

Wirtschaftswissenschaftliches Prüfungsamt

Diplomprüfung / Bachelorprüfung

Ökonometrie

Sommersemester 2011, Erster Prüfungstermin, 03. August 2011

Prof. Dr. Ralph Friedmann

Name, Vorname: _____

Matrikelnummer: _____

B i t t e b e a c h t e n S i e :

- (a) Kleben Sie bitte Ihr Namensschild auf die dafür vorgesehene **Markierung auf dem Deckblatt des Klausurhefts!**
- (b) Legen Sie Ihren Lichtbild- und Ihren Studierendenausweis an Ihrem Platz aus.
- (c) Die Klausur besteht aus fünf Aufgaben mit jeweils zwei untergliederten Teilen. Die vollständige Lösung einer Aufgabe wird mit 40 Punkten bewertet. Die Klausur enthält keinen Tabellenanhang. Die Aufgaben 1 und 2 sind verpflichtend. Von den Aufgaben 3, 4 und 5 ist genau eine zu bearbeiten. Sollten Sie von diesen Aufgaben mehr als eine bearbeiten, wird in der Reihenfolge der Bearbeitung lediglich die erste gültige gewertet.
- (d) Bei jeder Multiple-Choice Aufgabe ist genau eine Alternative richtig.
- (e) Bearbeiten Sie Teil 1 jeder Aufgabe im Klausurenheft. Die Lösungen der Multiple-Choice Aufgaben sind in der Aufgabensammlung mit einem Kreuz zu markieren. Diese ist daher am Ende der Klausur ebenfalls abzugeben.
- (f) Die Reihenfolge der Bearbeitung der Aufgaben kann beliebig gewählt werden, beginnen Sie aber für jede Aufgabe eine neue Seite.
- (g) Überprüfen Sie die Vollständigkeit Ihres Klausurexemplares. Spätere Reklamationen können nicht berücksichtigt werden.

- (h) Die Benutzung von zwei beidseitig beschriebenen bzw. vier einseitig beschriebenen DIN A4-Blättern sowie (auch programmierbaren) Taschenrechnern ist erlaubt.
- (i) Bei allen statistischen Tests sind die Hypothesen, die Teststatistik sowie deren Verteilung unter H_0 , der kritische Bereich, die Realisation der Teststatistik sowie die Testentscheidung anzugeben. Ist das Signifikanzniveau nicht explizit angegeben, so ist $\alpha = 0,05$ zu verwenden.

Aufgabe 1 [Teil 1: 30 Punkte; Teil 2: 10 Punkte]

Teil 1:

- (a) Gegeben sei ein einfaches lineares Regressionsmodell der Art

$$Y_i = \beta_0 + \beta_1 X_i + u_i \quad .$$

Erklären Sie, was man im Rahmen eines solchen Modells unter dem *Bestimmtheitsmaß* versteht und wie man es berechnet. Gehen Sie ferner auf den Bereich möglicher Werte ein und ziehen Sie einen Vergleich zum *adjustierten Bestimmtheitsmaß*. Diskutieren Sie, inwieweit das Bestimmtheitsmaß geeignet ist, um Aussagen über

- (i) Kausalität der exogenen Variablen hinsichtlich der endogenen,
- (ii) Signifikanz der Regressionskoeffizienten,
- (iii) Erklärung der Varianz der abhängigen Variablen durch das Modell oder
- (iv) Probleme hinsichtlich vergessener Variablen

zu treffen.

Zeigen Sie ferner den formalen Zusammenhang zwischen dem Bestimmtheitsmaß und dem Korrelationskoeffizienten im Rahmen eines einfachen linearen Regressionsmodells.

