Wirtschaftswissenschaftliches Prüfungsamt

Master of Economics, Finance and Philosophy
Diplomprüfung

Econometric Methods and Applications

Wintersemester 2012/2013, Erster Prüfungstermin, 20. Februar 2013

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 vier Aufgaben. Die Aufgaben 1 und 2 sind obligatorisch. Von den Aufgaben 3 und 4 ist **genau eine** zu bearbeiten.

(d) Bei jeder Multiple-Choice Aufgabe ist **genau eine** Alternative richtig.

(e) Die Reihenfolge der Bearbeitung der Aufgaben kann beliebig gewählt werden, beginnen Sie aber für jede Aufgabe eine neue Seite.

(f) Überprüfen Sie die Vollständigkeit Ihres Klausurexemplares. Spätere Reklamationen können nicht berücksichtigt werden.

(g) Die Benutzung von zwei beidseitig beschriebenen bzw. vier einseitig beschriebenen DIN A4-Blättern sowie (auch programmierbaren) Taschenrechnern ist erlaubt.

(h) **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 [20 Punkte]

Für jede richtige Auswahl: 2 Punkte

Für jede falsche Auswahl: -1 Punkt

Unbeantwortet: 0 Punkte

(a) Für den Erwartungswert der 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

- ☐ eine positive Zahl, da $\hat{\boldsymbol{\beta}}$ mit Hilfe der quadrierten Residuen ermittelt wird;
- ☐ Null;
- ☐ eine Zahl, deren Wert davon abhängt, ob die endogene Variable vorwiegend positiv oder negativ ist;
- ☐ eine stark negative Zahl, weil nach einem Minimum gesucht wird.

(b) Wenn der Steigungskoeffizient im einfachen Regressionsmodell, $\hat{\beta}_1$, Null ist, dann gilt:

- ☐ $R^2 = \overline{Y}$;
- ☐ $0 < R^2 < 1$;
- ☐ $R^2 = 0$;
- ☐ $R^2 > SSR/TSS$.

(c) Im multiplen Regressionsmodell gilt für das adjustierte Bestimmtheitsmaß:

- ☐ Es kann nicht negativ werden;
- ☐ Es kann nicht größer als das multiple Bestimmtheitsmaß sein;
- ☐ Es kann nicht kleiner werden, wenn zusätzliche exogene Variablen zum Modell hinzugefügt werden;
- ☐ Bei Hinzunahme zusätzlicher exogener Variablen entwickelt es sich proportional zum multiplen Bestimmtheitsmaß.

(d) Wenn der Regressor $\boldsymbol{X}$ mit einer Variablen $\boldsymbol{O}$ korreliert, die im Regressionsmodell nicht berücksichtigt wurde, obwohl sie Y zumindest in Teilen determiniert, dann

- ☐ kann man den Effekt der vergessenen Variable zwar nicht messen, aber die OLS-Schätzer bleiben hiervon unberührt;
- ☐ bleiben die OLS-Schätzer unverzerrt, ihre Varianz wird durch den Ausschluss von $\boldsymbol{O}$ allerdings größer;
- ☐ sind die OLS-Schätzer verzerrt;
- ☐ sind die OLS-Schätzer immer nach unten verzerrt.

(e) Gegeben sei folgendes lineares Modell:

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

Dabei seien:

$$Y_i : \text{ Einkommen der Untersuchungsperson } i \quad ;$$

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

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

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

- ☐ $\hat{\beta}_1$ wird positiv sein, $\hat{\beta}_2$ negativ;
- ☐ $|\hat{\beta}_0 - \hat{\beta}_1| = |\hat{\beta}_0 - \hat{\beta}_2|$;
- ☐ Probleme hinsichtlich Multikollinearität;
- ☐ $\hat{\boldsymbol{\beta}}$ ist aufgrund vergessener Variablen verzerrt.

(f) Der Unterschied zwischen einem unbalancierten und einem balancierten Panel ist, dass

- ☐ nicht gleichzeitig *fixed effects* hinsichtlich der Untersuchungseinheiten und hinsichtlich der Zeit vorliegen können;
- ☐ in einem unbalancierten Panel für wenigstens eine Untersuchungseinheit die Daten nicht für alle Zeitperioden vorliegen;
- ☐ in einem balancierten Panel alle Regressoren in etwa den gleichen Einfluss auf die endogene Variable haben;
- ☐ in einem balancierten Panel alle Regressoren in etwa gleich stark mit der Störgröße korrelieren.

(g) In einem *fixed effects* Modell der Art $Y_{i,t} = \beta_0 + \beta_1 X_{i,t} + \gamma_1 D_{1,i} + \ldots + \gamma_n D_{n,i} + u_{i,t}$

- ☐ schwanken die Steigungskoeffizienten zwischen den Untersuchungseinheiten, der Achsenabschnitt ist aber konstant;
- ☐ können nicht zusätzlich *fixed effects* hinsichtlich der Zeit ergänzt werden;
- ☐ ist der OLS-Schätzer für $\hat{\boldsymbol{\beta}}$ prinzipiell verzerrt;
- ☐ besteht ein Multikollinearitätsproblem.

