

## Diplomprüfung im Fach Ökonometrie im WS 05/06 - Aufgabenteil

<b>Name, Vorname</b>	
<b>Matrikelnr.</b>	
<b>Studiengang</b>	
<b>Semester</b>	
<b>Datum</b>	<b>06.03.2006</b>
<b>Raum, Sitzplatz-Nr.</b>	<b>0.224/225,</b>
<b>Unterschrift</b>	

### Vorbemerkungen:

**Anzahl der Aufgaben:** Die Klausur besteht aus 7 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 beigelegt)
- 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.
- Aufgabe 5 und 6 sind im Aufgabenteil zu beantworten, die restlichen Aufgaben im Lösungsteil. Verwenden Sie hierbei für jede Aufgabe ein neues Blatt.

**Aufgabe 1:****[24 Punkte]**

In der Mitte der 1970er versuchte die US-Regierung, Kfz-Treibstoffe anhand von u.a. Verbrauchswerten von Vorperioden und soziodemographischen Charakteristika auf 50 US-Bundesstaaten zu allozieren. Ein diesem Vorhaben implizit zugrunde liegendes ökonometrisches Modell könnte dabei wie folgt gelaundet haben:

$$PCON_i = \beta_0 + \beta_1 \cdot REG_i + \beta_2 \cdot TAX_i + e_i,$$

wobei

- $PCON_i$ : Mineralöl-Konsum im i-ten Bundesstaat (in Mill. Energieeinheiten)  
 $REG_i$ : Kfz-Zulassungen im i-ten Bundesstaat (in Tsd.)  
 $TAX_i$ : Mineralöl-Steuer im i-ten Bundesstaat (US-Cents pro Gallone Treibstoff)  
 $POP_i$ : Bevölkerung im i-ten Bundesstaat (in Tsd.)

Mit Daten aus dem Jahr 1985 wurde das Modell in R geschätzt:

```
Call:
lm(formula = PCON ~ REG + TAX)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 551.68799   186.27095      ?     0.00479 **
REG           0.18613      ?      15.883 < 2e-16 ***
TAX           ?         16.85588   -3.179  0.00261 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 253 on 47 degrees of freedom
Multiple R-Squared:  0.8664,    Adjusted R-squared:  0.8607
F-statistic: 152.4 on 2 and 47 DF,  p-value: < 2.2e-16
```

- a) Berechnen Sie unter Angabe des Rechenwegs die im Output fehlenden Werte für (2,5 Punkte)
- den t-Wert für  $b_0$ ;
  - den Standardfehler für  $b_1$ ;
  - den Koeffizienten für  $b_2$ ;
  - die geschätzte Fehlertermvarianz;
- b) Worin unterscheiden sich das  $R^2$  und das angepasste  $R^2$ ? (1,5 Punkte)
- c) Interpretieren Sie den Koeffizienten des Parameters  $b_1$  statistisch und inhaltlich. (2 Punkte)
- d) Um Fehlspezifikation vorzubeugen, werden sodann die nachfolgenden Schritte durchgeführt. (8 Punkte)
- Unter der Annahme, dass die Größe der Bevölkerung im i-ten Bundesstaat ebenfalls eine Rolle spiele, wird die Modellspezifikation um den Indikator POP erweitert und erneut geschätzt.

```
Call:
lm(formula = PCON ~ REG + TAX + POP)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 554.28355   188.09310   2.947  0.00503 **
REG           0.16379    0.05855    2.797  0.00750 **
TAX          -53.63094   17.01040   -3.153  0.00285 **
POP            0.01492    0.03830    0.390  0.69870
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 255.3 on 46 degrees of freedom
Multiple R-Squared:  0.8668,    Adjusted R-squared:  0.8581
F-statistic: 99.79 on 3 and 46 DF,  p-value: < 2.2e-16
```

d11) Begründen Sie, welche der beiden Modellspezifikationen (aus Teilaufgabe a oder aus d) Sie bevorzugen.

d12) Kommen Sie zur selben Entscheidung, wenn Sie anhand der AIC- und BIC-Werte urteilen? Berechnen Sie diese (Hinweis: Im ersten Modell ist  $\log \frac{1}{N} \sum e_i^2 = 11.00491$ , im zweiten Modell ist  $\log \frac{1}{N} \sum e_i^2 = 11.00162$ ).

d2) Es soll ein RESET-Test für die Spezifikation aus Teilaufgabe a) durchgeführt werden. Stellen Sie die Hilfsregression für eine Variante dieses Tests auf.