- (b) Gegeben seien die folgenden fünf alternativen Modellspezifikationen:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \tag{1}$$

$$\log(Y_i) = \beta_0 + \beta_1 X_i + u_i \tag{2}$$

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i \tag{3}$$

$$\log(Y_i) = \beta_0 + \beta_1 \log(X_i) + u_i \tag{4}$$

$$Y_i = \beta_0 + \beta_1 \log(X_i) + u_i \tag{5}$$

- (i) Interpretieren Sie die Steigungsparameter der fünf Modellspezifikationen indem Sie zeigen, wie sich eine Änderung von X auf Y auswirkt.
- (ii) KQ-Schätzungen liefern folgende Bestimmtheitsmaße:
(1) $R^2 = 0,58$; (2) $R^2 = 0,61$; (3) $R^2 = 0,74$; (4) $R^2 = 0,82$; (5) $R^2 = 0,79$.
Inwieweit können Sie aus diesen Werten Schlüsse für die Auswahl einer geeigneten Modellspezifikation ziehen?
- (c) Skizzieren Sie für ein einfaches lineares Regressionsmodell ein Streudiagramm für die Punkte (X, Y) und eine zugehörige Regressionsgerade für den Fall, dass X keinen Einfluss auf Y hat.

Teil 2: Welche der folgenden Aussagen sind richtig? (Markieren Sie Ihre Antwort durch eindeutiges Ankreuzen.)

- (a) Wenn der Steigungskoeffizient im einfachen Regressionsmodell, $\hat{\beta}_1$, Null ist, dann gilt:
- (i) $R^2 = \bar{Y}$;
 - (ii) $0 < R^2 < 1$;
 - (iii) $R^2 = 0$;
 - (iv) $R^2 > SSR/TSS$.
- (b) Im linearen Regressionsmodell erhält man den OLS-Schätzer für β durch Minimieren
- (i) der Summe der Regressionsresiduen;
 - (ii) der Summe der absoluten Regressionsresiduen;
 - (iii) der Summe der quadrierten Abweichungen zwischen den Beobachtungen der endogenen Variable und der dafür durch das Modell prognostizierten Werte;
 - (iv) der geschätzten Varianz der Regressionskoeffizienten.
- (c) In einem einfachen linearen Regressionsmodell, welches die in der Vorlesung besprochenen Annahmen erfüllt, gilt:
- (i) $\text{Var}(Y_i) = \beta_0^2 + \beta_1^2 \text{Var}(X_i) + \text{Var}(u_i)$;
 - (ii) $\text{Var}(Y_i) = \text{Var}(u_i)$;
 - (iii) $\text{Var}(Y_i) = \beta_1^2 \text{Var}(X_i) + \text{Var}(u_i)$;
 - (iv) $\text{Var}(Y_i) = \beta_1 \text{Var}(X_i) + \text{Var}(u_i)$.
- (d) Für die Summe der Residuen im Rahmen einer OLS-Schätzung bei einem Modell des Typs $Y_i = \beta_0 + \beta_1 X_i + u_i$ ergibt sich
- (i) eine positive Zahl, da $\hat{\beta}$ mit Hilfe der quadrierten Residuen ermittelt wird;
 - (ii) Null;
 - (iii) eine Zahl, deren Wert davon abhängt, ob die endogene Variable hauptsächlich positiv oder negativ ist;
 - (iv) eine stark negative Zahl, weil nach einem Minimum gesucht wird.

(e) Eine Multiplikation der endogenen Variable mit 10 und der exogenen Variable mit 100 hat bei einer OLS-Schätzung in einem einfachen linearen Modell keine Auswirkungen auf

- (i) $\hat{\beta}_1$;
- (ii) $\hat{\beta}_0$;
- (iii) R^2 ;
- (iv) $\text{Var}(\hat{\beta}_0)$, $\text{Var}(\hat{\beta}_1)$.

