Aufgabe 1: **[17 Punkte]**

Sie interessieren sich für den Zusammenhang zwischen Einkommen und Konsum. Ihr Datensatz enthält monatliche Informationen über einen Zeitraum von 12 Jahren zu folgenden Variablen:

C_t **realer Pro Kopf Konsum in Deutschland in Periode t (in €)**

Y_t **reales Pro Kopf Einkommen in Deutschland in Periode t (in €)**

i_t **Dummy-Variable (= 1 für Monate in denen der Zinssatz für Spareinlagen größer war als die Inflationsrate; sonst = 0)**

Sie schätzen das Modell $C_t = \beta_{Kons} + \beta_Y \cdot Y_t + \beta_i \cdot i_t + \varepsilon_t$ und erhalten folgenden Regressionsoutput:

```
Call:
lm(formula = C ~ Y + i)

Coefficients:
            Estimate   Std. Error t value Pr(>|t|)
(Intercept) 320.58834    219.731   1.459   0.147
Y             0.81230    ..???..   4.466   2.81e-05
i           -49.32590     26.764   1.843   0.068
---
Residual standard error: 50.692 on ??? degrees of freedom
Multiple R-Squared: 0.715,      Adjusted R-squared: ???
```

a) **Berechnen Sie unter Angabe des Rechenwegs:** **(4 Punkte)**

a1) den Standardfehler zu b_Y

$= b_Y / t_{b_Y}$ = 0,81230/4,466 = 0,18188

a2) ein 95% Konfidenzintervall für b_i

= [b_i ± 1,96*se(b_i)] = [-49,326 ± 1,96*26,764] = [-101,783 ; 3,132]

a3) das angepasste R^2.

$$= 1 - \frac{(1-R^2)(N-1)}{(N-K)} = 1 - \frac{(1-0{,}715)(144-1)}{(144-3)} = 0{,}71095 \approx 0{,}711$$

b) **Interpretieren Sie den Koeffizienten b_Y inhaltlich und statistisch.** **(2 Punkte)**

steigt das Pro Kopf Einkommen um 1 Euro so steigt der Pro Kopf Konsum im Erwartungswert c.p. um 81,2 Cent

der Parameter ist am 1%-Niveau statistisch signifikant von Null verschieden

c) **Wie lauten der Punkt- und der Intervallschätzer (α = 5%) für C_t bei einem Pro-Kopf-Einkommen von 3.000 Euro, einem Zinssatz von 2,5 Prozent und einer Inflationsrate von 1 Prozent? Die geschätzte Varianz des Vorhersagefehlers an diesem Punkt beträgt: $\widehat{\mathrm{Var}}(C_0 - \widehat{C_0}) = 2809{,}3$.** **(4 Punkte)**

Vorhersage: $\hat{C}_0$ = 320,58834+0,81230*3000 – 49,3259 = 2708,158 Euro

Konfidenzintervall:

$=\widehat{C_0} \pm 1{,}96 \cdot \widehat{se}(C_0 - \widehat{C_0}) = 2708{,}158 \pm 1{,}96 \cdot \sqrt{2809{,}3} = [2604{,}278; 2812{,}083]$

d) **Sie vermuten, dass der marginale Effekt des Einkommens Y_t auf die Konsumausgaben C_t niedriger ist, wenn der Zinssatz für Spareinlagen die Inflationsrate übersteigt.**

