

Diplomprüfung im Fach Ökonometrie im SS 2009 - Aufgabenteil

Name, Vorname	
Matrikelnr.	
Studiengang	
Semester	
Datum	
Raum, Sitzplatz-Nr.	
Unterschrift	
Email	

Vorbemerkungen:

Anzahl der Aufgaben:

Die Klausur besteht aus 6 Aufgaben, von denen alle bearbeitet werden müssen.

Bewertung:

Die Prüfung dauert 120 Minuten, es können maximal 120 Punkte erworben werden. Die Punktzahl ist für jede Aufgabe in Klammern angegeben.

Erlaubte Hilfsmittel:

- **1 DIN A4-Blatt** mit Notizen (Vorder- und Rückseite, also max. 2 DIN A4-Seiten)
- Tabellen der statistischen Verteilungen (sind der Klausur beige-fügt)
- Taschenrechner
- Fremdwörterbuch

Wichtige Hinweise:

- Sollte es vorkommen, dass die statistischen Tabellen, die dieser Klausur beiliegen, den exakten Wert der Freiheitsgrade nicht ausweisen, machen Sie dies kenntlich und verwenden Sie den nächstgelegenen Wert.
- Sollte es vorkommen, dass bei einer Berechnung eine erforderliche Annahme oder Angabe fehlt, machen Sie dies kenntlich und treffen Sie für den fehlenden Wert eine plausible Annahme.
- Die Aufgaben 5 und 6 sind im Aufgabenteil zu beantworten, die restlichen Aufgaben im Lösungsteil. Verwenden Sie hierbei für jede Aufgabe ein neues Blatt.

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

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

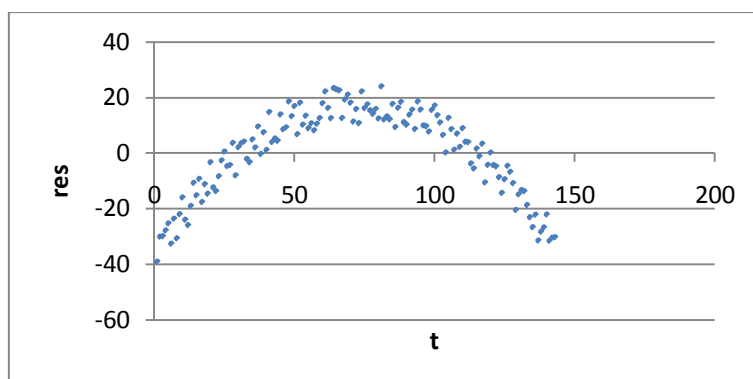
Coefficients:
            Estimate   Std. Error t value Pr(>|t|)
(Intercept) 320.58834    219.731    1.459  0.147
Y             0.81230     ..???..    4.466 2.81e-05
i            -49.32590     26.764    1.843  0.068
---
Residual standard error: 50.692 on ??? degrees of freedom
Multiple R-Squared:  0.715,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

- a) Berechnen Sie unter Angabe des Rechenwegs: (4 Punkte)
 - a1) den Standardfehler zu b_Y
 - a2) ein 95% Konfidenzintervall für b_i
 - a3) das angepasste R^2 .
- b) Interpretieren Sie den Koeffizienten b_Y inhaltlich und statistisch. (2 Punkte)
- c) Wie lauten der Punkt- und der Intervallschätzer ($\alpha = 5\%$) 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: $\widehat{\text{Var}}(C_0 - \widehat{C}_0) = 2809,3$. (4 Punkte)
- 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. (7 Punkte)
 - 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.
 - 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.

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)

- 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_1 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)
- 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)

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:

<i>children</i>	Anzahl der Kinder einer Frau, die das erste Lebensjahr überlebt haben
<i>age</i>	Alter der Frau in Jahren
<i>age²</i>	Alter der Frau in Jahren, quadriert
<i>med</i>	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 + age^2 + 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
age^2        -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)
- b) Ein Kollege behauptet, dass die Variable *med* endogen sei. (6 Punkte)
- b1) Welche Annahme des linearen Modells wäre verletzt? Nennen Sie die Annahme formal.
- b2) Welche Konsequenzen hätte die Verletzung der Annahme für die Qualität Ihres Schätzers?
- 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.
- 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. (7 Punkte)
- c1) Wie könnten Sie diese Variable nutzen, um β_3 zu schätzen? Beschreiben Sie verbal ein mögliches zweistufiges Verfahren.