Aufgabe 2 [Teil 1: 32 Punkte; Teil 2: 8 Punkte]

Teil 1:

- (a) Bei einer Untersuchung an kalifornischen Grundschulen stand u.a. der kausale Effekt des Schüler-Lehrer-Verhältnisses (`stratio`) auf das Abschneiden in einem standardisierten Leistungstest (`score`) im Fokus. Ein Teilergebnis der Untersuchung lautet:

Call:

```
lm(formula = score ~ stratio + expenditure + english, data = CASchools)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-51.340 -10.111   0.293  10.318  43.181
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 649.577947  15.205717  42.719 < 2e-16 ***
stratio     -0.286399         ???  -0.596  0.55149
expenditure  0.003868   0.001412   ???  0.00643 **
english           ???   0.039106 -16.776 < 2e-16 ***
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 14.35 on 416 degrees of freedom

Multiple R-squared: 0.4366, Adjusted R-squared: 0.4325

F-statistic: ??? on ??? and ??? DF, p-value: < 2.2e-16

Hierbei stehen

- `expenditure` für die Ausgaben pro Schüler,
- `english` für den Anteil der Englisch-Lerner (= Nicht-Muttersprachler),
- `???` für fehlende Werte.

- (i) Erklären Sie, was hier gemacht wurde. Gehen Sie dabei auf das unterstellte Modell, abhängige und unabhängige Variablen sowie unterstellte Nullhypothesen ein. Geben Sie den Stichprobenumfang n an und berechnen Sie die fünf fehlenden Werte.

(Anm.: Gehen Sie bei Ihren Berechnungen von homoskedastischen Störgrößen aus.)

- (ii) Interpretieren Sie alle Schätzergebnisse hinsichtlich ihrer Signifikanz.
 - (iii) Welche Gründe können Sie dafür anführen, die Variablen `expenditure` und `english` einzuführen, wenn eigentlich nur der Einfluss von `stratio` interessiert?
- (b) Zur Diskussion steht nun, ob die Variablen `stratio` und `expenditure` gemeinsam einen Einfluss auf `score` haben. Die Analyse führt zu folgendem Output:

Analysis of Variance Table

