

---

# Diplomprüfung im Fach Ökonometrie im SS 07 - Aufgabenteil

Name, Vorname	
Matrikelnr.	
Studiengang	
Semester	
Datum	23.07.2007
Raum, Sitzplatz-Nr.	
Unterschrift	

## 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:**

- 2 DIN A4-Blätter mit Notizen (Vorder- und Rückseite, also max. 4 DIN A4-Seiten)
- Tabellen der statistischen Verteilungen (sind der Klausur beigefügt)
- 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 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:****[25 Punkte]**

Sie interessieren sich für die Determinanten der schulischen Bildung. Sie vermuten, dass die schulische Ausbildungszeit abhängt von der Bildung der Eltern und der Anzahl der Geschwister, die eine Person hat. Sie unterstellen folgendes Modell (Modell A):

$$educ_i = \beta_0 + \beta_1 \cdot sibs_i + \beta_2 \cdot meduc_i + \beta_3 \cdot feduc_i + \beta_4 \cdot age_i + \beta_5 \cdot age_i^2 + \beta_6 \cdot urban_i + \varepsilon_i$$

- educ<sub>i</sub>*: schulische Ausbildungszeit der *i*-ten Person (in Jahren)  
*sibs<sub>i</sub>*: Anzahl der Geschwister der *i*-ten Person  
*meduc<sub>i</sub>*: schulische Ausbildungszeit der Mutter der *i*-ten Person (in Jahren)  
*feduc<sub>i</sub>*: schulische Ausbildungszeit des Vaters der *i*-ten Person (in Jahren)  
*age<sub>i</sub>*: Alter in Jahren  
*urban<sub>i</sub>*: Dummy-Variable (=1, wenn Person *i* in städtischen Regionen aufgewachsen ist)

Die Auswertung einer Stichprobe von 722 erwerbstätigen Männern im Alter von 28 bis 37 Jahren ergibt folgenden Output:

```
Call:
lm(formula = educ ~ sibs + meduc + feduc + age + I(age^2) + urban)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -19.594847   9.917491  -1.976  0.04856 *
sibs         -0.084677   0.034378  -2.463  0.01401 *
meduc        0.135033    ???         4.151 3.71e-05 ***
feduc        ???        0.027542    7.605 8.98e-14 ***
age          1.769902    0.600087    2.949  0.00329 **
I(age^2)     -0.026054    0.009032   -2.885  0.00404 **
urban        0.156258    0.165193    0.946  0.34451
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.974 on 715 degrees of freedom
Multiple R-Squared:  ???, Adjusted R-squared:  0.2209
F-statistic: 35.06 on 6 and 715 DF, p-value: < 2.2e-16
```

- a) Berechnen Sie unter Angabe des Rechenwegs (5 Punkte)
- a1) den Standardfehler für  $b_2$  ;
  - a2) den geschätzten Koeffizienten für  $\beta_3$  ;
  - a3) ein 95%-Konfidenzintervall für  $b_6$  ;
  - a4)  $R^2$ .
- b) Interpretieren Sie die Koeffizienten  $b_1$  und  $b_2$  inhaltlich und statistisch. (2 Punkte)
- c) Berechnen Sie den marginalen Effekt des Alters am Stichprobenmittel (= 32,9 Jahre) und interpretieren Sie diesen inhaltlich. (3 Punkte)
- d) Um mögliche Kohorteneffekte angemessener abbilden zu können, werden statt *age* und *age*<sup>2</sup> Dummy-Variablen (in 2er-Jahresschritten: *age1* – Alter zwischen 28 und unter 30; *age2* – Alter zwischen 30 und unter 32, etc.) in das Modell aufgenommen (Modell B): (5 Punkte)

```

Call:
lm(formula = educ ~ sibs + meduc + feduc + age1 + age2 + age3 + age4 + urban)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.156069   0.385269   26.361 < 2e-16 ***
sibs         -0.085994   0.034442   -2.497  0.0128 *
meduc        0.136387   0.032563    4.188 3.16e-05 ***
feduc        0.211012   0.027579    7.651 6.47e-14 ***
age1        -0.565485   0.237234   -2.384  0.0174 *
age2         0.009061   0.205658    0.044  0.9649
age3         0.291775   0.217727    1.340  0.1806
age4         0.125481   0.245191    0.512  0.6090
urban        0.172182   0.166497    1.034  0.3014
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.975 on 713 degrees of freedom
Multiple R-Squared:  0.2286,    Adjusted R-squared:  0.22
F-statistic: 26.41 on 8 and 713 DF,  p-value: < 2.2e-16