d3) Das Modell wird in modifizierter Spezifikation geschätzt. Interpretieren Sie die Steigungsparameter der folgenden Gleichung:

$$\log(PCON) = 6.15 + 0.009 \cdot REG - 0.042 \cdot \log(TAX).$$

e) Es kann weiterhin argumentiert werden, dass mit ansteigender Zahl der Kfz-Zulassungen die Variation im Mineralöl-Konsum steigt, was vermuten lässt, dass Heteroskedastie vorliegt. Es werden daher für das ursprüngliche, nicht logarithmierte Modell aus Teilaufgabe a) der Breusch-Pagan Test und der White-Test auf Heteroskedastie durchgeführt. (10 Punkte)

e1) Der Breusch-Pagan Test hat eine Test-Statistik von 14.03 bei einem Freiheitsgrad von d.f.=1. Testen Sie auf dem 5%-Signifikanzniveau, ob Heteroskedastie vorliegt und dokumentieren Sie Ihre Vorgehensweise.

e2) Stellen Sie die Gleichung der Hilfsregression des White-Tests auf und testen Sie mit dem ermittelten  $R^2$ -Wert von 0.6645 auf dem 5%-Signifikanzniveau auf Heteroskedastie. Dokumentieren Sie Ihre Vorgehensweise.

e3) Erläutern Sie kurz für das zugrunde liegende Modell die Vorgehensweise eines GLS-Schätzers zur Beseitigung der beschriebenen Heteroskedastie.

e4) Zur Korrektur von Heteroskedastie unbekanntem Ursprungs werden gelegentlich White-Standardfehler berechnet. Erläutern Sie verbal je einen Vor- und einen Nachteil dieses Verfahrens.

## Aufgabe 2:

[12 Punkte]

Unterstellen Sie das folgende einfache Modell der Geldnachfrage:

$$\log(MI_t) = \beta_0 + \beta_1 \cdot \log(GDP_t) + \beta_2 \cdot \log(CPI_t) + e_t,$$

wobei

$MI_t$ : nominaler M1-Geldbestand im Jahr t

$GDP_t$ : reales Bruttonettoprodukt im Jahr t (in Mrd. US-\$)

$CPI_t$ : Preisindex im Jahr t.

Das Modell wird mit US-Quartalsdaten von 1950.1 bis 1999.4 geschätzt, ein (hier nicht dargestellter) Residuen-Plot lässt Autokorrelation vermuten. Aus einer KQ-Schätzung ergibt sich:

$$\sum_{t=2}^T \hat{e}_t \hat{e}_{t-1} = 1.357 \quad \text{und} \quad \sum_{t=2}^T \hat{e}_{t-1}^2 = 1.381$$

a) Berechnen Sie den Autokorrelationskoeffizienten  $\rho$  und die (approximative) Durbin-Watson-Statistik  $d$ . (4 Punkte)

b) Unterstellen Sie eine Durbin-Watson-Statistik von 1,0 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)

- c) Zeigen Sie für ein lineares Modell  $y_i = x_i' \beta + \varepsilon_i$  mit Autokorrelation zweiter Ordnung, also mit  $\varepsilon_i = \rho_1 \varepsilon_{i-1} + \rho_2 \varepsilon_{i-2} + \nu_i$ , wobei  $\nu_i \sim i.i.d.(0, \sigma^2)$ , wie eine Änderung der Spezifikation das Autokorrelationsproblem beseitigt. (4 Punkte)

**Aufgabe 3:**

**[16 Punkte]**

Unterstellen Sie das ‚wahre‘ Modell  $y_{it} = x_{it}' \beta + \alpha_i + \varepsilon_{it}$  mit  $\alpha_i$  als nicht-beobachtbarer, nicht-stochastischer zeitkonstanter Größe.

- a) Das Modell werde ohne Berücksichtigung der Panelnatur der Daten im Querschnitt geschätzt. Zeigen Sie, dass der KQ-Schätzer  $b = \sum (x_{it}' x_{it})^{-1} x_{it}' y_{it}$  verzerrt ist. (4 Punkte)
- b) Erläutern Sie kurz formal und verbal, wie man durch die Verwendung von Paneldaten  $\beta$  unverzerrt schätzen kann; unterstellen Sie  $t = 1, 2$ . (3 Punkte)
- c) Aus dem US-amerikanischen NLSY79 werden Daten von 545 vollzeitbeschäftigten Männern, die 1980 ihren Schulabschluss gemacht haben und über 8 Jahre beobachtet wurden, gezogen. Der logarithmierte Lohn wird auf die Merkmale Schuljahre, Arbeitsmarkterfahrung und quadrierte Arbeitsmarkterfahrung, Gewerkschaftsmitgliedschaft, Familienstand (verheiratet), Beschäftigung im öffentlichen Sektor und ethnische Zugehörigkeit regressiert. Es ergeben sich bei Anwendung unterschiedlicher Verfahren die folgenden Schätzergebnisse: (9 Punkte)

