- $\hat{\theta} = \frac{3}{4}$ maximiert die Wahrscheinlichkeit, die Stichprobe zu erhalten.

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

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

3.	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.

4.	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**

5.	Für die Schätzgleichung $\hat{y}_i = 2,5 - 4 \cdot x_{i2} + 2 \cdot x_{i3}$ und die Beobachtung $(y_i, x_{i2}, x_{i3}) = (-1, 1, 1)$ beträgt das Residuum
a	-2,5.
b	-1,5. **X**
c	-0,5.
d	0,5.

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

7.	Gegeben ist folgendes Modell: $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i$. Die aufgestellten Hypothesen lauten $H_0 : \beta_3 \leq 3$ vs. $H_1 : \beta_3 > 3$. Auf einem Signifikanzniveau von 5% führt eine t-Teststatistik von $1,72$
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.

8.	Gegeben ist folgendes Modell: $y_i = \beta_1 + \beta_2 x_{i2} + u_i$. b_2 beträgt -4 und $\widehat{V}(b_2) = 0,25$. Die aufgestellten Hypothesen lauten $H_0 : \beta_2 \leq -5$ vs. $H_1 : \beta_2 > -5$. Die dazugehörige t-Teststatistik beträgt
a	$t^{empirisch} = 36$.
b	$t^{empirisch} = -4$.
c	$t^{empirisch} = 4$.
d	$t^{empirisch} = 2$. **X**

9.	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.

10.	In welchem Fall wird b_2 im Modell $y_i = \beta_1 + \beta_2 x_{2i} + \varepsilon_i$, in dem die Variable x_{3i} mit Koeffizient β_3 ausgelassen wurde, nach oben verzerrt geschätzt?
a	$\beta_3 < 0$ und $\text{Cov}(x_{i2}; x_{i3}) > 0$.
b	$\beta_2 > 0$ und $\text{Cov}(x_{i2}; x_{i3}) < 0$.
c	$\beta_3 < 0$ und $\text{Cov}(x_{i2}; x_{i3}) < 0$. **X**
d	$\beta_2 > 0$ und $\text{Cov}(x_{i2}; x_{i3}) > 0$.

11.	Von einem überidentifizierten Modell spricht man, wenn
a	auch die exogenen Regressoren in der ersten Stufe stark mit der endogenen Variable korrelieren.
b	der F-Wert eines Instruments in der ersten Stufe > 10 ist.
c	mehr Instrumente als endogene Regressoren vorliegen. **X**
d	sich OLS und 2SLS Koeffizienten signifikant unterscheiden.

12.	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**

13.	Unter welcher Annahme ist der KQ-Schätzer konsistent?
a	$\{X_1, ..., X_N\}$ und $\{\varepsilon_1, ...\varepsilon_N\}$ sind statistisch unabhängig. **X**
b	$V\{\varepsilon_i\} = \sigma^2$.
c	$\varepsilon_i \sim NID(0, \sigma^2)$.
d	$\varepsilon_t \sim IID(0, \sigma^2)$.

14.	In welchem Fall ist die kausale Interpretation einer KQ-Schätzung nicht gerechtfertigt?
a	Bei der Verwendung einer binären abhängigen Variable.
b	Bei Heteroskedastie.
c	Bei Vorliegen eines omitted variable bias. **X**
d	Bei der Verwendung robuster Standardfehler.

15.	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**

16.	Sie schätzen das Modell $ln(Stundenlohn_i) = \beta_1 + \beta_2 migrant_i + \beta_3 educ_i + \beta_4 (migrant_i \cdot educ_i) + \varepsilon_i$ mittels einer KQ-Schätzung ($educ_i$ misst Bildung in Jahren, $migrant_i = 1$ falls Migrant). Welche Aussage über die Schätzung trifft zu?
a	b_3+b_4 gibt die geschätzte Bildungsrendite für Migranten an. **X**
b	b_4 gibt die geschätzte Bildungsrendite für Migranten an.
c	b_1 gibt den durchschnittlichen Stundenlohnen für Migranten mit 0 Bildungsjahren an.
d	b_2 wird als Lohnunterschied zwischen Migranten und der restlichen Stichprobe in % interpretiert.

17.	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 Instrumentenvariable.
d	die vorhergesagten Werte der abhängigen Variable aus der ersten Stufe. **X**

18.	Welche Bedingung muss eine Instrumentenvariable unter anderem erfüllen?
a	Endogenität.
b	Exogenität.**X**
c	Korrelation mit der abhängigen Variable in der zweiten Stufe.
d	Kleine Varianz.