```

d1) Interpretieren Sie den Koeffizienten der Variable *age1* inhaltlich.

d2) Welches der beiden Modelle ist zu bevorzugen? Ziehen Sie hierzu die AIC-Werte der beiden Modelle heran; die Summe der quadrierten geschätzten Residuen beträgt 2786.123 in Modell A und 2781.146 in Modell B.

e) Sie vermuten, dass die Fehlertermvarianz mit steigendem Alter variiert:  $\text{var}(\varepsilon_i) = \sigma_i^2 = \sigma^2 \cdot \text{age}_i^2$ . (10 Punkte)

e1) Wie kann man Modell (A) so transformieren, dass die Heteroskedastie eliminiert wird? Stellen Sie zudem formal die Auswirkung der Transformation auf die Varianz des Störterms dar.

e2) Beschreiben Sie kurz die Vorgehensweise des Breusch-Pagan Tests und testen Sie mit den folgenden Angaben am 5% Niveau auf Heteroskedastie. Geben Sie hierzu auch den kritischen Wert an.

```

Studentized Breusch-Pagan test

data:  mod.A
BP = 20.1235, df = 6

```

e3) Welchen allgemeineren Test kann man als Alternative zum Breusch-Pagan Test heranziehen? Skizzieren Sie kurz einen Vor- und einen Nachteil.

**Aufgabe 2:** **[10 Punkte]**

Sie untersuchen, ob die Wirtschaftsleistung in den USA Auswirkung hat auf die Erwerbsbeteiligung in Puerto Rico. Ihnen liegen hierfür Daten des Zeitraums zwischen 1965 bis einschließlich 2004 zur Verfügung. Sie schätzen folgendes Modell:

$$\log(\text{share}_t) = \beta_0 + \beta_1 \cdot \log(\text{usgnp}_t) + \beta_2 \cdot \log(\text{prgnp}_t) + \varepsilon_t$$

- mit
- share<sub>t</sub>*: Anteil der erwerbstätigen Personen an der Bevölkerung zum Zeitpunkt *t*
  - usgnp<sub>t</sub>*: Bruttosozialprodukt in den USA zum Zeitpunkt *t* (in Mrd. 1985-US-Dollar)
  - prgnp<sub>t</sub>*: Bruttosozialprodukt in Puerto Rico zum Zeitpunkt *t* (in Mrd. 1985-US-Dollar)

- a) Wie berechnen Sie die Vorhersage für die Erwerbsbeteiligung, wenn Ihnen Schätzergebnisse vorliegen? (2 Punkte)
- b) Der geschätzte Autokorrelationskoeffizient betrage 0.3495. Berechnen Sie die Teststatistik des Durbin-Watson Tests und testen Sie auf dem 5% Niveau, ob positive Autokorrelation 1. Ordnung vorliegt. Geben Sie hierzu auch die Nullhypothese, die Alternativhypothese sowie die Freiheitsgrade und die jeweiligen kritischen Werte an. (5 Punkte)

- c) Wie ändern Sie das Testprozedere, wenn zusätzlich die Variable  $y_{t-1}$  im Modell berücksichtigt wird? Erläutern Sie Ihre Vorgehensweise. (3 Punkte)

### Aufgabe 3:

[11 Punkte]

Sie untersuchen die Renditen der Bildung. Auf Grundlage der Humankapitaltheorie schätzen Sie hierzu für 428 erwerbstätige Frauen folgende Verdienstfunktion:

$$lwage_i = \beta_0 + \beta_1 \cdot educ_i + \beta_2 \cdot exper_i + \beta_3 \cdot exper_i^2 + \varepsilon_i,$$

wobei

- lwage<sub>i</sub>*: logarithmierter Verdienst (in €) von Person *i*  
*educ<sub>i</sub>*: Jahre der (schulischen) Bildung von Person *i*  
*exper<sub>i</sub>*: betriebspezifische Erfahrung von Person *i* (in Jahren)  
*fatheduc<sub>i</sub>*: Jahre der (schulischen) Bildung des Vaters von Person *i*  
*motheduc<sub>i</sub>*: Jahre der (schulischen) Bildung der Mutter von Person *i*

In der nachfolgenden Tabelle sind die Ergebnisse einer KQ-Schätzung (1), einer Hilfsregression (2), einer IV-Schätzung (3) sowie einer KQ-Schätzung einer reduzierten Gleichung (4) dargelegt:

Abh. Variable	(1)	(2)	(3)	(4)
	lwage	lwage	lwage	educ
constant	-0.522*** (0.20)	0.0743 (0.41)	0.0481 (0.40)	9.480*** (0.32)
educ	0.107*** (0.014)	0.0609* (0.031)	0.0614* (0.031)	—
exper	0.0416*** (0.013)	0.0415*** (0.013)	0.0442*** (0.013)	—
expersq	-0.000811** (0.00039)	-0.000840** (0.00039)	-0.000899** (0.00040)	—
v	—	0.0586* (0.035)	—	—
fatheduc	—	—	—	0.188*** (0.034)
motheduc	—	—	—	0.156*** (0.036)
R <sup>2</sup>	0.16	0.16	—	0.21

Standardfehler in Klammern

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

- a) Sie vermuten, dass *educ* eine endogene Variable ist. Welche Annahmeverletzung liegt dem Endogenitätsproblem zugrunde? Was wäre die Konsequenz? (2 Punkte)
- b) Welcher Test liegt der Hilfsregression in Spalte (2) zugrunde? Beschreiben Sie diesen kurz. Wie ist das Testergebnis hier? (4 Punkte)
- c) Die IV-Schätzung in Spalte (3) zieht *fatheduc* und *motheduc* als Instrumente heran. Ist dies eine gute Wahl? Nennen Sie die allgemeinen Bedingungen und interpretieren Sie diesbezüglich die Ergebnisse der reduzierten Gleichung in Spalte (4). (3 Punkte)
- d) Was versteht man unter Überidentifikation? (2 Punkte)

### Aufgabe 4:

[14 Punkte]

Sie interessieren sich für die Determinanten der an einer großen US-amerikanischen Universität erbrachten Leistungen. Ihnen stehen für 728 Studierende für zwei Semester Daten zu folgenden Variablen zur Verfügung:

- gpa*: Durchschnittsnoten-Index (zwischen 0-mangelhaft und 4-sehr gut) von StudentIn *i*  
*black*: =1, wenn StudentIn *i* Afro-AmerikanerIn ist  
*female*: =1, wenn StudentIn *i* eine Frau ist  
*hssize*: Größe der High School-Abschlussklasse von StudentIn *i*  
*football*: =1, wenn StudentIn *i* FootballspielerIn in der Universitätsmannschaft ist

Bei Anwendung unterschiedlicher Verfahren ergeben sich folgende Schätzergebnisse:

Abhängige Variable: gpa				
	gepooltes KQ	between	within	random effects
Constant	2.407*** (0.058)	2.405*** (0.073)	2.694*** (0.37)	2.416*** (0.072)
black	-0.513*** (0.065)	-0.527*** (0.081)	–	-0.515*** (0.081)
female	0.433*** (0.064)	0.435*** (0.080)	–	0.424*** (0.080)
hssize	-0.0354 (0.13)	-0.0375 (0.16)	0.120 (0.10)	-0.0337 (0.16)
football	-0.210*** (0.061)	-0.199** (0.077)	-1.252*** (0.46)	-0.228*** (0.076)

Anmerkungen: Standardfehler in Klammern; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

- a) Erläutern Sie ausführlich, wieso bei der within Schätzung keine Koeffizienten für *black* und *female* ausgewiesen werden. (3 Punkte)
- b) Wieso können sich die Ergebnisse der within und random effects Schätzung erheblich unterscheiden? Führen Sie zwei mögliche Begründungen an. (3 Punkte)
- c) Zeigen Sie, dass der gepoolte KQ-Schätzer  $b = \sum (x_{it} x'_{it})^{-1} x_{it} y_{it}$  verzerrt ist, wenn das ‚wahre‘ Modell  $y_{it} = x'_{it} \beta + \alpha_i + \varepsilon_{it}$  sei, mit  $\alpha_i$  als nicht-beobachtbarer, nicht-stochastischer zeitkonstanter Größe. (4 Punkte)
- d) Der EGLS-Schätzer des random effects Modells kann bestimmt werden über eine KQ-Schätzung des folgenden transformierten Modells: (4 Punkte)

$$(y_{it} - \rho \bar{y}_i) = \mu(1 - \rho) + (x'_{it} - \rho \bar{x}_i)' \beta + u_{it}$$

- d1) Der Parameter  $\rho$  liegt zwischen 0 und 1. Welche Schätzer resultieren, wenn  $\rho = 0$  bzw.  $\rho = 1$ ?
- d2) Berechnen Sie  $\rho$ , gegeben dass  $\hat{\sigma}_\alpha^2 = 0.3510$  und  $\hat{\sigma}_\varepsilon^2 = 0.2095$ . Hierbei sind  $\alpha_i$  und  $\varepsilon_{it}$  die Fehlertermkomponenten des zugrunde liegenden random effects Modells.