Abhängige Variable: ln(Wage)				
	KQ	between	fixed effects	random effects
Constant	-0.0344 (0.065)	0.490** (0.22)	—	-0.104 (0.11)
Schooling	0.0994*** (0.0047)	0.0948*** (0.011)	—	0.101*** (0.0089)
Experience	0.0891*** (0.010)	-0.0502 (0.050)	0.116*** (0.0084)	0.112*** (0.0083)
Experience^2	-0.00285*** (0.00071)	0.00511 (0.0032)	-0.00429*** (0.00061)	-0.00406*** (0.00059)
Union member	0.180*** (0.017)	0.274*** (0.047)	0.0812*** (0.019)	0.106*** (0.018)
Married	0.108*** (0.016)	0.145*** (0.041)	0.0451** (0.018)	0.0625*** (0.017)
Black	-0.144*** (0.024)	-0.139*** (0.049)	—	-0.144*** (0.048)
Hispanic	0.0157 (0.021)	0.00548 (0.043)	—	0.0197 (0.043)
Public sector	0.00355 (0.037)	-0.0563 (0.11)	0.0349 (0.039)	0.0302 (0.036)

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

- c1) Wieso werden für den fixed effects-Schätzer keine Parameterwerte für die Regressoren Schooling, Black und Hispanic ausgewiesen? Erläutern Sie ausführlich.
- c2) Erläutern Sie, warum sich die Schätzergebnisse für between- und fixed effects-Schätzer teils erheblich unterscheiden können.
- c3) Geben Sie sowohl die Bedingung an, unter der der random effects-Schätzer konsistent ist als auch die konkrete Hypothese, anhand derer dies im Rahmen des Hausman-Tests überprüft wird.

**Aufgabe 4:****[7 Punkte]**

Gegeben sei ein lineares Modell, das das Gewicht von Neugeborenen auf verschiedene soziodemographische Merkmale regressiert:

$$\log(\text{birthweight}) = \beta_0 + \beta_1 \text{male} + \beta_2 \text{parity} + \beta_3 \log(\text{fam\_inc}) + \beta_4 \text{packs} + u,$$

wobei

*birthweight*: Geburtsgewicht (in Unzen);

*male*: =1, falls das Neugeborene ein Junge ist;

*parity*: das wievielte Kind das Neugeborene ist;

*fam\_inc*: Familieneinkommen (in 1988 1000 US-\$);

*packs*: durchschnittliche Anzahl der während der Schwangerschaft pro Tag gerauchten Zigaretten (in Packungen).

- a) Im angeführten Modell könnte Endogenität eine Rolle spielen. Unter welcher Bedingung sind erklärende Variablen exogen? Erläutern Sie verbal und formal. (2 Punkte)
- b) Mit Daten einer US-amerikanischen Stichprobe ( $N=1388$ ) wird das Modell mit KQ und, um möglicher Endogenität Rechnung zu tragen, mit IV geschätzt. Für den IV-Schätzer wird *cigprice*, der Preis pro Packung Zigaretten in 1988 US-Cent im jeweiligen Bundesstaat, als Instrument für *packs* herangezogen. (5 Punkte)

Abhängige Variable:	KQ log(birthweight)	IV log(birthweight)	KQ packs
Constant	4.676*** (0.022)	4.468*** (0.26)	0.137 (0.10)
male	0.0262*** (0.010)	0.0298* (0.018)	-0.00473 (0.016)
parity	0.0147*** (0.0057)	-0.00124 (0.022)	0.0181** (0.0089)
log(faminc)	0.0180*** (0.0056)	0.0636 (0.057)	-0.0526*** (0.0087)
packs	-0.0837*** (0.017)	0.797 (1.09)	—
cigprice	—	—	0.000777 (0.00078)
$R^2$	0.04	—	0.03

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

- b1) Was lernen Sie aus dem Vergleich der Koeffizienten für die Variable *packs* in der KQ- und IV-Schätzung des Geburtsgewichts?
- b2) Diskutieren Sie die Qualität des vorliegenden IV-Schätzers.