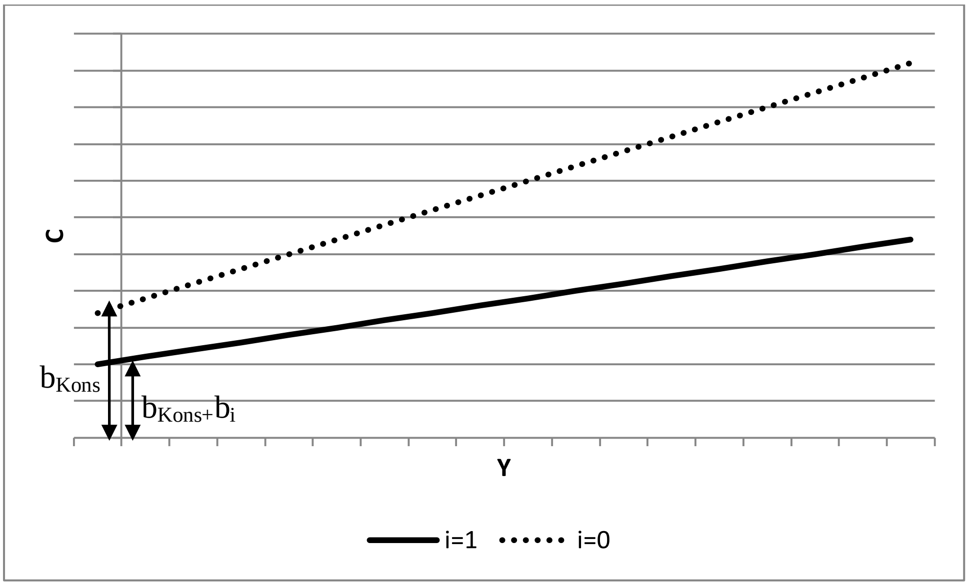
d1) Wie können Sie diese Vermutung in *einer* Schätzgleichung überprüfen? Schreiben Sie das zu schätzende Modell auf und erläutern Sie genau, wie Sie die Ergebnisse in Bezug auf Ihre Vermutung interpretieren würden. (3 Punkte)

es kann ein Interaktionsterm $i_t \cdot Y_t$ eingefügt werden

das Modell lautet dann $C_t = \widetilde{\beta_{Kons}} + \widetilde{\beta_Y} \cdot Y_t + \widetilde{\beta_\iota} \cdot i_t + \gamma \cdot Y_t \cdot i_t + u_t$

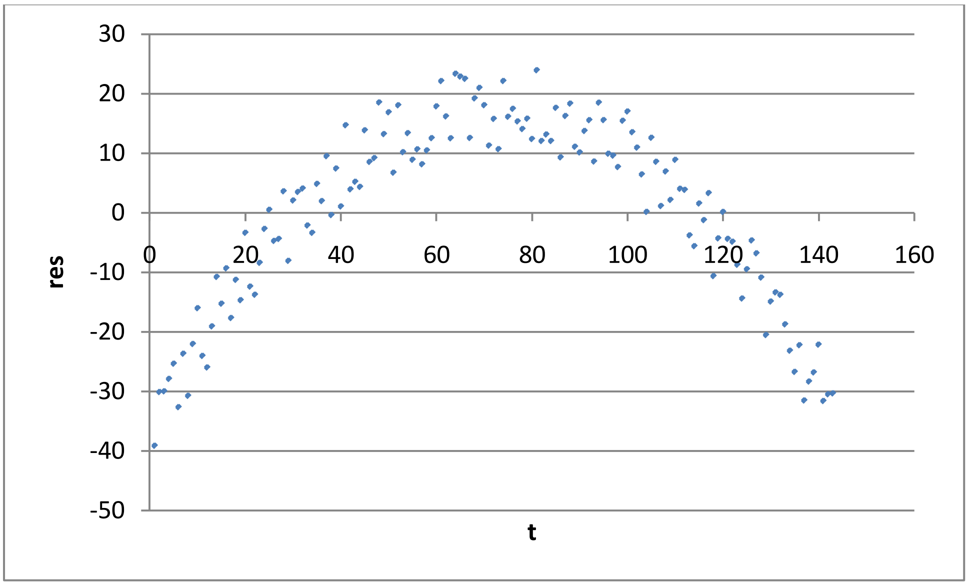
ist $\hat{\gamma}$ <0, bestätigt dies die Vermutung, ist $\hat{\gamma}$ > 0 spricht das gegen die Vermutung

d2) Unterstellen Sie, dass Ihre Vermutung zutrifft. Stellen Sie den Zusammenhang zwischen C_t und Y_t für beide möglichen Ausprägungen von i_t graphisch dar. (4 Punkte)



Aufgabe 2: [11 Punkte]

a) **Sie stellen das Residuum aus der Regression in Aufgabe 1a) im Zeitablauf dar und erhalten folgende Grafik:**