- 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.
- d) Nehmen Sie an, Ihnen lägen Paneldaten vor. (4 Punkte)
- d1) Stellen Sie für das Beispiel die Schätzgleichung für den within-Schätzer dar. Unterstellen Sie hierfür anstelle des quadratischen einen linearen Einfluss des Alters.
- 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.

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)

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. (6 Punkte)

- b1) Was versteht man unter beiden Begriffen? Erläutern sie verbal.
- b2) Welcher Unterschied und welche Gemeinsamkeit in der Nutzung von between- und within-Variation charakterisieren das Random-Effects- und das KQ-Verfahren?

- c) Sie erhalten folgende Ergebnisse für Ihre Stichprobe mit N=4316 und T=4: (6 Punkte)

	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.
- c2) Sind die Koeffizienten aus c1) im Random-Effects-Modell und im gepoolten KQ-Modell signifikant auf dem 5%-Niveau? Begründen Sie.
- c3) Welche Ursache vermuten Sie für die in c2) gefundenen Signifikanzmuster? Begründen Sie kurz.

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.

	Für Matrizen gilt allgemein: $(A')' = A$
	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.
	Berücksichtigt man im Modell irrelevante erklärende Variablen, so sinkt die Streuung der geschätzten Parameter.
	Das Auslassen einer relevanten erklärenden Variablen kann zu verzerrten Schätzern führen.
	GMM-Schätzer des linearen Modells sind inkonsistent bei Autokorrelation.
	GMM Verfahren benötigen keine Verteilungsannahmen.
	Ein Parameterschätzer ist effizient, wenn er gegen seinen Erwartungswert konvergiert.
	Bei Heteroskedastie ist der KQ-Schätzer verzerrt.
	Ein Vektor ist eine Reihe oder Spalte von Zahlen.
	Bei Paneldaten lassen sich Probleme ausgelassener zeitkonstanter Variablen durch den Prais-Winsten Schätzer lösen.
	Ein niedriger AIC-Wert weist auf ein besseres Regressionsmodell hin als ein hoher.
	Wenn es mehr Instrumente als zu schätzende Parameter gibt, ist das Modell überidentifiziert und kann daher nicht geschätzt werden.
	Mit steigender Zahl von Freiheitsgraden konvergiert die t-Verteilung zur F-Verteilung.
	Um k Parameter zu identifizieren, benötigt man mindestens k Momentenbedingungen.
	Die wahre Regressionskonstante hat eine Varianz von Null.
	Wenn der p-Wert größer ist als das Signifikanzniveau eines Tests, wird die Nullhypothese verworfen.
	Beim Random-Effects Modell enthält der Störterm zwei Komponenten, die beide über die Dimensionen i und t variieren.
	Die Annahme $\varepsilon_i \sim i. i. d. (0, \sigma^2)$ schließt sowohl Heteroskedastie als auch Autokorrelation aus.
	Der White-Test ist allgemeiner als der Test auf Gleichheit zweier Varianzen (Goldfeldt-Quandt).
	An der Parametermitte ist ein KQ-Schätzer überidentifiziert.
	Das BIC Kriterium zeigt eine umso bessere Schätzgüte, je kleiner die Fehlerquadratsumme bei gegebener Parameter- und Beobachtungszahl ist.
	AR(1) Störterme sind homoskedastisch.
	Eine Matrix P ist symmetrisch wenn $P = P'$.
	Ein Typ II Fehler ist unwahrscheinlicher als ein Typ I Fehler.
	Bei Autokorrelation sind die mit den KQ-Schätzern ausgewiesenen p-Werte ungültig.
	Der "PE Test" verwendet die t-Teststatistik.
	Die Gleichheit zweier Varianzen kann mittels eines F-Tests überprüft werden.
	Die χ^2 -Verteilung ist eine einparametrische Verteilung.
	Bei Querschnittsdaten bilden heteroskedastische Störterme die Schocks vergangener Perioden ab.
	Auf Basis linearer Modelle geschätzte Koeffizienten sind kausale Effekte.
	Mit Hilfe des Goldfeld-Quandt Tests lassen sich fixed- und random effects-Schätzer gegeneinander testen.
	Die optimale Gewichtungsmatrix des GMM Modells W ist eine Funktion der Varianz-Kovarianzmatrix der Koeffizienten.
	Wenn X deterministisch ist, gilt $E\{\varepsilon X\} = E\{\varepsilon\}$.

