Masterprüfung WS 2017/2018 - 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: **[17 Punkte]**

Sie möchten die Determinanten des Radwegeausbaus in deutschen Gemeinden analysieren. Dazu liegt Ihnen ein Querschnittsdatensatz mit Beobachtungen zu 500 Kommunen in Deutschland mit folgenden Merkmalen vor:

$Fahrradwege_i$	Länge des Fahrradwegenetzes in Kommune i, in km
$Einwohnerzahl_i$	Einwohnerzahl in Kommune i
$Verkehrsausgaben_i$	Verkehrsausgaben von Kommune i, in 1000 Euro
$Entfernung_i$	Entfernung von Kommune i zur nächsten Großstadt, in km
$Anteil\_Grüne_i$	Stimmenanteil grüner Parteien bei der letzten Kommunalwahl in Kommune i, in Prozent

Sie schätzen folgendes Modell mit der KQ-Methode:

$$ln(Fahrradwege_i) = \beta_1 + \beta_2 Einwohnerzahl_i + \beta_3 ln(Verkehrsausgaben_i) + \beta_4 Entfernung_i + \beta_5 Entfernung_i^2 + \varepsilon_i$$

```
Source               |       SS           df       MS      Number of obs   =       500
---------------------+----------------------------------   F(4, 495)       =    434.01
Model                |  1498.62906         4  374.657265   Prob > F        =    0.0000
Residual             |  427.305562       495   .86324356   R-squared       =    0.7781
---------------------+----------------------------------   Adj R-squared   =    0.7763
Total                |  1925.93462       499  3.85958842   Root MSE        =    .92911

--------------------------------------------------------------------------------------
ln_fahrradwege       |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------------+----------------------------------------------------------------
einwohnerzahl        |   .0001067   2.95e-06    36.20   0.000     .0001009    .0001125
ln_verkehrsausgaben  |   .0851434   .0227248     3.75   0.000     .0404944    .1297924
entfernung           |   .0151998   .0185444     0.82   0.413    -.0212356    .0516352
entfernung2          |  -.0010167   .0002298    -4.42   0.000    -.0014681   -.0005652
_cons                |   2.153906   .3611879     5.96   0.000     1.444256    2.863557
--------------------------------------------------------------------------------------
```

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

1.1 Interpretieren Sie den Koeffizientenschätzer $\hat{\beta}_3$ inhaltlich und statistisch. [2 Punkte]

- Steigen die Verkehrsausgaben um 1%, so erhöht sich Länge des Fahrradwegenetzes c.p. im Mittel um 0,085%.
- Der Effekt ist am 1%-Niveau statistisch signifikant.

1.2 Berechnen Sie den marginalen Effekt der Variable $Entfernung_i$ auf die abhängige Variable. Bei welchem Wert von $Entfernung_i$ beträgt der marginale Effekt 0? [2 Punkte]

- marginaler Effekt: $0{,}015 - 2 \cdot 0{,}001 \cdot Entfernung_i = 0{,}015 - 0{,}002 \cdot Entfernung_i$
- $0{,}015 - 0{,}002 \cdot Entfernung_i = 0 \Rightarrow Entfernung_i = 7{,}5$

1.3 Benennen Sie den Inhalt einer Annahme, die im Fall von Endogenität einer erklärenden Variablen verletzt ist, sowie die daraus resultierenden Konsequenzen für die KQ-Schätzung. [2 Punkte]

- Bei Endogenität ist die Annahme $E[x_{ij}\varepsilon_i] = 0$ verletzt, d.h. eine erklärende Variable ist mit dem Störterm korreliert. (Alternativ können Annahme 2 oder Annahme 10 genannt werden.)
- In der Folge ist der KQ-Schätzer verzerrt und inkonsistent.

1.4 Nennen Sie zwei mögliche Ursachen für Endogenität und erläutern Sie diese kurz am Beispiel der Variable $Verkehrsausgaben_i$. [3 Punkte]

- Omitted variable bias: Eine relevante erklärende Variable, die mit $Verkehrsausgaben_i$ korreliert, wird nicht im Modell berücksichtigt.
- Beispiel: geografische Lage von Kommunen, die sowohl die Radwegnetzlänge beeinflusst als auch insgesamt mit den Verkehrsausgaben korreliert ist.
- Umgekehrte Kausalität: Die abhängige Variable beeinflusst die erklärende Variable.
- Beispiel: ein großes Radwegenetz erfordert Instandhaltung und hat somit einen Effekt auf die Verkehrsausgaben.