Unterstellen Sie hier eine konstante Störtermvarianz. Würde die Teststatistik eines Durbin Watson Tests einen Wert größer, kleiner oder gleich 2 ergeben? Begründen Sie kurz. (2 Punkte)

die Grafik zeigt positive Autokorrelation der Residuen.

die DW-Testatistik ist bei positiver Autokorrelation 1. Ordnung kleiner als 2.

b) Auf Vorschlag eines Kommilitonen hin schätzen Sie das Modell $\hat{\varepsilon}_t^2 = \alpha_0 + \alpha_1 \cdot Y_t + \vartheta_t$, wobei $\hat{\varepsilon}_t$ das geschätzte Residuum des Modells aus Aufgabe 1a) ist. Für ϑ_t gilt: $\vartheta_t \sim i.i.d.(0, \sigma^2)$. Sie schätzen für α_i den Wert 0,078 mit einem Standardfehler von 0,012. Auf welches Problem deutet dieser Wert hin? Was bedeutet das für die Schätzung Ihres Modells? (4 Punkte)

deutet auf das Vorliegen von Heteroskedastie hin

dies führt zu falsch berechneten Standardfehlern, und daher zu falschen t- und F-Statistiken, sowie zu einer ineffizienten Schätzung

c) Wie können Sie bei Vorliegen eines Autokorrelationsprozesses 1. Ordnung einen BLUE Schätzer erzeugen? Beschreiben Sie formal Ihre Vorgehensweise und erläutern Sie kurz deren Auswirkung auf den Störterm. (Ignorieren Sie die erste Periode.) (5 Punkte)

AR(1): $\varepsilon_t = \rho \cdot \varepsilon_{t-1} + u_t$ mit $u_t \sim i.i.d.(0, \sigma^2)$

GLS Transformation

1) $C_t = \beta_{Kons} + \beta_Y \cdot Y_t + \beta_i \cdot i_t + \varepsilon_t$

2) $\rho C_{t-1} = \rho\beta_{Kons} + \beta_Y \cdot \rho Y_{t-1} + \beta_i \cdot \rho i_{t-1} + \rho\varepsilon_{t-1}$

3)=1)-2):

$$C_t - \rho C_{t-1} = (1-\rho)\beta_{Kons} + \beta_Y \cdot (Y_t - \rho Y_{t-1}) + \beta_i \cdot (i_t - \rho i_{t-1}) + (\varepsilon_t - \rho\varepsilon_{t-1}) \Leftrightarrow$$

$$C_t^* = \beta_{Kons}^* + \beta_Y^* \cdot Y_t^* + \beta_i^* \cdot i_t^* + \varepsilon_t^*$$

durch die Transformation ist der Störterm in Gleichung 3) $\varepsilon_t^* = u_t \sim i.i.d.(0, \sigma^2)$ und es liegt daher keine Autokorrelation mehr vor

Aufgabe 3: [19 Punkte]

Sie untersuchen den Zusammenhang zwischen der medizinischen Versorgung in Botswana und der Zahl der Kinder, die dort das erste Lebensjahr überleben. Ihnen liegt ein Datensatz mit Informationen über 4361 Frauen vor. Der Datensatz enthält unter anderem folgende Variablen:

children	**Anzahl der Kinder einer Frau, die das erste Lebensjahr überlebt haben**
age	**Alter der Frau in Jahren**
age²	**Alter der Frau in Jahren, quadriert**
med	**Dummyvariable (=1, wenn die Frau medizinische Versorgung während und nach ihren Schwangerschaften in Anspruch genommen hat, =0 sonst)**

Um die Wirksamkeit der medizinischen Versorgung zu untersuchen, schätzen Sie zunächst folgendes Modell:

$$children_i = \beta_0 + \beta_1 \cdot age_i + \beta_2 \cdot age_{it}^2 + \beta_3 \cdot med_i + \varepsilon_i.$$