```

Model 1: score ~ english
Model 2: score ~ stratio + expenditure + english
  Res.Df  RSS Df
1     418 89000
2     416 85700  2

```

Formulieren, testen und interpretieren Sie die entsprechende Nullhypothese. (Anm.: Auch hier können Sie von homoskedastischen Störgrößen ausgehen. Interpretieren Sie Werte der entsprechenden Teststatistik > 5 als signifikant.)

- (c) Die bisherigen Ergebnisse beruhen auf der Annahme homoskedastischer Störgrößen. Erklären Sie, was man hierunter versteht und nehmen Sie kritisch dazu Stellung.
- (d) Unter Berücksichtigung heteroskedastischer Störgrößen ergibt sich für die Varianz-Kovarianz-Matrix von $\hat{\beta}$:

	(Intercept)	stratio	expenditure	english
(Intercept)	245.50573620	-6.8268126913	-2.135819e-02	8.528940e-02
stratio	-6.82681269	0.2376688839	4.115836e-04	-2.530641e-03
expenditure	-0.02135819	0.0004115836	2.583757e-06	-1.065003e-05
english	0.08528940	-0.0025306410	-1.065003e-05	1.031327e-03

Führen Sie die Koeffiziententests unter Aufgabenteil (a) erneut durch (abgesehen vom F -Test), indem Sie die t -Werte berechnen. Fällen Sie eine Testentscheidung zu einem Signifikanzniveau von $\alpha = 0,01$ und vergleichen Sie Ihre Ergebnisse mit den vorherigen.

(Anm.: `qnorm(0.995)` ergibt den Wert 2.575829.)

- (e) Erklären Sie anhand eines Beispiels im Zusammenhang mit dem in dieser Aufgabe diskutierten Modell, was man unter *Interaktionstermen* versteht.

Teil 2: Welche der folgenden Aussagen sind richtig? (Markieren Sie Ihre Antwort durch eindeutiges Ankreuzen.)

- (a) Im multiplen Regressionsmodell gilt für das adjustierte Bestimmtheitsmaß:
 - (i) Es kann nicht negativ werden;
 - (ii) Es kann nicht größer als das multiple Bestimmtheitsmaß sein;
 - (iii) Es kann nicht kleiner werden, wenn zusätzliche exogene Variablen zum Modell hinzugefügt werden;
 - (iv) Bei Hinzunahme zusätzlicher exogener Variablen entwickelt es sich proportional zum multiplen Bestimmtheitsmaß.

- (b) Wenn der Regressor \mathbf{X} mit einer Variablen \mathbf{O} korreliert, die im Regressionsmodell nicht berücksichtigt wurde, obwohl sie Y zumindest in Teilen determiniert, dann
 - (i) kann man den Effekt der vergessenen Variable nicht messen, die OLS-Schätzer bleiben hiervon allerdings unberührt;
 - (ii) bleiben die OLS-Schätzer unverzerrt, ihre Varianz wird durch den Ausschluss von \mathbf{O} allerdings größer;
 - (iii) sind die OLS-Schätzer verzerrt;
 - (iv) sind die OLS-Schätzer immer nach unten verzerrt.

(c) Gegeben sei folgendes lineares Modell:

$$Y_i = \beta_0 + \beta_1 D_{1,i} + \beta_2 D_{2,i} + u_i \quad .$$

Dabei seien:

Y_i : Einkommen der Untersuchungsperson i ;

$$D_{1,i} = \begin{cases} 1 & \text{wenn } i \text{ ein Mann ist} \\ 0 & \text{sonst} \end{cases} \quad ;$$

$$D_{2,i} = \begin{cases} 1 & \text{wenn } i \text{ eine Frau ist} \\ 0 & \text{sonst} \end{cases} \quad .$$

Was erwarten Sie im Rahmen einer OLS-Schätzung für $\hat{\beta}$?

- (i) $\hat{\beta}_1$ wird positiv sein, $\hat{\beta}_2$ negativ;
- (ii) $|\hat{\beta}_0 - \hat{\beta}_1| = |\hat{\beta}_0 - \hat{\beta}_2|$;
- (iii) Probleme hinsichtlich Multikollinearität;
- (iv) $\hat{\beta}$ ist aufgrund vergessener Variablen verzerrt.

(d) Im Fall eines linearen Modells mit zwei stetigen exogenen Variablen

- (i) determiniert β eine Regressionsebene;
- (ii) können die individuellen Auswirkungen einer Änderung nur einer exogenen Variable auf die endogene nicht mehr ermittelt werden;
- (iii) kann das adjustierte Bestimmtheitsmaß nicht mehr sinnvoll interpretiert werden;
- (iv) kann es keine Probleme mit vergessenen Variablen geben.

Aufgabe 3 [Teil 1: 32 Punkte; Teil 2: 8 Punkte]

Teil 1:

Betrachtet wird ein balanciertes Panel für alle US-Bundesstaaten über die Jahre 1977-1999 zu den folgenden Variablen:

<code>state</code>	Indikator des Bundesstaates
<code>year</code>	Indikator des Jahres
<code>violent</code>	Gewaltverbrechen pro 100.000 Einwohner
<code>law</code>	Indikator für gesetzlich erlaubten Waffenbesitz
<code>afam</code>	Bevölkerungsanteil schwarz
<code>cauc</code>	Bevölkerungsanteil weiß
<code>density</code>	Bevölkerung pro Quadratmeile/1.000
<code>income</code>	Durchschnittseinkommen in 1.000 USD
<code>male</code>	Anteil der männlichen Bevölkerung
<code>population</code>	Bevölkerung in Millionen
<code>prisoners</code>	Vorjahreswert für den Bevölkerungsanteil in Gefängnissen pro 100.000 Einwohner

- (a) Eine einfache Regression von $(\log(\text{violent}))$ auf `law` liefert folgendes Ergebnis:

t test of coefficients:

```
                Estimate Std. Error  t value  Pr(>|t|)
(Intercept)  6.134919   0.019309  317.7199 < 2.2e-16 ***
lawyes                ???   0.047636  -9.2989 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- (i) Erklären Sie, was man unter einem Panel versteht und gehen Sie auf Gemeinsamkeiten und Unterschieden zu anderen Daten- bzw. Untersuchungstypen ein.
- (ii) Berechnen und interpretieren Sie den Koeffizienten von `law`.
- (iii) In diesem Modell ist `law` der einzige echte Regressor. Auf welches Problem sollten Sie deshalb achten?