1.5 Jemand schlägt vor, die potentiell endogene Variable $Verkehrsausgaben_i$ mit der Variable $Anteil\_Grüne_i$ zu instrumentieren. Erklären Sie kurz die Bedingungen, die die Variable $Anteil\_Grüne_i$ erfüllen muss, um ein geeignetes Instrument für $Verkehrsausgaben_i$ zu sein. Sind diese erfüllt? Begründen Sie Ihre Entscheidung. [4 Punkte]

- Relevanz: Die Variable $Anteil\_Grüne_i$ muss mit der endogenen Variable $Verkehrsausgaben_i$ korreliert sein. Dies ist wahrscheinlich, da grüne Parteien häufig eine Erhöhung der Ausgaben für Verkehr fordern oder durchsetzen (eine gegenteilige Korrelation ist ebenso denkbar).
- Exogenität: Die Variable $Anteil\_Grüne_i$ darf nicht mit dem Störterm ε_i korrelieren. Dies ist fraglich, da der Stimmenanteil grüner Parteien vom Ausbauzustand des Radwegenetzes abhängen könnte (umgekehrte Kausalität). (Alternative Begründung: ausgelassene relevante Variabe, welche mit $Anteil\_Grüne_i$ korreliert.) $Anteil\_Grüne_i$ wäre endogen und somit kein gültiges Instrument.

1.6 Nehmen Sie an, dass die Variable $Anteil\_Grüne_i$ ein gültiges Instrument für $ln(Verkehrsausgaben_i)$ darstellt. Beschreiben Sie, wie Sie mittels des Durbin-Wu-Hausman-Tests die Variable $ln(Verkehrsausgaben_i)$ auf Endogenität prüfen können. Geben Sie dazu auch die zwei benötigten Regressionsgleichungen an. [4 Punkte]

- In einer Hilfsregression wird die potentiell endogene Variable auf die exogenen erklärenden Variablen und das Instrument regressiert:

$$ln(Verkehrsausgaben_i) = \theta_1 + \theta_2 Einwohnerzahl_i + \theta_3 Entfernung_i + \theta_4 Entfernung_i^2 + \theta_5 Anteil\_Grüne_i + \eta_i$$

- Die Residuen $\hat{\eta}_i$ werden bestimmt und folgendes Modell mit KQ geschätzt:

$$ln(Fahrradwege_i) = \beta_1 + \beta_2 Einwohnerzahl_i + \beta_3 ln(Verkehrsausgaben_i) + \beta_4 Entfernung_i + \beta_5 {Entfernung_i}^2 + \gamma\hat{\eta}_i + \varepsilon_i$$

- Ist $\hat{\gamma}$ statistisch signifikant von Null verschieden, weist dies auf Endogenität der Variable $ln(Verkehrsausgaben_i)$ hin. Kann $H_0 : \gamma = 0$ nicht verworfen werden, wird $ln(Verkehrsausgaben_i)$ als exogen angenommen.

Aufgabe 2: **[14,5 Punkte]**

Sie interessieren sich für die Determinanten des Schneefalls an Weihnachten. Dazu liegt Ihnen ein Datensatz mit Beobachtungen zu Wetter und geografischen Gegebenheiten von 100 Urlaubsorten in Deutschland in 2017 vor:

$snow_i$	=1 wenn es am 24.12.2017 in Ort i Schneefall gab, =0 sonst
$height_i$	Höhe des Urlaubsortes i, in 100m über NN
$seadist_i$	Entfernung des Ortes i zum nächstgelegenen Meer, in km
$temp_i$	Durchschnittstemperatur in Ort i im Dezember 2017, in Grad Celsius

Die dazugehörigen deskriptiven Statistiken sind in der unteren Tabelle dargestellt:

Variable	Obs.	Mean	Std.Dev.	Min.	Max.
snow	100	0,560	0,4989	0	1
height	100	4,6644	3,3890	0,03	16,97
seadist	100	279,31	126,0827	14	499
temp	100	0,2	-2,6931	-8	6

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

$$P(snow_i = 1|\mathbf{x_i}) = \Lambda(\beta_1 + \beta_2 height_i + \beta_3 seadist_i + \beta_4 temp_i)$$

```
Logistic regression                               Number of obs     =        100
                                                  LR chi2(3)        =      85.61
                                                  Prob > chi2       =     0.0000
Log likelihood = -25.786634                       Pseudo R2         =     0.6241

------------------------------------------------------------------------------
snow         |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
height       |   1.330191   .2900191     4.59   0.000     .7617645    1.898618
temp         |  -.3084108   .1775614    -1.74   0.082    -.6564246    .0396031
seadist      |   .0019583   .0029037     0.67   0.500    -.0037328    .0076494
_cons        |  -5.334966   1.321067    -4.04   0.000    -7.924209   -2.745722
------------------------------------------------------------------------------
```

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

2.1 Berechnen und interpretieren Sie den marginalen Effekt der Variable $height_i$ 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:
- $\overline{x_i}'\beta = -5,335 + 1,330 \cdot 4,664 - 0,308 \cdot 0,2 + 0,002 \cdot 279,31 = 1,365$
- $\frac{\partial E[y|X]}{\partial age_i} = \frac{e^{1,365}}{(1+e^{1,365})^2} \cdot 1,330 = 0,216$
- Ceteris paribus erhöhen 100 zusätzliche Höhenmeter am Mittel der Daten die Wahrscheinlichkeit für Schneefall um 21,6 Prozentpunkte.

2.2 Wie hoch ist die vorhergesagte Wahrscheinlichkeit für Schneefall an Heiligabend 2017 in Oberstdorf (Höhe über NN: 813 m, Durchschnittstemperatur Dezember 2017: -1,3 Grad Celsius, Entfernung zum nächstgelegenen Meer: 283 km)? [3 Punkte]

- $(x'\beta|height_i = 8,13, temp_i = -1,3, seadist_i = 283) = -5,335 + 1,330 \cdot 8,13 - 0,308 \cdot (-1,3) + 0,002 \cdot 283 = 6,444$.
- $Prob(snow_i = 1|height_i = 8,13, temp_i = -1,3, seadist_i = 283) = \frac{1}{1+e^{-6,444}} = 0,998 \approx 100\%$.

2.3 Nennen Sie einen Grund, der gegen eine Aufnahme der Variable $temp_i$ in logarithmierter Form spricht. [1 Punkt]

- Die Variable nimmt auch negative Werte an. Da diese nicht logarithmiert werden können ergeben sich fehlende Werte. Dadurch reduziert sich die Anzahl der Beobachtungen in der ohnehin schon kleinen Stichprobe.

2.4 Ein befreundeter Geologe weist Sie auf ein potentielles Multikollinearitätsproblem hin. Nennen Sie eine mögliche Kombination kollinearer Variablen im vorliegenden Modell. Was ist die Konsequenz hoher Kollinearität? [1,5 Punkte]

- Temperatur und Höhe, alternativ: Entfernung zum Meer und Höhe, Temperatur und Entfernung zum Meer
- unpräzise Schätzung, hohe Standardfehler der Schätzung

2.5 Erläutern Sie das Prinzip des Likelihood-Ratio-Tests. Berechnen Sie anschließend basierend auf den gegebenen Informationen den Wert der Log-Likelihood $lnL(\tilde{\beta})$ eines vollständig restringierten Modells (eines Modells, welches nur mit der Konstanten als erklärende Variable geschätzt wird) im vorliegenden Fall. [4 Punkte]

- Der Likelihood-Ratio-Test misst die relative Güte zweier (genesteter) Modelle durch Vergleichen der Werte der Log-Likelihood.
- Ist der Wert der Likelihood des unrestringierten Modells signifikant größer (betragsmäßig kleiner) als der des restringierten Modells, gelten die im restringierten Modell unterstellten Restriktionen nicht und das unrestringierte Modell ist zu bevorzugen.
- Die LR-Teststatistik für den Test der Nullhypothese $\beta_2 = \beta_3 = \beta_4 = 0$ gegen die Alternativhypothese, dass mindestens einer der Parameter von Null verschieden ist, beträgt 85,61.
- Log-Likelihood des restringierten Modells: $lnL(\tilde{\beta}) = -\frac{85,61}{2} - 25,786 = -68,591$