**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.

	Zu den Gauss-Markov-Annahmen gehört, dass der Erwartungswert der Regressionskonstante Null ist.
	Marginale Effekte beschreiben den Zusammenhang zwischen der abhängigen und einer erklärenden Variablen, gegeben der Wert aller anderen im Modell berücksichtigten erklärenden Variablen.
	Bei Heteroskedastie können KQ-Schätzer nicht mehr unverzerrt sein.
	Die Annahme $\varepsilon_i \sim i.i.d.(0, \sigma^2)$ schließt sowohl Heteroskedastie als auch Autokorrelation aus.
	Der "Encompassing Test" wird genutzt, um alternative Spezifikationen der abhängigen Variablen gegeneinander zu testen.
	Beim random effects Modell enthält der Störterm zwei Komponenten, wovon eine ein reiner Zufallsstörterm ist.
	Die Streuung der Steigungsparameter im Kleinstquadrateschätzer variiert mit der Anzahl der Stichprobenelemente.
	Der KQ Schätzer ist nur dann konsistent, wenn erklärende Variablen und Störterm unkorreliert sind.

	Bei falschen Standardfehlern sind Schlussfolgerungen auf Basis von $t$ -Tests ungültig.
	Das "nonlinear least squares" Verfahren erlaubt es, solche Spezifikationen zu schätzen, die nicht-linear in den Parametern sind.
	Der Wert des $R^2$ kann nicht kleiner als Null sein.
	Auf Basis linearer Modelle geschätzte Koeffizienten können immer als Kausaleffekte interpretiert werden.
	Der within Schätzer verwendet als Regressoren die beobachtungsspezifischen (i,t) Abweichungen vom Mittelwert der Variable pro Beobachtungseinheit (i).
	Heteroskedastieprobleme können mit einem verallgemeinerten KQ-Schätzer gelöst werden.
	Unter einem Interaktionsterm versteht man das Produkt zweier erklärender Variablen.
	Messfehler in den erklärenden Variablen führen zu inkonsistenten KQ-Schätzern.
	Wird ein log-lineares Modell geschätzt, so spielt bei der Vorhersage der nicht-logarithmierten abhängigen Variable die Annahme an die Verteilung des Störterms eine Rolle.
	Die Werte der $t$ -Verteilung sind symmetrisch um den Wert 1.96 verteilt.
	Beim einfachen $t$ -Test werden Hypothesen zu Eigenschaften des Störterms überprüft.
	Charakteristisch für Heteroskedastie ist, dass die Störterme nicht zufällig um den Wert Null schwanken.
	$t$ -Tests werden nicht verwendet, um Hypothesen, die mehrere Koeffizienten gleichzeitig betreffen, zu überprüfen.
	Wenn der Wahrscheinlichkeitsgrenzwert eines Schätzers dem wahren Wert entspricht, nennt man den Schätzer effizient.
	Im einfachen Regressionsmodell führt ein Messfehler in der erklärenden Variable dazu, dass der mit KQ geschätzte Steigungsparameter gegen Null hin verzerrt ist.
	AIC und BIC Kriterium bewerten die Güte von Schätzmodellen auf Basis der nicht erklärten Variation, der Anzahl der Parameter sowie der Anzahl der Beobachtungen.
	Der Typ II Fehler beschreibt eine Situation in der eine nicht zutreffende Nullhypothese verworfen wird.
	Homoskedastische Störterme haben eine Varianz von 1.
	Die Gleichheit zweier Varianzen kann mittels eines F-Tests überprüft werden.
	Werden erklärende Variablen endogen aus dem Modellrahmen heraus bestimmt, so führt dies zu inkonsistenten KQ-Schätzern.
	Der GLS Schätzer im random effects Modell stellt einen Mittelwert zwischen random effects und fixed effects Schätzer dar.
	Die Teststatistik des Breusch-Pagan Tests ist asymptotisch Chi-quadrat verteilt.
	Die Transponierte eine Matrix ist ebenfalls eine Matrix.
	Der (Durbin-Wu-) Hausman Test bietet eine Möglichkeit, die Endogenität einer erklärenden Variablen zu überprüfen.
	Der "PE Test" wird genutzt, um nicht genestete Modelle zu vergleichen.