Die Auswertung der Stichprobe ergibt folgenden Output:

```
Call:
lm(formula =children ~ age + age² + prenat)

Coefficients:
            Estimate Std.Err. t value  Pr(>|t|)
(Intercept) -4.865   0.2388   -20.37   < 2e-16
age          0.350   0.0168    20.77   < 2e-16
age²        -0.003   0.0003    -9.85   < 2e-16
med         -0.383   0.045     -8.47   < 2e-16
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Residual standard error: 2.2089 on 4357 degrees of freedom
Multiple R-Squared: 0.5529,    Adjusted R-squared: 0.5526
F-statistic: 1796.16 on 3 and 4357 DF,  p-value: < 2.2e-16
```

a) Unterstellen Sie, dass die Gauss-Markov-Annahmen erfüllt sind und interpretieren Sie den Koeffizienten β_3 inhaltlich und statistisch. (2 Punkte)

Inanspruchnahme medizinischer Versorgung während der Schwangerschaft senkt die Zahl der Kinder, die das erste Lebensjahr überleben c.p. im Erwartungswert um 0,38. Der Koeffizient ist statistisch signifikant von Null verschieden auf dem 1%-Niveau.

b) Ein Kollege behauptet, dass die Variable *med* endogen sei.

b1) Welche Annahme des linearen Modells wäre verletzt? Nennen Sie die Annahme formal. (1 Punkt)

A7: $E(\varepsilon \cdot med) = 0$ ist nicht erfüllt.

b2) Welche Konsequenzen hätte die Verletzung der Annahme für die Qualität Ihres Schätzers? (2 Punkte)

verzerrt und inkonsistent

b3) Erläutern Sie verbal, was die Annahme bedeutet, und unter welchen Umständen sie allgemein verletzt wäre. Nennen Sie ein plausibles Beispiel, wodurch ihre Verletzung im vorliegenden Fall verursacht werden könnte. (3 Punkte)

Die Annahme bedeutet, dass der Störterm nicht mit der erklärenden Variable korreliert sein darf.

Sie wird verletzt, wenn eine unbeobachtete Variable gemeinsam die Zahl der Kinder wie auch die Inanspruchnahme medizinischer Versorgung determiniert.

Hier könnte dies zum Beispiel Einkommen oder Bildung der Frau sein.

c) „Ärzte ohne Grenzen" hat in manche Regionen Botswanas Hebammen entsandt, deren Dienste kostenlos in Anspruch genommen werden können. Ihr Datensatz enthält eine weitere Variable *prog*. Es handelt sich um eine Dummy-Variable, die den Wert 1 annimmt, wenn die Frau in einer Region lebt, in der das Programm durchgeführt wurde, und 0 sonst.

c1) Wie könnten Sie diese Variable nutzen, um β_3 zu schätzen? Beschreiben Sie verbal ein mögliches zweistufiges Verfahren. (3 Punkte)

prog als Instrument einsetzen:

med auf prog regressieren

vorhergesagte Werte med_hat ermitteln

Ursprungsregression mit med_hat anstelle von med durchführen.

c2) Nennen Sie die zwei formalen Annahmen, die *prog* erfüllen müsste, damit Ihr Vorgehen zu einer konsistenten Schätzung führt. Halten sie die Annahmen für erfüllt? Begründen Sie Ihre Antwort am Beispiel: (4 Punkte)

prog muss korreliert sein mit med → vermutlich erfüllt, da die Wahrscheinlichkeit, medizinisch versorgt zu werden steigen sollte, wenn man nicht dafür bezahlen muss.

prog muss unkorreliert sein mit dem Störterm ε → wäre verletzt, wenn das Projekt z.B. gezielt in Regionen mit besonders kinderreichen Haushalten eingesetzt worden wäre, oder irgendetwas anderes Plausibles. („Kein ersichtlicher Grund von Verletzung auszugehen" kann auch akzeptiert werden.)

d) Nehmen Sie an, Ihnen lägen Paneldaten vor.