Aufgabe 3: **[17 Punkte]**

Unterstellen Sie das folgende einfache Modell der Geldnachfrage:

$$log(M1_t) = \beta_1 + \beta_2 log(GDP_t) + \beta_3 log(CPI_t) + \varepsilon_t$$

wobei

$M1_t$	nominaler Wert der M1-Geldmenge im Jahr t (in Mrd. US-$)
GDP_t	reales Bruttosozialprodukt im Jahr t (in Mrd. US-$)
CPI_t	Preisindex im Jahr t

Das Modell wird mit US-Quartalsdaten von 1980Q1 bis 1999Q4 geschätzt. Aus einer KQ-Schätzung ergibt sich: $\sum_{t=2}^{T} \hat{e}_t \hat{e}_{t-1} = 1,264$ und $\sum_{t=2}^{T} \hat{e}_{t-1}^2 = 1,323$

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

$\hat{\rho} = \frac{\sum_{t=2}^{T} \hat{e}_t \hat{e}_{t-1}}{\sum_{t=2}^{T} \hat{e}_{t-1}^2} = \frac{1{,}264}{1{,}323} = 0{,}955$.

$dw \approx 2(1-\hat{\rho}) = 2 - 2 \cdot 0{,}955 = 0{,}09$.

3.2 Unterstellen Sie für die Durbin-Watson-Statistik einen Wert von 1,45 und testen Sie auf dem 5%-Signifikanzniveau auf positive Autokorrelation erster Ordnung. Geben Sie hierzu auch die Nullhypothese, die Alternativhypothese sowie die Freiheitsgrade und die jeweiligen kritischen Werte an. [4 Punkte]

- Nullhypothese: $H_0 : \rho = 0$
- Alternativhypothese: $H_1 : \rho > 0$
- Freiheitsgrade: T=80 und K=3.
- Kritische Werte: bei T=80 und K=3: $d_L = 1{,}59; d_U = 1{,}69$
- Die Durbin-Watson Statistik von 1,45 liegt unter der unteren Grenze von 1,59; die Nullhypothese, dass keine positive Autokorrelation erster Ordnung vorliegt kann auf dem 5%-Signifikanzniveau somit verworfen werden.

3.3 Führen Sie einen Breusch-Pagan-Test auf Heteroskedastie für das Modell auf einem Signifikanzniveau von 10% 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 0,044. [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: $e_t^2 = \alpha_1 + \alpha_2 log(GDP_t) + \alpha_3 log(CPI_t) + \nu_t$
- Abhängige Variable: $e_t^2 = (y_t - x_t' b)^2$, sind die quadrierten Residuen aus der ursprünglichen Schätzung.
- Teststatistik: $\chi^2_{emp} = T \cdot R^2 = 80 \cdot 0{,}044 = 3{,}52$ (wobei T die Anzahl der Beobachtungen bezeichnet; R^2 ist das Bestimmtheitsmaß aus der Hilfsregression).
- Freiheitsgrade: J=2.
- Kritischer Wert: $\chi^2_{J;\alpha} = \chi^2_{2;0,10} = 4{,}61$.
- Da $\chi^2_{empirisch} = 3{,}52 < 4{,}61 = \chi^2_{kritisch}$ wird die Nullhypothese auf dem 10%-Niveau nicht verworfen. Der Test weist nicht auf Heteroskedastie in dem vorliegenden Modell hin.