	Mit steigender Zahl von Freiheitsgraden konvergiert die t-Verteilung zur F-Verteilung.
	Die Teststatistik des White-Tests ist asymptotisch Chi-quadrat verteilt.
	Liegen in einem Modell statt einer drei endogene Variable vor, so reicht eine zusätzliche Instrumentvariable aus, um das Problem zu lösen.
	Wenn zwei oder mehr im Zeitreihensinn aufeinander folgende Störterme korreliert sind, sprechen wir von Autokorrelation.
	Random effects Schätzer sind inkonsistent, wenn der beobachtungsspezifische Teil des Störterms (der random effect) mit der abhängigen Variable korreliert ist.
	Besonders präzise Vorhersagen erhält man, wenn diese für solche Werte der erklärenden Variablen erfolgen, die in der Nähe des Mittelwertes der beobachteten Ausprägungen liegen.
	Bei Autokorrelation erster Ordnung wird der Wert eines Störterms unmittelbar vom Wert des Störterms in den beiden vorausgehenden Perioden beeinflusst.
	Die Normalgleichungen im einfachen Regressionsmodell sind zwei Momentenbedingungen.
	Die Standardnormalverteilung hat eine Standardabweichung von Null.
	Bei stationären Störtermprozessen sind die Störterme mit wachsendem zeitlichen Abstand zunehmend schwächer korreliert.
	Ein Modell mit hohem Wert des BIC-Kriteriums passt besser zu den Daten als eins mit niedrigem BIC-Wert.
	Das Gauss-Markov-Theorem beschreibt die Eigenschaften des KQ-Schätzers unter bestimmten Annahmen.
	Ein schwaches Instrument liegt dann vor, wenn die Instrumentvariable eine Varianz unter 10 hat.
	Die Summe quadrierter, standardnormalverteilter Zufallsvariablen ist Chi-quadrat verteilt.
	Der Durbin-Watson Test auf Autokorrelation ist nur anwendbar, wenn das Modell mindestens zwei Steigungsparameter enthält.
	Die Breite des Vorhersageintervalls für die abhängige Variable variiert mit der Streuung des Störterms.
	Bei überidentifizierten GMM Modellen wird eine Gewichtungsmatrix benutzt.
	Eine Normalgleichung beschreibt die Zielfunktion des KQ-Schätzers.
	Newey-West Standardfehler basieren auf einer modifizierten Autokorrelationsannahme.
	Liegt keine Autokorrelation vor, so ergibt sich für die Durbin-Watson Teststatistik ein Wert von 2.
	Zeitreihenmodelle beschreiben Beziehungen zwischen zu verschiedenen Zeitpunkten gemessenen Werten einzelner Größen.
	GMM Verfahren benötigen keine Verteilungsannahmen.
	GIVE Schätzer können in Situationen genutzt werden, in denen für die Schätzer keine analytischen Lösungen vorliegen.
	Wenn der $p$ -Wert größer als das Signifikanzniveau $\alpha$ ist, wird $H_0$ verworfen.
	Der kritische Wert der Durbin-Watson Teststatistik variiert mit der Anzahl der Regressoren im Modell.

	Quadratische Matrizen haben keine Inverse.
	Bei nichtlinearen Funktionen $g$ gilt für eine Zufallsvariable $Y$ , dass der Erwartungswert für $g(Y)$ gleich ist dem Funktionswert von $E(Y)$ , also $g(E(Y))$ .

**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**.

	Eine $t$ -verteilte Zufallsvariable kann auf Basis zweier Zufallsvariablen mit unterschiedlichen Verteilungen berechnet werden.
	Die Annahme statistischer Unabhängigkeit zwischen zwei Zufallsvariablen $X$ und $Y$ ist allgemeiner als die Annahme, dass der auf $Y$ bedingte Erwartungswert von $X$ identisch ist mit dem unbedingten Erwartungswert von $X$ .
	Mithilfe des linearen Regressionsmodells können Elastizitäten geschätzt werden.
	Unter Ausnutzung von Werten zur Summe quadrierter Residuen lassen sich Hypothesen überprüfen.
	Solange bei autokorrelierten Störtermen keine Heteroskedastie vorliegt, sind KQ Schätzer effizient.
	Bei autokorrelierten Störtermen führen verzögerte erklärende Variablen in der Modellspezifikation zu verzerrten Schätzern.
	Wenn die Störterme des Regressionsmodells alle gleich Null sind, so beträgt der $R^2$ Wert 1.



	Damit sich eine Variable $Z$ als Instrumentvariable eignet, müssen zwei Bedingungen erfüllt sein.
	Fixed effects und random effects Schätzer führen im Normalfall zu sehr ähnlichen Schätzergebnissen.
	Heteroskedastie und Autokorrelation sind verwandte Phänomene.