d1) Stellen Sie für das Beispiel die Schätzgleichung für den within-Schätzer dar. Unterstellen Sie hierfür anstelle des quadratischen einen linearen Einfluss des Alters. (2 Punkte)

$$\text{children}_{it} - \overline{\text{children}_i} = \beta_1 \cdot (\text{age}_{it} - \overline{\text{age}_i}) + \beta_3 \cdot (\text{med}_{it} - \overline{\text{med}_i}) + (\alpha_i - \overline{\alpha_i}) + \varepsilon_{it}$$

oder

$$\text{children}_{it} - \overline{\text{children}_i} = \beta_1 \cdot (\text{age}_{it} - \overline{\text{age}_i}) + \beta_3 \cdot (\text{med}_{it} - \overline{\text{med}_i}) + 0 + \varepsilon_{it}$$

d2) Wäre dieses Schätzmodell geeignet, um das von Ihnen in b3) beschriebene Problem zu lösen? Erläutern Sie (gegebenenfalls an Ihrem Beispiel aus b3)), unter welchen Umständen das Vorgehen geeignet wäre oder nicht. (2 Punkte)

Wenn die in b3) genannte Variable zeitkonstant ist, lässt sich das Endogenitätsproblem durch die within-Transformation lösen, da α herausgerechnet wird. Ist die in b3) genannte Variable zeitvariabel, geht dies nicht.

Aufgabe 4 [13 Punkte]

Betrachten Sie erneut den Datensatz aus Aufgabe 3. Ihnen stehen weitere Variablen zur Verfügung, die die sozioökonomische Situation der Frauen beschreiben:

tv **Dummy-Variable (=1, wenn im Haushalt ein Fernseher vorhanden ist)**
bicycle **Dummy-Variable (=1, wenn die Frau ein Fahrrad besitzt)**
elec **Dummy-Variable (=1, wenn die Wohnung mit Elektrizität ausgestattet ist)**

Weiterhin liegen Ihnen nun Paneldaten zu 4 Zeitpunkten vor. Sie schätzen das Modell:

$$children_{it} = \beta_0 + \beta_1 \cdot age_{it} + \beta_2 \cdot age_{it}^2 + \beta_3 \cdot med_{it} + \beta_4 \cdot tv_{it} + \beta_5 \cdot bicycle_{it} + \beta_6 \cdot elec_{it} + \mu_{it}$$

mit $\mu_{it} = \alpha_i + \varepsilon_{it}$ einmal als gepooltes KQ-Modell und einmal als Random-Effects-Modell.

a) **Welche formale Annahme bezüglich α_i muss sowohl im Random-Effects-Modells als auch im gepoolten KQ-Modell erfüllt sein, damit die Schätzer konsistent sind? (1 Punkt)**

Beide Modelle unterstellen, dass α und die erklärenden Variablen unkorreliert sind.

Gehen Sie im Folgenden davon aus, dass die notwendigen Annahmen erfüllt sind, damit beide Methoden konsistente Schätzer darstellen.

b) **In Paneldaten liegt sowohl sogenannte „between-Variation“ als auch „within-Variation“ vor.**

b1) Was versteht man unter beiden Begriffen? Erläutern sie verbal. (4 Punkte)

within-Variation ist die Abweichung einer Variable vom individuenspezifischen Mittelwert für diese Variable in jeder Periode. Sie beschreibt, wie sich die Ausprägungen einer Variable für ein Individuum über die Zeit hinweg unterscheiden.

between-Variation ist die Abweichung des individuenspezifischen Mittelwerts einer Variable vom Mittelwert dieser Variable für die gesamte Stichprobe. Sie beschreibt, wie sich die Ausprägungen einer Variable über die Individuen hinweg unterscheiden.

b2) Welcher Unterschied und welche Gemeinsamkeit in der Nutzung von between- und within-Variation charakterisieren das Random-Effects- und das KQ-Verfahren? (2 Punkte)

Gemeinsamkeit: Beide Modelle nutzen sowohl within-, als auch between-Variation.