3.4 Zeigen Sie für ein lineares Modell $y_t = x_t\beta + \varepsilon_t$ mit Autokorrelation zweiter Ordnung, also $\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + \nu_t$ mit $\nu_t \sim iid(0, \sigma^2)$, wie eine Spezifikationsänderung das Autokorrelationsproblem beseitigen kann. [4 Punkte]

- In das lineare Modell die mit ρ_1 und ρ_2 vormultipizierten verzögerten abhängigen Variablen und die verzögerten erklärenden Variablen mit aufnehmen:
- $y_t - \rho_1 y_{t-1} - \rho_2 y_{t-2} = \beta(x_t - \rho_1 x_{t-1} - \rho_2 x_{t-2}) + (\varepsilon_t - \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2})$ [3P]
- Da $\varepsilon_t - \rho_1\varepsilon_{t-1} - \rho_2\varepsilon_{t-2} = \nu_t \sim iid(0, \sigma^2)$ ist der neue Störterm nicht mehr autokorreliert.

Aufgabe 4: **[11,5 Punkte]**

4.1 Zeigen Sie in Matrizenschreibweise, dass der KQ-Schätzer $b = (X'X)^{-1}X'y$ für das Modell $y = X\beta + \varepsilon$ unverzerrt ist. Gehen Sie von einer deterministischen X Matrix aus und kennzeichnen Sie eindeutig, an welcher Stelle der Herleitung Sie welche Annahme verwenden. [3 Punkte]

i. $E[b] = E[(X'X)^{-1}X'y]$ [Aufgabenstellung]
ii. $E[b] = E[(X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon]$
iii. $E[b] = E[\beta + (X'X)^{-1}X'\varepsilon]$
iv. $E[b] = \beta + (X'X)^{-1}X'E[\varepsilon]$
v. $E[b] = \beta$

- Für Schritt v) wird A1 ($E[\varepsilon] = 0$) benötigt.

4.1.1 Was ändert sich in Ihrer Antwort zu Teilaufgabe 4.1) bei der Herleitung, wenn die Matrix X stochastisch ist? [2 Punkte]

- Bei stochastischen X wird von Schritt iii) auf iv) zusätzlich die Annahme A2 (Unabhängigkeit von X und ε) oder die A10 ($E[\varepsilon|X] = 0$) benötigt.

4.2 Für das Modell $y = X\beta + \varepsilon$ entspricht die Varianz des Kleinstquadrateschätzers $Var[b] = Var[(X'X)^{-1}X'\varepsilon]$. Gehen Sie von einer deterministischen $(X'X)$ Matrix aus. Leiten Sie die Varianz unter der Annahme von homoskedastischen, nicht-autokorrelierten Störtermen her. Kennzeichnen Sie die Verwendung der Annahmen. (4 Punkte)

i. $Var[b] = Var[(X'X)^{-1}X'\varepsilon]$ [Aufgabenstellung]
ii. $= (X'X)^{-1}X'Var[\varepsilon]X(X'X)^{-1}$
iii. $= (X'X)^{-1}X'(\sigma^2 I_N)X(X'X)^{-1}$
iv. $= \sigma^2(X'X)^{-1}X'X(X'X)^{-1}$
v. $= \sigma^2(X'X)^{-1}$

- Schritt iii): $Var[\varepsilon]$ kann durch $\sigma^2 I_N$ ersetzt werden, wenn Homoskedastie und keine Autokorrelation vorliegen.

4.3 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: $Var(\varepsilon_i) = \sigma^2$ (o.ä. Darstellung).
- 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.)

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.	Wenn zwei Zufallsvariablen in einem nicht-linearen Zusammenhang stehen, dann
a	kann dieser Zusammenhang korrekt mit der Kovarianz abgebildet werden.
b	kann die Kovarianz negativ sein.**X**
c	kann dieser Zusammenhang korrekt mit dem Korrelationskoeffizienten abgebildet werden.
d	ist die Kovarianz stets 0.

2.	Wenn für zwei Zufallsvariablen A und B gilt, dass $E(B\|A) = E(B) = 0$, dann
a	ist $Var(B\|A) = Var(B) = 0$.
b	ist $E(A\|B) = E(A) = 0$.
c	ist die Kovarianz von A und B gleich 0. **X**
d	sind die Variablen A und B statistisch unabhängig.

3.	Die Stärke (Power) eines Tests bezeichnet die Wahrscheinlichkeit,
a	eine wahre Nullhypothese abzulehnen.
b	eine falsche Nullhypothese abzulehnen. **X**
c	eine wahre Nullhypothese nicht abzulehnen.
d	eine falsche Nullhypothese nicht abzulehnen.