(h) Betrachtet werde eine Panelregression für die Arbeitslosenzahlen der G7 (Stand: 1998) Staaten für die Zeitperiode von 1976-1998 (jährliche Daten). Wie groß ist die Anzahl der zusätzlich zum Absolutglied benötigten Dummyvariablen, wenn im Modell *fixed effects* sowohl hinsichtlich der Untersuchungseinheiten als auch hinsichtlich der Zeit berücksichtigt werden sollen?

- ☐ 6;
- ☐ 27;
- ☐ 28;
- ☐ 29.

(i) Die Instrumentalvariablenschätzung verwendet Instrumente, um

- [ ] Probleme im Zusammenhang mit heteroskedastischen Störgrößen zu vermeiden;
- [ ] das Bestimmtheitsmaß des Modells zu erhöhen;
- [ ] Interaktionen zwischen erklärenden Variablen zu berücksichtigen;
- [ ] Variationen in X zu isolieren, die nicht mit der Störgröße korrelieren.

(j) In dem Modell

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 X_t + u_t \quad , \quad t = 1, \ldots, T$$

wird angenommen, dass die Störgrößen autokorreliert sind gemäß

$$u_t = \alpha u_{t-1} + \epsilon_t \quad , \quad |\alpha| < 1 \quad , \quad \epsilon_t \ iid \ N(0, \sigma^2)$$

und X_t exogen ist. Dann

- [ ] ist der LS-Schätzer $\hat{\beta}$ konsistent;
- [ ] ist Y_{t-1} mit u_t korreliert und X_t ein geeignetes Instrument für Y_{t-1};
- [ ] ist Y_{t-1} mit u_t korreliert und Y_t ein geeignetes Instrument für Y_{t-1};
- [ ] ist Y_{t-1} mit u_t korreliert und X_{t-1} ein geeignetes Instrument für Y_{t-1}.

Aufgabe 2 [**Teil 1**: 7 + 3 + 10 = 20 Punkte;
Teil 2: 4 + 16 + 6 + 4 + 7 + 7 + 6 = 50 Punkte]

Teil 1:

(a) Gegeben sei ein einfaches 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 *Bestimmheitsmaß* versteht und wie man es berechnet. Gehen Sie ferner auf den Bereich möglicher Werte ein.
Für welche Fragestellung ist das *adjustierte Bestimmtheitsmaß* vorzuziehen?
Welcher Zusammenhang besteht zwischen dem Bestimmtheitsmaß und dem Korrelationskoeffizienten im *einfachen* linearen Modell?

(b) Stimmt das Bestimmtheitsmaß einer Regression von Y auf X mit dem einer Regression von X auf Y überein? Begründen Sie Ihre Antwort.

(c) Gegeben seien die folgenden fünf alternativen Modellspezifikationen:

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

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

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

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

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

Geben Sie die marginalen Effekte von X auf Y an und interpretieren Sie diese kurz.

Teil 2:

Zahlreiche deutsche Städte erstellen sogenannte Mietspiegel, um Mietern, Vermietern, Mietberatungsstellen und Sachverständigen eine objektive Entscheidungshilfe in Mietfragen zur Verfügung zu stellen. Die Mietspiegel werden dabei insbesondere zur Ermittlung der ortsüblichen Vergleichsmiete (Nettomiete in Abhängigkeit von Wohnungsgröße, -ausstattung, -alter, etc.) herangezogen.

Bei der Erstellung von Mietspiegeln wird aus der Gesamtheit aller in Frage kommenden Wohnungen eine repräsentative Zufallsstichprobe gezogen und die interessierenden Daten werden von Interviewern anhand von Fragebögen ermittelt.
Der hier verwendete Datensatz bezieht sich auf die Stadt München (Jahr 2003) und enthält folgende Variablen.

`nmqm`	Nettomiete pro m^2
`wfl`	Wohnfläche
`rooms`	Anzahl der Zimmer
`bj`	Baujahr
`wohnschlecht`	Schlechte Wohnlage? (Ja = 1, Nein = 0)
`wohngut`	Gute Wohnlage? (Ja = 1, Nein = 0)
`wohnbest`	Beste Wohnlage? (Ja = 1, Nein = 0)
`ww0`	Warmwasserversorgung vorhanden? (Ja = 0, Nein = 1)
`zh0`	Zentralheizung vorhanden? (Ja = 0, Nein = 1)
`badkach0`	Gekacheltes Badezimmer? (Ja = 0, Nein = 1)
`badextra`	Zusatzausstattung im Bad vorhanden? (Ja = 1, Nein = 0)
`kueche`	Gehobene Küchenausstattung? (Ja = 1, Nein = 0)