(b) Ein erweitertes Modell führt zu folgendem Ergebnis:

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	2.9817e+00	6.1364e-01	4.8591	1.340e-06	***
lawyes	???	3.4985e-02	-10.5298	< 2.2e-16	***
afam	8.0853e-02	2.0274e-02	3.9880	7.080e-05	***
cauc	3.1200e-02	9.8246e-03	3.1758	0.001534	**
density	2.6688e-02	1.5011e-02	1.7779	0.075681	.
income	1.2051e-06	7.3806e-06	0.1633	0.870324	
male	8.8709e-03	1.2317e-02	0.7202	0.471545	
population	4.2710e-02	3.2120e-03	13.2969	< 2.2e-16	***
prisoners	1.6126e-03	1.9205e-04	8.3972	< 2.2e-16	***

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(i) Berechnen Sie auch hier den Koeffizienten für law und vergleichen Sie ihn mit Ihrem Ergebnis aus (a)(ii).

(ii) Begründen Sie, weshalb es auch hier zu Problemen hinsichtlich vergessener Variablen kommen kann. Stellen Sie in diesem Zusammenhang eine Möglichkeit vor, wie Sie für vergessene Variablen kontrollieren können, ohne diese überhaupt beobachtet zu haben.

(c) Für den folgenden Output wurde das Modell um *fixed effects* hinsichtlich der Untersuchungseinheiten ergänzt¹:

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	4.0368e+00	3.9748e-01	10.1559	< 2.2e-16	***
lawyes	-4.6141e-02	2.0516e-02	-2.2490	0.0247065	*
afam	1.0428e-01	1.7370e-02	6.0033	2.612e-09	***
cauc	4.0861e-02	5.5417e-03	7.3734	3.239e-13	***
density	-1.7229e-01	1.2074e-01	-1.4269	0.1538767	
income	-9.2037e-06	6.9417e-06	-1.3259	0.1851575	
male	-5.0273e-02	8.0479e-03	-6.2466	5.957e-10	***
population	1.1525e-02	1.0270e-02	1.1222	0.2620300	
prisoners	-7.1008e-05	1.0210e-04	-0.6954	0.4869220	

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

¹Die Koeffizienten der entsprechenden Dummies interessieren hier nicht und werden deshalb nicht aufgeführt.

- (i) Erklären Sie, was man unter dem Begriff *fixed effects* hinsichtlich der Untersuchungseinheiten versteht und skizzieren Sie eine entsprechende Regressionsgleichung.
 - (ii) Wie verändern sich die Untersuchungsergebnisse, insbesondere hinsichtlich der ursprünglich interessierenden Variable `law`?
 - (iii) Welches Problem bleibt weiterhin relevant und wie könnten Sie es in den Griff bekommen?
(Anm.: Denken Sie an die Relevanz zeitlicher Veränderungen.)
- (d) Die Berücksichtigung von *fixed effects* hinsichtlich der Zeit liefert folgendes Ergebnis:²

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	3.9720e+00	4.6903e-01	8.4686	< 2.2e-16	***
lawyes	-2.7994e-02	2.0099e-02	-1.3928	0.1639752	
afam	2.9186e-02	2.2189e-02	1.3154	0.1886647	
cauc	9.2500e-03	8.9102e-03	1.0381	0.2994367	
density	-9.1555e-02	7.2131e-02	-1.2693	0.2046065	
income	9.5859e-07	7.5268e-06	0.1274	0.8986815	
male	7.3326e-02	1.9693e-02	3.7234	0.0002066	***
population	-4.7545e-03	7.1039e-03	-0.6693	0.5034640	
prisoners	7.5994e-05	8.7346e-05	0.8700	0.3844738	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

- (i) Erklären Sie, was man unter dem Begriff *fixed effects* hinsichtlich der Zeit versteht und skizzieren Sie ein entsprechendes Regressionsmodell.
- (ii) Zu welcher Schlussfolgerung gelangen Sie nun?
- (iii) Welche Mängel könnte das Modell immer noch aufweisen?

²Auch hier werden die Koeffizienten der nicht-interessierenden Dummies nicht aufgeführt.