4.	Schätzgleichungen mit quadratischen erklärenden Variablen
a	erfordern nichtlineare Schätzverfahren.
b	können nicht für logarithmierte abhängige Variablen berechnet werden.
c	haben stets ein höheres R^2 als eine Schätzung ohne quadratische Terme.
d	können mit KQ berechnet werden. **X**

5.	Sie schätzen das Modell $ln(y_i) = \beta_1 + \beta_2 ln(x_i) + \varepsilon_i$ mit KQ und erhalten $b_2 = 1$. Ceteris paribus und im Mittel gilt: wenn x um...
a	1% steigt, so steigt y um 1%. **X**
b	1% steigt, so steigt y um 100 Einheiten.
c	eine Einheit steigt, so steigt y um 1%.
d	eine Einheit steigt, so steigt y um 1 Einheit.

6.	Sie schätzen folgendes Modell: $y_i = \beta_1^* + \beta_2^* x_i + \varepsilon_i^*$, obwohl das wahre Modell $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$ lautet. Welcher Ausdruck beschreibt den Erwartungswert von b_2^*?
a	$E[b_2^*] = \beta_1 + \beta_2 \frac{Cov(z,y)}{Var(x)}$.
b	$E[b_2^*] = \beta_2 + \beta_3 \frac{Cov(z,y)}{Var(x)}$.
c	$E[b_2^*] = \beta_2 + \beta_3 \frac{Cov(x,z)}{Var(x)}$. **X**
d	$E[b_2^*] = \beta_1 + \beta_2 \frac{Cov(x,z)}{Var(x)}$.

7.	Das korrigierte Bestimmtheitsmaß
a	enstpricht dem quadrierten Korrelationskoeffizienten der abhängigen Variablen und den vorhergesagten Werten.
b	berücksichtigt die zur Schätzung verwendeten Freiheitsgrade. **X**
c	ist genauso groß wie das unkorrigierte Bestimmtheitsmaß.
d	kann bei hohem Erklärungsgehalt des Modells größer als 1 sein.

8.	Die Annahme $\varepsilon_\mathbf{i} \sim (0, \sigma_i^2 I)$ mit $\sigma_i \neq \sigma_j$ für $i \neq j$, wobei I der Einheitsmatrix entspricht, impliziert dass
a	die Störterme autokorreliert und homoskedastisch sind.
b	die Störterme autokorreliert, aber nicht homoskedastisch sind.
c	die Störterme homoskedastisch und nicht autokorreliert sind.
d	die Störterme nicht homoskedastisch und nicht autokorreliert sind. **X**

9.	Wenn der Störterm einer Regression nicht normalverteilt ist, dann
a	ist die Annahme $E[\mathbf{x}\varepsilon] = 0$ verletzt.
b	sinkt die Effizienz der Schätzung.
c	minimiert das KQ-Verfahren nicht die Summe der quadrierten Residuen.
d	sind bei KQ-Schätzern die t- und F-Tests asymptotisch zutreffend. **X**

10.	Die Nullhypothese im White Test besagt,
a	dass Autokorrelation vorliegt.
b	dass Homoskedastie vorliegt.**X**
c	dass weder Autokorrelation noch Homoskedastie vorliegen.
d	dass Autokorrelation und Heteroskedastie vorliegen.

11.	Bei Autokorrelation in Form von moving average Störprozessen
a	sind alle Störterme miteinander korreliert.
b	gibt es Fehlerterme, die nicht miteinander korreliert sind.**X**
c	folgen die Störterme einem AR(1) Prozess.
d	steigt der Grad der Korrelation über die Zeit an.

12.	Autokorrelation im Störterm kann behoben werden durch
a	logarithmieren der abhängigen Variable.
b	die Aufnahme von relevanten erklärenden Variablen. **X**
c	eine Vergrößerung der Stichprobe.
d	Quadrierung der abhängigen Variable.