Unterschied: Das RE-Modell gewichtet between- und within-Variation effizient, während sie in die KQ-Regression immer gleich gewichtet eingehen.

c) **Sie erhalten folgende Ergebnisse für Ihre Stichprobe mit N=4316 und T=4:**

	pooled		RE	
	Koeffizient	t-Wert	Koeffizient	t-Wert
constant	-0.382	-0.152	-0.396	-0.204
age	0.497	1.520	0.455	1.827
age^2	-0.006	-1.500	-0.003	-3.583
med	0.287	1.577	0.266	2.486
tv	-0.666	-1.908	-0.403	-2.343
bicycle	-0.088	-0.481	-0.091	-1.492

elec	-0.564	-2.968	-0.400	-3.419

c1) Interpretieren Sie inhaltlich den Koeffizienten zu *med* im KQ-Modell. (1 Punkt)

med: nimmt eine Frau während und nach der Schwangerschaft medizinische Versorgung in Anspruch, ist die erwartete Zahl der Kinder, die das erste Lebensjahr überleben, c.p. um 0,287 höher als für Frauen, die das nicht tun.

c2) Sind die Koeffizienten aus c1) im Random-Effects-Modell und im gepoolten KQ-Modell signifikant auf dem 5%-Niveau? Begründen Sie. (3 Punkte)

kritischer Wert für 5%-Niveau ist 1,96 da N=4316 und T=4

Im KQ sind die t-Werte zu β_3 und β_4 jeweils <1,96, die Koeffizienten sind also nicht statistisch signifikant auf 5%-Niveau.

Im RE-Modell sind die t-Werte zu β_3 und β_4 jeweils >1,96, die Koeffizienten sind also statistisch signifikant auf 5%-Niveau.

c3) Welche Ursache vermuten Sie für die in c2) gefundenen Signifikanzmuster? Begründen Sie kurz. (2 Punkte)

Die Standardfehler im RE-Modell sind niedriger, da es das effizientere Verfahren ist. Deshalb sind die t-Werte größer.

Aufgabe 5: [45 Punkte]

Wahr oder falsch? Tragen Sie für jede der folgenden Aussagen ein „w“ für „wahr“ oder ein „f“ für „falsch“ ein. Für jede richtige Antwort gibt es 0.75 Punkte, für jede falsche Antwort werden 0.75 Punkte abgezogen. Die Gesamtpunktzahl kann nicht negativ werden.

W	Für Matrizen gilt allgemein: (A')' = A
W	Die Nullhypothese H_0: $\beta \geq c$ wird bei 1500 Freiheitsgraden am 5 Prozentniveau verworfen, wenn als Teststatistik der t-Wert kleiner als – 1,645 ist.
F	Berücksichtigt man im Modell irrelevante erklärende Variablen, so sinkt die Streuung der geschätzten Parameter.
W	Das Auslassen einer relevanten erklärenden Variablen kann zu verzerrten Schätzern führen.
F	GMM-Schätzer des linearen Modells sind inkonsistent bei Autokorrelation.
W	GMM Verfahren benötigen keine Verteilungsannahmen.
F	Ein Parameterschätzer ist effizient, wenn er gegen seinen Erwartungswert konvergiert.
F	Bei Heteroskedastie ist der KQ-Schätzer verzerrt.
W	Ein Vektor ist eine Reihe oder Spalte von Zahlen.
F	Bei Paneldaten lassen sich Probleme ausgelassener zeitkonstanter Variablen durch den Prais-Winsten Schätzer lösen.
W	Ein niedriger AIC-Wert weist auf ein besseres Regressionsmodell hin als ein hoher.
F	Wenn es mehr Instrumente als zu schätzende Parameter gibt, ist das Modell überidentifiziert und kann daher nicht geschätzt werden.
F	Mit steigender Zahl von Freiheitsgraden konvergiert die t-Verteilung zur F-Verteilung.
W	Um k Parameter zu identifizieren, benötigt man mindestens k Momentenbedingungen.
W	Die wahre Regressionskonstante hat eine Varianz von Null.
F	Wenn der p-Wert größer ist als das Signifikanzniveau eines Tests, wird die Nullhypothese verworfen.
F	Beim Random-Effects Modell enthält der Störterm zwei Komponenten, die beide über die Dimensionen i und t variieren.