Teil 2: Welche der folgenden Aussagen sind richtig? (Markieren Sie Ihre Antwort durch eindeutiges Ankreuzen.)

- (a) Der Unterschied zwischen einem unbalancierten und einem balancierten Panel ist, dass
- (i) nicht gleichzeitig *fixed effects* hinsichtlich der Untersuchungseinheiten und hinsichtlich der Zeit vorliegen können;
 - (ii) in einem unbalancierten Panel für wenigstens eine Untersuchungseinheit und eine Zeitperiode keine Daten vorliegen;
 - (iii) in einem balancierten Panel alle Regressoren in etwa den gleichen Einfluss auf die endogene Variable haben;
 - (iv) in einem balancierten Panel alle Regressoren in etwa gleich stark mit der Störgröße korrelieren.
- (b) In einem *fixed effects* Modell der Art $Y_{i,t} = \beta_0 + \beta_1 X_{i,t} + \gamma_1 D_{1,i} + \dots + \gamma_n D_{n,i} + u_{i,t}$
- (i) schwanken die Steigungskoeffizienten zwischen den Untersuchungseinheiten, der Achsenabschnitt ist aber konstant;
 - (ii) können nicht zusätzlich *fixed effects* hinsichtlich der Zeit ergänzt werden;
 - (iii) ist der OLS-Schätzer für $\hat{\beta}$ prinzipiell verzerrt;
 - (iv) besteht ein Multikollinearitätsproblem.
- (c) Betrachtet werde eine Panelregression für die Arbeitslosenzahlen der G7 Staaten für die Zeitperiode von 1976-1998 (jährliche Daten). Wie groß ist die Anzahl der zusätzlich zu β_0 benötigten Dummyvariablen, wenn im Modell *fixed effects* sowohl hinsichtlich der Untersuchungseinheiten als auch hinsichtlich der Zeit berücksichtigt werden sollen?
- (i) 6;
 - (ii) 27;
 - (iii) 28;
 - (iv) 29.
- (d) Die Koeffizienten in einem Probit-Modell werden typischerweise geschätzt mittels
- (i) Transformation der OLS-Schätzer aus dem linearen Wahrscheinlichkeitsmodell;
 - (ii) einem zweistufigen Kleinste-Quadrate-Verfahren;
 - (iii) Maximum-Likelihood-Schätzung;
 - (iv) Logit-Ansatz.

Aufgabe 4 [Teil 1: 32 Punkte; Teil 2: 8 Punkte]

Teil 1:

(a) Beschreiben Sie die Konsequenzen einer OLS-Schätzung bei endogenen erklärenden Variablen und erklären Sie, wie man die entsprechenden Probleme mittels Instrumentalvariablenschätzung (IV-Schätzung) in den Griff bekommen kann. Formulieren Sie dazu ein entsprechendes allgemeines IV-Modell und benennen Sie die verwendeten Variablen sowie die Voraussetzungen, die gelten müssen, um konsistente Schätzergebnisse zu erhalten. Gehen Sie dabei insbesondere auf die Begriffe *Validität* von Instrumenten, *Identifikation* und *Annahmen* der IV-Schätzung ein.

(b) Gegeben sei ein einfaches lineares Modell der Art

$$Y_i = \beta X_i + u_i \quad \text{mit} \quad \text{Cov}(X_i, u_i) = 0 \quad ,$$

welches die drei KQ-Annahmen erfüllt [$E(u_i|X_i) = 0$; $(X_i, Y_i) \stackrel{\text{i.i.d.}}{\sim} (X, Y)$; $0 < E(X_i^4) < \infty$, $0 < E(Y_i^4) < \infty$].

(i) Warum ist X ein valides Instrument?

(ii) Warum ist Y kein valides Instrument?

(iii) Zeigen Sie, dass der IV-Schätzer in diesem Fall identisch ist mit dem OLS-Schätzer.