13.	Der Prais-Winsten Schätzer
a	gehört zur Klasse der BLUE Schätzer (best-linear-unbiased estimator).
b	kann bei Vorliegen von Heteroskedastie nicht bestimmt werden.
c	wird unter Verwendung einer Instrumentalvariable berechnet.
d	ist effizienter als der Cochrane-Orcutt Schätzer.**X**

14.	Der Durbin-Watson Test ist bei Schätzung mit Konstante und deterministischen Regressoren
a	nur durchführbar, wenn mindestens 15 Zeitperioden vorliegen.
b	sowohl asymptotisch als auch in kleinen Stichproben gültig. **X**
c	zum Testen auf Heteroskedastie geeignet.
d	bei positiver Autokorrelation nicht durchführbar.

15.	Das Informationskriterium AIC
a	wird verwendet, um zu überprüfen, ob die Log-Likelihood Funktion korrekt spezifiert ist.
b	hat weniger Informationsgehalt als das R^2.
c	bewertet die Effizienz einer Schätzung.
d	kann zum Vergleich genesteter Modelle verwendet werden. **X**

16.	In welchem der folgenden Schätzverfahren können die mittels Maximum Likelihood geschätzten Koeffizienten $\hat{\beta}$ als marginale Effekte interpretiert werden, wenn keine Interaktionsterme enthalten sind?
a	Probit.
b	Logit.
c	Im linearen Modell mit normalverteilten Störtermen. **X**
d	Im Probit und im Logit.

17.	Im Falle perfekter Kollinearität zweier erklärender Variablen im linearen Regressionsmodell
a	ist die Kreuzproduktmatrix $(X'X)$ singulär. **X**
b	liegt der Wert des Bestimmtheitsmaßes R^2 nahe 1.
c	können beide betroffenen Variablen aufgenommen werden, solange das Modell mit Konstante geschätzt wird.
d	konvergiert die Störtermvarianz gegen Unendlich.

18.	Welche der folgenden Aussagen bezüglich der Maximum-Likelihood-Methode ist korrekt?
a	Die Methode maximiert die Wahrscheinlichkeit, die Residuenquadratsumme zu minimieren.
b	Die Methode benötigt keine Annahme über die Verteilung der Störterme.
c	Eine Annahme bezüglich der Verteilung der zu schätzenden Parameter ist erforderlich.
d	Die Methode kann zur Schätzung des linearen Regressionsmodells verwendet werden. **X**

19.	Im Modell $y_i = \beta_1 + \beta_2 x_{i1} + \beta_3 x_{i2} + \varepsilon_i$ instrumentieren Sie die endogene Variable x_1 nur mit sich selbst und schätzen mittels 2SLS. Die Folge ist
a	Äquivalenz der KQ- und 2SLS-Koeffizientenschätzer. **X**
b	ein Problem schwacher Instrumente.
c	perfekte Kollinearität auf der zweiten Stufe der 2SLS-Schätzung.
d	ein erwartungstreuer Schätzer für β_2.

20.	Ein Anzeichen für schwache Instrumente liegt vor wenn
a	der IV-Schätzer $\hat{\beta}_{IV}$ kleiner als der KQ-Schätzer $\hat{\beta}_{KQ}$ ist.
b	ein niedriger Wert des Bestimmtheitsmaßes R^2 der zweiten Stufe der 2SLS-Schätzung auftritt.
c	der Wert der F-Statistik eines Tests auf gemeinsame Signifikanz der Instrumente auf der ersten Stufe der 2SLS-Schätzung geringer als 10 ist. **X**
d	eine hohe Korrelation zwischen den Instrumenten und der endogenen Variable besteht.

21.	Welche Aussage bezüglich des Modells $ln(wage_i) = \beta_1 + \beta_2 educ_i + \beta_3 exper_i + \beta_4 exper_i^2 + \varepsilon_i$ ist korrekt?
a	Der Effekt der Variable $exper_i$ auf $wage_i$ ist positiv, wenn $\beta_4 = 0$.
b	Der Effekt der Variable $exper_i$ auf $wage_i$ folgt einem konkaven Verlauf, wenn $\beta_3 + \beta_4 < 0$.
c	Der Effekt der Variable $exper_i$ auf $wage_i$ folgt einem konvexen Verlauf, wenn $\beta_3 + \beta_4 > 1$.
d	Der Effekt der Variable $exper_i$ auf $wage_i$ folgt einem konkaven Verlauf, wenn $\beta_4 < 0$. **X**

22.	Welche Aussage ist korrekt? Das Logit Verfahren
a	wird im Falle einer logarithmierten abhängigen Variable angewandt.
b	trifft die gleiche Annahme bezüglich der Verteilung der Störterme wie das Probit-Modell.
c	unterstellt eine logistisch verteilte Dichtefunktion im Modell mit einer binären abhängigen Variable. **X**
d	impliziert konstante marginale Effekte.