	Im Gegensatz zum KQ-Schätzer optimieren GMM (generalized method of moments) Schätzer keine Zielfunktion.
	Der marginale Effekt einer erklärenden Variablen kann je nach Spezifikation des Modells unterschiedlich ausfallen.
	Auch bei exakter Multikollinearität kann der Kleinstquadrateschätzer unverzerrt geschätzt werden.
	Bei negativer Autokorrelation zweiter Ordnung ist der Durbin-Watson Test nicht gültig.
	Auf Basis linearer Modelle geschätzte Koeffizienten können immer als Elastizität interpretiert werden.
	Zur Unverzerrtheit des fixed effects-Schätzers ist keine Aussage möglich.
	Der Typ II Fehler beschreibt eine Situation in der eine nicht zutreffende Nullhypothese verworfen wird.
	Die Teststatistik des White-Tests ist asymptotisch Chi-quadrat verteilt.
	Newey-West Standardfehler korrigieren sowohl für Heteroskedastie unbekanntes Ursprungs als auch für Autokorrelationsmuster, die auf H Perioden beschränkt sind.
	Die χ^2 -Verteilung ist eine symmetrische Verteilung
	Die Inkonsistenz eines Steigungsparameters führt im linearen Modell nicht zur Inkonsistenz der gleichzeitig geschätzten Regressionskonstanten.
	Die Nullhypothese $\beta \geq c$ wird am 5 Prozent Signifikanzniveau bei 500 Freiheitsgraden verworfen, wenn der empirische t-Wert größer als 1,645 ist.
	Wird die Nullhypothese des Sargan-Tests verworfen, so treffen nicht alle Momentenbedingungen zu.
	Der between-Schätzer ist konsistent, wenn der individuenspezifische Störterm nicht mit den individuenspezifischen Mittelwerten der erklärenden Variablen korreliert ist.
	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).
	Bei strikter Exogenität der erklärenden Variablen ist der fixed effects-Schätzer konsistent.
	Der RESET Test nutzt Potenzen der vorhergesagten Werte von y, um ein Modell auf Fehlspezifikation zu überprüfen.
	Autokorrelation führt nicht zu Inkonsistenz des KQ-Schätzers.
	Das angepasste R^2 einer Schätzung kann bei Hinzufügen von erklärenden Variablen sinken.
	Wenn man eine konkrete Form der Heteroskedastie unterstellt, kann man FGLS-Schätzer verwenden.
	Der Durbin-Wu-Hausman Test auf Endogenität einer erklärenden Variablen wird durchgeführt, indem der Regressionsgleichung eine zusätzliche erklärende Variable hinzugefügt wird.
	Bei Autokorrelation in Form von moving average Störprozessen gibt es Fehlerterme, die nicht miteinander korreliert sind.
	Bei Messfehlern in der abhängigen Variable sind die Parameterschätzer auf Null hin verzerrt.
	In einem Modell mit Konstante beschreibt das R^2 den Anteil der erklärten Variation der abhängigen Variable an der gesamten Variation der abhängigen Variable.
	Unter einem Interaktionsterm versteht man das Produkt zweier erklärender Variablen.
	Die Dichtefunktion der t-Verteilung hat ihr Maximum bei 1,96.
	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.

	Unter den Gauss-Markov-Annahmen sind KQ Schätzer nicht exakt normalverteilt.
	Bei Vorliegen von Heteroskedastie gilt das Gauss-Markov-Theorem nicht.
	Ob "schwache Instrumente" vorliegen, lässt sich durch eine Hilfsregression überprüfen.
	Alle Fragestellungen der Paneldatenanalyse können auch mit Querschnittsdatenanalysen beantwortet werden.
	Die Varianz des geschätzten Achsenabschnittsparameters muss 0 sein.
	Der verallgemeinerte Kleinstquadrateschätzer kann zu Effizienzsteigerung beitragen.
	Damit sich eine Variable Z als Instrumentvariable eignet, müssen zwei Bedingungen erfüllt sein.
	Bei Vorliegen von Heteroskedastie sollte ein fixed effects Schätzer verwendet werden.
	Nutzt man das KQ Verfahren, um Elastizitäten zu schätzen, muss die abhängige Variable logarithmiert sein.
	Der Wert des R^2 kann nicht größer sein als 1.