(c) Gegeben sei folgendes Regressionsmodell ohne Konstante:

$$Y_i = \beta X_i + u_i \quad \text{mit} \quad \text{Cov}(X_i, u_i) \neq 0 \quad .$$

Ferner existiere für jede Beobachtung i eine weitere Variable Z_i mit

$$\text{Cov}(Z_i, u_i) = 0 \quad , \quad \text{Cov}(Z_i, X_i) \neq 0 \quad .$$

Zeigen Sie, dass eine 2SLS-Schätzung mit dem Instrument Z zu

$$\hat{\beta}^{(2SLS)} = \frac{\sum_{i=1}^n Y_i Z_i}{\sum_{i=1}^n X_i Z_i}$$

führt, wenn Sie auf der ersten Stufe das Modell

$$X_i = \gamma Z_i + e_i$$

schätzen.

Teil 2: Welche der folgenden Aussagen sind richtig? (Markieren Sie Ihre Antwort durch eindeutiges Ankreuzen.)

Hinweis: Lösen Sie die folgenden Aufgaben im Kontext eines linearen Regressionsmodells, in dem endogene Regressoren auftreten.

- (a) Die Verfügbarkeit vieler relevanter Instrumentalvariablen
 - (i) ist problematisch, weil das Regressionsmodell dann möglicherweise überidentifiziert ist;
 - (ii) ist prinzipiell positiv zu bewerten, weil dann mehr Informationen zur Verfügung stehen;
 - (iii) führt typischerweise zu größeren Standardfehlern der 2SLS-Schätzer;
 - (iv) ist nur nützlich, wenn ebenso viele exogene Variablen zur Verfügung stehen.
- (b) Instrumentalvariablenschätzung verwendet Instrumente, um
 - (i) Probleme im Zusammenhang mit heteroskedastischen Störgrößen zu vermeiden;
 - (ii) das Bestimmtheitsmaß des Modells zu erhöhen;
 - (iii) Probleme im Zusammenhang mit autokorrelierten Störgrößen zu vermeiden;
 - (iv) Variationen in X zu isolieren, die nicht mit der Störgröße korrelieren.
- (c) In praktischen Anwendungen ist eines der größten Probleme im Zusammenhang mit IV-Schätzung
 - (i) valide Instrumente zu finden;
 - (ii) die Durchführung des zweistufigen Regressionsansatzes;
 - (iii) die Berechnung von Standardfehlern und F -Statistiken;
 - (iv) exogene Instrumente zu finden, denn relevante sind leicht zu finden.

- (d) Folgendes Gleichungssystem beschreibe die Arbeitsnachfrage L^D und das Arbeitsangebot L^S in Abhängigkeit vom Reallohn (W/P) und staatlichen Transferleistungen T in verschiedenen Wirtschaftssektoren i :

$$L_i^D = \beta_0 + \beta_1 \left(\frac{W}{P} \right)_i + u_i \quad ;$$

$$L_i^S = \gamma_0 + \gamma_1 \left(\frac{W}{P} \right)_i + \gamma_2 T_i + v_i \quad .$$

Im Gleichgewicht gilt $L_i = L_i^D = L_i^S$.

- (i) Der Reallohn ist exogen, die gleichgewichtige Arbeitsmenge endogen.
- (ii) Sowohl L als auch (W/P) sind endogen.
- (iii) Der Reallohn ist endogen, die gleichgewichtige Arbeitsmenge exogen.
- (iv) Sowohl L als auch (W/P) sind exogen.

Aufgabe 5 [Teil 1: 32 Punkte; Teil 2: 8 Punkte]

Teil 1:

- (a) Gegeben sei ein stationärer Prozess der Art

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t \quad .$$

- (i) Was versteht man unter dem Konzept der Stationarität?
 (ii) Wie bezeichnet man einen solchen Prozess und welche Parameterrestriktionen ermöglichen die Stationarität?
 (iii) Zeigen Sie, dass $E(Y_t) = \frac{\beta_0}{1 - \beta_1}$ gilt.

- (b) Ein Index der Industrieproduktion (IP_t) misst monatlich die Menge der produzierten Industriegüter einer Volkswirtschaft. Sei $Y_t := 1200 \cdot \ln \left(\frac{IP_t}{IP_{t-1}} \right)$.