23.	Eine Matrix **A** ist idempotent, wenn gilt **AA=A**. Auf welche der folgenden Matrizen trifft diese Eigenschaft <u>nicht</u> zu?
a	$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. **X**
b	$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.
c	$A = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$.
d	$A = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$.

24.	Die Varianz-Kovarianzmatrix der Störterme eines Regressionsmodells sei $\sigma^2\Omega$, wobei $\Omega = \begin{pmatrix} 1 & 0,2 \\ 0,2 & 1,2 \end{pmatrix}$. Worauf können Sie schließen?
a	Homoskedastie und positive Autokorrelation
b	Heteroskedastie und positive Autokorrelation **X**
c	Homoskedastie und keine Autokorrelation
d	Heteroskedastie und keine Autokorrelation

25.	Sie schätzen folgendes Regressionsmodell $y_i = \beta_1 + \beta_2 x_{i1} + \beta_3 x_{i2} + \beta_4 x_{i2}^3 + \varepsilon_i$ und erhalten $\hat{\beta}_1 = 0,5$, $\hat{\beta}_2 = -2,2$, $\hat{\beta}_3 = -1,5$, $\hat{\beta}_4 = 0,5$. Bei welchem Wert beträgt der marginale Effekt der Variable x_{i2} null?
a	x_{i2} = 1 **X**
b	x_{i2} = 2
c	x_{i2} = 0,5
d	x_{i2} = -1,5

26.	Eine Verletzung der Annahme $E[\varepsilon_i x_{ij}] = 0$
a	kann mit dem AIC-Wert überprüft werden.
b	hat keine Konsequenz im Falle einer Maximum-Likelihood-Schätzung.
c	kann durch Erhöhung der Stichprobengröße verhindert werden.
d	kann Folge eines Messfehlers in der erklärenden Variablen sein. **X**

27.	Der marginale Effekt einer binären erklärenden Variable x_{ij}
a	kann als ceteris paribus Mittelwertsdifferenz interpretiert werden. **X**
b	beträgt $\frac{1}{n}\sum_i^N x_{ij}$.
c	am Mittel der Daten entspricht im Logit-Modell stets dem mittleren marginalen Effekt.
d	liegt immer im Intervall (0,1).

28.	Der F-Test auf gemeinsame Signifikanz von J Koeffizienten
a	testet gegen die Alternativhypothese $H_1 : \beta_1 = \beta_2 = ... = \beta_J \neq 0$.
b	lehnt die Nullhypothese nur ab, wenn mindestens $\frac{J}{2}$ Koeffizienten nicht am 10%-Niveau signifikant sind.
c	berücksichtigt die mögliche Korrelation zwischen den J Parameterschätzern. **X**
d	basiert auf dem Vergleich des angepassten R^2 in einem restringierten und einem unrestringierten Modell.

29.	Die lineare Umskalierung einer erklärenden Variablen
a	hat keinen Einfluss auf die Schätzgüte eines Modells. **X**
b	ändert den p-Wert, nicht jedoch den Standardfehler des Schätzkoeffizienten.
c	ermöglicht die Schätzung im Falle von perfekter Kollinearität.
d	kann Heteroskedastie verringern oder erhöhen.

30.	In welchem Fall ist die kausale Interpretation der Schätzkoeffizienten einer KQ-Schätzung <u>nicht</u> gerechtfertigt?
a	Im linearen Wahrscheinlichkeitsmodell.
b	Bei Vorliegen eines omitted variable bias. **X**
c	Bei der Verwendung robuster Standardfehler.
d	Wenn Kollinearität besteht.