W	Die Annahme $\varepsilon_i \sim i.i.d.(0, \sigma^2)$ schließt sowohl Heteroskedastie als auch Autokorrelation aus.
W	Der White-Test ist allgemeiner als der Test auf Gleichheit zweier Varianzen (Goldfeldt-Quandt).
F	An der Parametermitte ist ein KQ-Schätzer überidentifiziert.
W	Das BIC Kriterium zeigt eine umso bessere Schätzgüte, je kleiner die Fehlerquadratsumme bei gegebener Parameter- und Beobachtungszahl ist.
W	AR(1) Störterme sind homoskedastisch.
W	Eine Matrix P ist symmetrisch wenn P = P'.
F	Ein Typ II Fehler ist unwahrscheinlicher als ein Typ I Fehler.
W	Bei Autokorrelation sind die mit den KQ-Schätzern ausgewiesenen p-Werte ungültig.
W	Der "PE Test" verwendet die t-Teststatistik.
W	Die Gleichheit zweier Varianzen kann mittels eines F-Tests überprüft werden.
W	Die χ^2-Verteilung ist eine einparametrige Verteilung.
F	Bei Querschnittsdaten bilden heteroskedastische Störterme die Schocks vergangener Perioden ab.
F	Auf Basis linearer Modelle geschätzte Koeffizienten sind kausale Effekte.
F	Mit Hilfe des Goldfeld-Quandt Tests lassen sich fixed- und random effects-Schätzer gegeneinander testen.
W	Die optimale Gewichtungsmatrix des GMM Modells W ist eine Funktion der Varianz-Kovarianzmatrix der Koeffizienten.
W	Wenn X deterministisch ist, gilt $E\{\varepsilon \mid X\} = E\{\varepsilon\}$.
F	Im Gegensatz zum KQ-Schätzer optimieren GMM (generalized method of moments) Schätzer keine Zielfunktion.
W	Der marginale Effekt einer erklärenden Variablen kann je nach Spezifikation des Modells unterschiedlich ausfallen.
F	Auch bei exakter Multikollinearität kann der Kleinstquadrateschätzer unverzerrt geschätzt werden.
W	Bei negativer Autokorrelation zweiter Ordnung ist der Durbin-Watson Test nicht gültig.
F	Auf Basis linearer Modelle geschätzte Koeffizienten können immer als Elastizität interpretiert werden.
F	Zur Unverzerrtheit des fixed effects-Schätzers ist keine Aussage möglich.
F	Der Typ II Fehler beschreibt eine Situation in der eine nicht zutreffende Nullhypothese verworfen wird.
W	Die Teststatistik des White-Tests ist asymptotisch Chi-quadrat verteilt.
W	Newey-West Standardfehler korrigieren sowohl für Heteroskedastie unbekannten Ursprungs als auch für Autokorrelationsmuster, die auf H Perioden beschränkt sind.
F	Die χ^2-Verteilung ist eine symmetrische Verteilung
F	Die Inkonsistenz eines Steigungsparameters führt im linearen Modell nicht zur Inkonsistenz der gleichzeitig geschätzten Regressionskonstanten.
F	Die Nullhypothese $\beta \geq c$ wird am 5 Prozent Signifikanzniveau bei 500 Freiheitsgraden verworfen, wenn der empirische t-Wert größer als 1,645 ist.
W	Wird die Nullhypothese des Sargan-Tests verworfen, so treffen nicht alle Momentenbedingungen zu.
W	Der between-Schätzer ist konsistent, wenn der individuenspezifische Störterm nicht mit den individuenspezifischen Mittelwerten der erklärenden Variablen korreliert ist.
F	Bei endogenen erklärenden Variablen ist eine GMM Schätzung dann effizient, wenn die Anzahl der Momentenbedingungen im KQ-Schätzer kleiner ist, als die Anzahl der zu schätzenden Parameter (K).
W	Bei strikter Exogenität der erklärenden Variablen ist der fixed effects-Schätzer konsistent.
W	Der RESET Test nutzt Potenzen der vorhergesagten Werte von y, um ein Modell auf Fehlspezifikation zu überprüfen.