- (i) Trifft folgende Aussage zu: „ Y_t misst in etwa die auf ein Jahr hochgerechnete prozentuale Veränderung des Index der Industrieproduktion.“ ? Begründen Sie Ihre Antwort.
 (ii) Die Schätzung eines AR(4)-Modells aus einem Datensatz mit Werten für die Industrieproduktion von Januar 1970 bis Dezember 2010 liefert das folgende Ergebnis:

$$\hat{Y}_t = 1,377 + 0,318Y_{t-1} + 0,123Y_{t-2} + 0,068Y_{t-3} + 0,001Y_{t-4} \quad .$$

Ferner stehen für Juli bis Dezember 2010 folgende Werte zur Verfügung:

Zeitpunkt	2010:07	2010:08	2010:09	2010:10	2010:11	2010:12
IP	147,595	148,650	148,973	148,660	148,206	147,300
Y						

Ermitteln Sie \hat{Y} und \widehat{IP} für Januar 2011.

- (c) Das Modell in (b)(ii) wird aufgrund ökonomischer Überlegungen auf der rechten Seite um Y_{t-12} ergänzt und der entsprechende Koeffizient mit 0,054 geschätzt bei einer geschätzten Varianz von 0,00281. Halten Sie die Berücksichtigung des Koeffizienten für sinnvoll? Untermauern Sie ihre Antwort mit statistischen Argumenten.
 (d) Ferner wird die Beschränkung auf die Lagordnung 4 kritisiert. Aus diesem Grund soll die optimale Laglänge anhand statistischer Kriterien ermittelt werden. Hierzu werden für $p = 1, \dots, 6$ folgende Residuenquadratsummen notiert:

p	1	2	3	4	5	6
SSR	29175	28538	28393	28391	28378	28652

- (i) Angenommen die Werte für $p = 1, \dots, 5$ sind richtig. Wieso ist dann der SSR -Wert für $p = 6$ nicht plausibel?
- (ii) Bestimmen Sie den Stichprobenumfang T für die Schätzung der $AR(p)$ -Modelle.
- (iii) Bestimmen Sie die optimale Lagordnung anhand eines geeigneten Informationskriteriums. Der korrekte SSR -Wert für $p = 6$ ist 28317.

Teil 2: Welche der folgenden Aussagen sind richtig? (Markieren Sie Ihre Antwort durch eindeutiges Ankreuzen.)

- (a) *ADL*-Modelle beinhalten
 - (i) aktuelle und verzögerte Werte der Störgröße;
 - (ii) verzögerte Werte der endogenen Variable sowie zusätzlicher exogener Regressoren;
 - (iii) aktuelle und verzögerte Werte der Residuen;
 - (iv) verzögerte und zukünftige Werte der endogenen Variable.

- (b) $\Delta \ln(Y_t)$ entspricht
 - (i) ΔY_t
 - (ii) $Y_{t+1} - Y_{t-1}$;
 - (iii) ungefähr der Wachstumsrate von Y , sofern die Wachstumsrate klein ist;
 - (iv) der Wachstumsrate von Y .

- (c) Betrachtet wird ein AR(1)-Prozess Y_t mit $\rho_1 := \text{Korr}(Y_t, Y_{t-1}) < 0$ und $E(Y_t) := \mu$. Dann
 - (i) gilt $Y_t < 0$ für alle $t = 1, \dots, T$;
 - (ii) ist der Prozess keinesfalls stationär;
 - (iii) kann man die Daten nicht logarithmieren;
 - (iv) folgt auf einen Wert des Prozesses in einer Periode, der oberhalb des Erwartungswertes liegt, in der nachfolgenden Periode tendenziell ein Wert, der unterhalb des Erwartungswertes liegt.

- (d) Um die Anzahl der Lags eines AR-Modells zu bestimmen, ist welches Konzept ungeeignet:
 - (i) F -Test;
 - (ii) Akaike-Informationskriterium;
 - (iii) Bayes-Informationskriterium;
 - (iv) Dickey-Fuller-Test?