**Aufgabe 5:****[40 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.

	Der J-Test wird als $t$ -Test durchgeführt.
	HAC Standardfehler unterstellen auf eine feste Anzahl von Perioden begrenzte Störtermkorrelationen.
	AR(1) Fehlerterme sind heteroskedastisch.
	Schätzgleichungen mit quadratischen erklärenden Variablen erfordern nichtlineare Schätzverfahren.
	Der Instrumentvariablenschätzer kann nicht mehr als ein Instrument pro Steigungsparameter nutzen.
	Die Qualität von Instrumentvariablenschätzern hängt von der Qualität der Instrumente ab.
	Im einfachsten Fall des KQ-Schätzers wird unterstellt, dass die erklärenden Variablen des linearen Modells keine Zufallsvariablen sind.
	Beim random effects-Schätzer wird unterstellt, dass der unbeobachtete Störterm der Beobachtungseinheit ( $\alpha_i$ ) mit dem Zufallsstörterm ( $\varepsilon_{it}$ ) korreliert ist.
	Die Koeffizienten endogener erklärender Variablen können als kausale Effekte interpretiert werden.
	Zur Unverzerrtheit des fixed effects-Schätzers ist keine Aussage möglich.
	Die kritischen Werte des Durbin-Watson-Tests sind für Tests auf positive Autokorrelation erster Ordnung definiert.
	Die Varianz des KQ-Schätzers lässt sich auf Basis von Information über die Störterme und die $X'X$ Matrix berechnen.
	Bei Autokorrelation in Form von moving average Störprozessen gibt es Fehlerterme, die nicht miteinander korreliert sind.
	Der within-Schätzer nutzt ausschließlich Unterschiede zwischen den Beobachtungseinheiten, um die Steigungsparameter zu identifizieren.
	Positive Autokorrelation kommt häufiger vor als negative.
	Je nach Wahl der Gewichtungsmatrix $W$ des GMM Schätzers ergeben sich unterschiedliche, aber stets konsistente Schätzer.
	Der Instrumentvariablenschätzer ist ein Spezialfall des GMM (generalized method of moments) Ansatzes.
	Grundidee des KQ-Schätzers ist, eine Linie so durch eine Punktwolke zu legen, dass die Summe der quadrierten horizontalen Abweichungen der beobachteten Werte von der Linie minimiert wird.
	Das least squares dummy variable (LSDV) Modell berücksichtigt für jede Beobachtungseinheit eine Dummyvariable.
	Der Breusch-Pagan Test auf Heteroskedastie ist eine Verallgemeinerung des White-Tests.
	Im Gegensatz zum KQ-Schätzer optimieren GMM (generalized method of moments) Schätzer keine Zielfunktion.
	Um $K$ Parameter zu identifizieren, benötigt man $K - 1$ Momentenbedingungen.

	Bei perfekter Multikollinearität ist die $XX$ Matrix nicht invertierbar.
	Der GIVE (generalized instrumental variables estimator) Schätzer kann in einem zweistufigen Verfahren angewendet werden.
	Newey-West Standardfehler korrigieren für Heteroskedastie unbekanntem Ursprungs ebenso wie für Autokorrelation.
	Wenn $X$ deterministisch ist, gilt $E\{\varepsilon   X\} = E\{\varepsilon\}$ .
	Bei Vorliegen von Paneldaten lassen sich Probleme ausgelassener Variablen durch den random effects Schätzer lösen.
	Im linearen Modell gibt die Regressionskonstante den Mittelwert der abhängigen Variable an.
	Je größer der Typ II Fehler eines Tests, umso größer muss der Typ I Fehler sein.
	Wird die Nullhypothese des Sargan-Tests verworfen, so ist das Schätzmodell überidentifiziert.
	Der marginale Effekt einer erklärenden Variablen kann je nach Spezifikation des Modells unterschiedlich ausfallen.
	Bei Messfehlern in der abhängigen Variable sind die Parameterschätzer auf Null hin verzerrt.
	Der Korrelationskoeffizient für die abhängige Variable und ihren vorhergesagten Wert ist immer größer als das $R^2$ der zugehörigen Schätzung.
	Das Residuum der linearen Regression ist orthogonal zu den vorhergesagten Werten der abhängigen Variablen.
	Mit dem $t$ -Test können mehrere lineare Hypothesen gleichzeitig getestet werden.
	Der Chow-Test prüft, ob die lineare oder loglineare Version der abhängigen Variable dem Modell besser entspricht.
	Insignifikante Koeffizienten können größer als 100 sein.
	Bei der Ableitung des KQ-Schätzers im linearen Modell erhält man so viele Normalgleichungen wie unbekannte Parameter vorliegen.
	Die within Transformation modifiziert alle Beobachtungen, indem die Varianz des individuellen Störterms von den erklärenden Variablen abgezogen wird.
	Mithilfe hedonischer Preisfunktionen lassen sich einzelne Eigenschaften eines Gutes bewerten.
	Man kann den $t$ -Test verwenden, um die Nullhypothese positiver Autokorrelation erster Ordnung zu überprüfen.
	Der between-Schätzer ist konsistent, wenn der beobachtungsspezifische Störterm nicht mit den beobachtungsspezifischen Mittelwerten der erklärenden Variablen korreliert ist.
	Die Nullhypothese $\beta \geq \beta_k$ wird am 5 Prozent Signifikanzniveau bei 500 Freiheitsgraden verworfen, wenn der empirische $t$ -Wert größer als 1,645 ist.
	Die Inkonsistenz eines Steigungsparameters führt nicht zur Inkonsistenz der gleichzeitig geschätzten Regressionskonstanten.
	Mit dem RESET Test können Probleme bei der Spezifikation des Regressionsmodells aufgedeckt werden.
	Monte Carlo Studien stützen sich auf per Zufallsziehung generierte Daten.
	Der Durbin-Watson-Test auf AR(1) Fehler ist nicht anwendbar, wenn das Modell verzögerte endogene Variablen als erklärende Variablen aufweist.