W	Autokorrelation führt nicht zu Inkonsistenz des KQ-Schätzers.
W	Das angepasste R^2 einer Schätzung kann bei Hinzufügen von erklärenden Variablen sinken.
W	Wenn man eine konkrete Form der Heteroskedastie unterstellt, kann man FGLS-Schätzer verwenden.
W	Der Durbin-Wu-Hausman Test auf Endogenität einer erklärenden Variablen wird durchgeführt, indem der Regressionsgleichung eine zusätzliche erklärende Variable hinzugefügt wird.
W	Bei Autokorrelation in Form von moving average Störprozessen gibt es Fehlerterme, die nicht miteinander korreliert sind.
F	Bei Messfehlern in der abhängigen Variable sind die Parameterschätzer auf Null hin verzerrt.
W	In einem Modell mit Konstante beschreibt das R^2 den Anteil der erklärten Variation der abhängigen Variable an der gesamten Variation der abhängigen Variable.
W	Unter einem Interaktionsterm versteht man das Produkt zweier erklärender Variablen.
F	Die Dichtefunktion der t-Verteilung hat ihr Maximum bei 1,96.
F	Der Chow-Test überprüft mittels einer F Teststatistik, ob vorhergesagte Werte der abhängigen Variable den Erklärungsgehalt des Modells erhöhen.

Aufgabe 6: **[15 Punkte]**

Wahr oder falsch? Begründen Sie Ihre Auffassung (Bsp.: "Stimmt, weil..." bzw. "Stimmt nicht, weil..."). Nur bei korrekter Begründung erhält jede richtige Antwort 1.5 Punkte; Angaben ohne Begründung werden nicht gewertet.

W	Unter den Gauss-Markov-Annahmen sind KQ Schätzer nicht exakt normalverteilt. → Aus den Gauss-Markov-Annahmen lässt sich nur asymptotische Normalverteilung ableiten.
W	Bei Vorliegen von Heteroskedastie gilt das Gauss-Markov-Theorem nicht. → Bei Heteroskedastie ist der KQ-Schätzer nicht mehr BLUE, da ein GLS-Schätzer in diesem Fall effizienter wäre.
W	Ob "schwache Instrumente" vorliegen, lässt sich durch eine Hilfsregression überprüfen. → Die endogene Variable wird auf die exogenen Variablen und das Instrument regressiert. Ist der geschätzte Parameter des Instruments signifikant von Null verschieden, ist das Instrument nicht schwach.
F	Alle Fragestellungen der Paneldatenanalyse können auch mit Querschnittsdatenanalysen beantwortet werden. → Mit Querschnittsanalysen können z.B. keine Aussagen zu zeitlichen Veränderungen innerhalb derselben Beobachtungseinheiten getroffen werden.
F	Die Varianz des geschätzten Achsenabschnittsparameters muss 0 sein. → Schätzwerte für einen Parameter sind immer Realisationen von Zufallsvariablen und haben daher eine Varianz >0.
W	Der verallgemeinerte Kleinstquadrateschätzer kann zu Effizienzsteigerung beitragen. → So kann z.B. die GLS Transformationen bei Autokorrelation oder Heteroskedastie zu Effizienzsteigerung beitragen.

W	Damit sich eine Variable Z als Instrumentvariable eignet, müssen zwei Bedingungen erfüllt sein. → Das Instrument muss mit endogener Variable korreliert sein. → Das Instrument muss mit dem Störterm unkorreliert sein.
F	Bei Vorliegen von Heteroskedastie sollte ein fixed effects Schätzer verwendet werden. → Ein Fixed Effects Schätzer ist keine Lösung bei Heteroskedastie. Ein verallgemeinerter Kleinstquadrateschätzer (GLS) kann eine Lösung sein.
W	Nutzt man das KQ Verfahren, um Elastizitäten zu schätzen, muss die abhängige Variable logarithmiert sein. → Schätzergebnisse mittels KQ lassen sich nur als prozentuale Änderungen der abhängigen Variable interpretieren wenn diese logarithmiert ist (gleichzeitig muss auch die erklärende Variable logarithmiert sein; ansonsten nur Semi-Elastizität).
W	Der Wert des R^2 kann nicht größer sein als 1. → R^2 misst den Anteil der durch das Modell erklärten Streuung an der Gesamtstreuung der abhängigen Variable. Da maximal die gesamte Streuung erklärt werden kann ist der maximale Anteil 1.