19.	Nicht-systematische Messfehler in einem Regressor in der KQ-Schätzung eines einfachen Modells führen zu
a	einer Verzerrung des geschätzten Koeffizienten gegen 0. **X**
b	einer Verzerrung des geschätzten Koeffizienten in eine unbekannte Richtung.
c	unverzerrten Koeffizientenschätzern.
d	einer konsistenten, aber nicht erwartungstreuen Schätzung des Steigungsparameters.

20.	Wann wird der Sargan-Test angewendet?
a	Bei jeder Schätzung mit Instrumentenvariablen.
b	Wenn mehr Instrumente als endogene Regressoren vorliegen. **X**
c	Bei ausgelassenen Variablen.
d	Bei autokorrelierten Störtermen.

21.	Gegeben ist folgendes Modell: $lnwage_i = \beta_1 + \beta_2 schooling_i + \beta_3 gender_i + \varepsilon_i$. Sie möchten einen White-Test auf Heteroskedastie durchführen. Für wie viele Freiheitsgrade müssen Sie die kritischen Werte der χ^2-Verteilung nachschlagen?
a	3.
b	5. **X**
c	6.
d	12.

22.	Eine große Differenz zwischen den geschätzten Koeffizienten im Logit- und Probit-Modell
a	ist ein Hinweis auf Missspezifikation des Modells.
b	bedeutet nicht notwendigerweise einen großen Unterschied in den marginalen Effekten. **X**
c	deutet auf eine Verletzung der getroffenen Verteilungsannahmen hin.
d	tritt auf, wenn die Standardfehler inkorrekt geschätzt werden.

23.	Sie schätzen die Lohngleichung $wage_i = \beta_1 + \beta_2 exper_i + \beta_3 exper_i^2 + \varepsilon_i$ und erhalten $b_1 = 5$, $b_2 = 0{,}5$ und $b_3 = -0{,}01$. Ab welchem Wert von *exper* ist der marginale Effekt von *exper* auf den Lohn negativ?
a	10.
b	25. **X**
c	50.
d	100.

24.	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**

25.	Welche Aussage bezüglich Inferenz im ML-Rahmen ist korrekt?
a	Der Wald-Test wird nach restringierter Schätzung durchgeführt.
b	Beim Likelihood-Ratio-Test werden das restringierte und das unrestringierte Modell verglichen. **X**
c	Asymptotisch liefern Wald-, Likelihood-Ratio- und Lagrange-Multiplier-Test unterschiedliche Ergebnisse.
d	Die Teststatistik des Wald-Tests folgt bei Zutreffen der Nullhypothese asymptotisch der F-Verteilung.

26.	Die Determinante der Matrix $\mathbf{A} = \begin{pmatrix} 2 & 5 \\ -5 & 1 \end{pmatrix}$ lautet
a	-23.
b	5.
c	15.
d	27. **X**

27.	Die Matrix $\mathbf{B} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ ist
a	singulär.
b	idempotent. **X**
c	eine Einheitsmatrix.
d	nicht identisch mit ihrer Transponenten.

28.	Welche Aussage bezüglich der Maximum-Likelihood-Schätzung ist zutreffend?
a	Unter den Gauß-Markow-Annahmen sind bei Annahme standardnormalverteilter Störterme der KQ- und der ML-Schätzer für die Steigungsparameter β identisch. **X**
b	In der ML-Schätzung wird die Varianz σ^2 unverzerrt und konsistent geschätzt.
c	Wenn der ML-Schätzer konsistent geschätzt werden kann, so erreicht der KQ-Schätzer stets die Cramer-Rao-Lower-Bound.
d	ML-Schätzer können im Gegensatz zu KQ-Schätzern stets analytisch bestimmt werden.

29.	Im Fall einer an der Stelle $\hat{\beta}$ stark gekrümmten global konkaven Log-Likelihood-Funktion
a	kann der ML-Schätzer nicht analytisch ermittelt werden.
b	kann β präzise geschätzt werden. **X**
c	liegt Heteroskedastie vor.
d	ist das Modell falsch spezifiziert.

30.	Gegeben ist folgendes Modell: $pruefungsnote_i = \beta_1 + \beta_2 buecher_i + \beta_3 lerntage_i + \varepsilon_i$ wobei *buecher* die Anzahl gelesener Bücher zur Prüfungsvorbereitung und *lerntage* die Anzahl an Lerntagen bezeichnet. Jeder Student in der Stichprobe lernt ausschließlich durch Bücherlesen und pro Lerntag werden 5 Bücher gelesen. Das Modell
a	kann mittels GLS-Transformation oder FGLS geschätzt werden.
b	kann mittels KQ- sowie ML-Methode geschätzt werden.
c	weist bei unabhängiger Stichprobenziehung Autokorrelation auf.
d	kann in der vorliegenden Form nicht geschätzt werden. **X**