	Bei einem gegebenen Schätzverfahren und Modell können sich für unterschiedliche Stichproben verschiedene Schätzwerte ergeben.
	Mit Hilfe des Goldfeld-Quandt Tests lassen sich fixed- und random effects-Schätzer gegeneinander testen.
	Bei Heteroskedastie ist der KQ-Schätzer verzerrt.
	Berücksichtigt man im Modell zu viele erklärende Variablen, so steigt die Streuung der geschätzten Parameter.
	Die Dichtefunktion der $t$ -Verteilung hat ihr Maximum bei 1,96.
	Der KQ-Schätzer erzeugt für Steigungsparameter aber nicht für die Regressionskonstante Zufallsvariablen.
	Bei GLS (generalized least squares) Schätzern sind $t$ - und F-Tests nicht einmal approximativ anwendbar.
	Simultane Gleichungssysteme in struktureller Form können mit dem KQ-Schätzer konsistent geschätzt werden, solange jede Gleichung einzeln betrachtet wird.
	Das BIC Kriterium fällt umso günstiger aus, je kleiner die Fehlerquadratsumme bei gegebener Parameter- und Beobachtungszahl ist.
	Bei strikter Exogenität der erklärenden Variablen ist der fixed effects-Schätzer konsistent.
	Der White-Schätzer benutzt gruppenspezifische Mittelwerte, um heteroskedastische Koeffizienten zu berechnen.
	Das angepasste $R^2$ einer Schätzung kann bei Hinzufügen von erklärenden Variablen sinken.
	Mithilfe eines linearen Regressionsmodells lassen sich Elastizitäten schätzen.

### 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**.

	Die Zielfunktion $S(\tilde{\beta}) = \sum_{i=1}^N (y_i - g(x_i, \tilde{\beta}))^2$ ergibt keinen KQ-Schätzer.
	Wenn $a$ ein konsistenter Schätzer für $\alpha$ ist, dann ist $\sqrt{a}$ ein konsistenter Schätzer für $\sqrt{\alpha}$ .
	Die Berücksichtigung einer verzögerten endogenen Variable als erklärender Variable kann zur Inkonsistenz des KQ-Schätzers führen.
	Der PE Test überprüft die Spezifikation des Modells.



	Unter den Annahmen des Gauss-Markov Theorems hat der KQ-Schätzer die kleinste Varianz unter allen linearen Schätzern.
	Der iterative Cochrane-Orcutt Schätzer wird in zwei Stufen geschätzt.
	Unter den Annahmen des Gauss-Markov Theorems sind die mittels KQ-Schätzer erzeugten Vorhersagen verzerrt, aber konsistent.
	Die Endogenität einer erklärenden Variablen lässt sich mit Hilfe eines $t$ -Tests überprüfen.
	Bei KQ-Schätzern sind $t$ - und F-Tests asymptotisch zutreffend, selbst wenn der Störterm der Regression nicht normalverteilt ist.
	random effects-, between- und within-Schätzer stehen in einem linearen Zusammenhang.

**Aufgabe 7:**

**[6 Punkte]**

Das lineare Modell  $y_i = x_i' \beta + \varepsilon_i$  kann genutzt werden, um an konkreten Ausprägungen von  $x_0$  Vorhersagen von  $y$  zu generieren. Leiten Sie die Varianz des Vorhersagefehlers ab und zeigen Sie formal formal, in welcher Form diese von  $\sigma^2$  abhängt.