Bei einer Modellschätzung mit der Statistik-Software R unter der Annahme homoskedastischer Störgrößen enstand folgender Output:

```
Call:
lm(formula = nmqm ~ wfl + rooms + bj + wohngut + wohnbest + badextra +
    kueche, data = miete03)

Residuals:
   Min     1Q Median     3Q    Max
-7.688 -1.425  0.006  1.483 10.007

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)        ???   4.026065  -8.694  < 2e-16 ***
wfl          -0.005698   0.003726  -1.529    0.126
rooms        -0.545786        ???  -5.914 3.91e-09 ***
bj            0.022800   0.002039  11.183  < 2e-16 ***
wohngut       0.922239   0.102166   9.027  < 2e-16 ***
wohnbest      1.947972   0.336010     ??? 7.78e-09 ***
badextra      0.716975   0.175253   4.091 4.46e-05 ***
kueche        1.369730   0.190666   7.184 9.45e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.195 on 2045 degrees of freedom
Multiple R-squared: 0.2112,Adjusted R-squared: ???
F-statistic: ??? on ??? and ??? DF,  p-value: < 2.2e-16
```

(a) Geben Sie die Modellformulierung an, welche dem obigen Output zugrundeliegt.

(b) Berechnen Sie die sieben fehlenden Werte im Output. (durch ??? gekennzeichnet)

(c) Welche Koeffizienten sind signifikant zu einem Signifikanzniveau $\alpha = 0.01$?
Welche Aussage können Sie über die Signifikanz des Erklärungsansatzes machen?
Wie lauten die zugehörigen Hypothesen?

(d) In der Stadt München wird zwischen schlechter, guter und bester Wohnlage differenziert.
Begründen Sie, warum die Variable `wohnschlecht` nicht in obigen Modellansatz aufgenommen wurde und geben Sie eine alternative Modellformulierung an in der alle drei Dummyvariablen vorhanden sind.
(`wohnschlecht`, `wohngut`, `wohnbest`)

(e) Treffen Sie eine Entscheidung bezüglich der Hypothese, dass sich eine gehobene Küchenausstattung stärker auf die Quadratmetermiete auswirkt als Zusatzausstattung im Bad.

$\text{Cov}(\hat{\beta}_{kueche}, \hat{\beta}_{badextra}) = -0.0023 \quad ; \quad H_1 : \beta_{kueche} - \beta_{badextra} > 0 \; ; \; \alpha = 0.05$

Verwenden Sie *eines* der folgende Quantile der Standard-Normal-Verteilung um Ihre Entscheidung zu treffen:

$$z_{0.95} = 1.645 \; ; \; z_{0.975} = 1.96 \; ; \; z_{0.99} = 2.326$$

(f) Wie können Sie den Modellansatz derart parametrisieren, dass eine Entscheidung bezüglich der Hypothese aus (e) direkt im `R`-Output abgelesen werden kann?

(g) Welche Teile des Outputs würden ihre Gültigkeit verlieren, wenn man die (restriktive) Annahme homoskedastischer Störgrößen aufgibt und fortan von einer LS-Schätzung bei Heteroskedastie ausgeht? (`Estimate`, `Std. Error`, `t value`, ...)

Wahlteil:

Aufgabe 3 [8 + 4 + 6 + 12 = 30 Punkte]

Betrachtet wird ein balanciertes Panel für 48 US-Bundesstaaten (ohne Alaska und Hawaii) für die Jahre 1982-1988 zu den folgenden Variablen:

`state`	Faktor: Indikatorvariablen für die 48 Bundesstaaten
`year`	Faktor: Indikatorvariablen für die Jahre 1982 – 1988
`unemp`	Arbeitslosenrate
`income`	Jahreseinkommen
`beertax`	Steuern pro Kasten Bier
`drinkage`	Mindestalter für legalen Alkoholkonsum
`youngdrivers`	Prozentsatz von jungen Fahrern im Alter von 15 bis 24 Jahren
`miles`	durchschnittlich gefahrene Kilometer pro Person und Jahr
`jail`	Wird bei Trunkenheit am Steuer eine Gefängnisstrafe verhängt? (Ja = 1, Nein = 0)
`service`	Müssen bei Trunkenheit am Steuer Sozialstunden geleistet werden? (Ja = 1, Nein = 0)
`fatal`	Anzahl der tödlichen Verkehrsunfälle pro 10000 Einwohner und Jahr

(a) Eine einfache Regression von `fatal` auf `beertax` liefert folgendes Ergebnis:

```
t test of coefficients:

            Estimate Std. Error t value  Pr(>|t|)
(Intercept) 1.853308   0.043567 42.5391 < 2.2e-16 ***
beertax          ???   0.062170  5.8647 1.082e-08 ***
---
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 Unterschiede zu Zeitreihen- beziehungsweise Querschnittsdaten ein.

(ii) Berechnen und interpretieren Sie den Koeffizienten von `beertax`.

(iii) In diesem Modell ist `beertax` der einzige echte Regressor. Auf welches Problem sollten Sie deshalb bei der Interpretation der Schätzergebnisse achten?

(b) Ein erweitertes Modell führt zu folgendem Ergebnis:

```
t test of coefficients:

               Estimate  Std. Error t value  Pr(>|t|)
(Intercept)   2.1240e+00  7.7054e-01  2.7565  0.006172 **
beertax              ???  5.7539e-02  1.6448  0.100978
unemp         6.4931e-03  1.2820e-02  0.5065  0.612863
drinkage     -2.5523e-03  2.8229e-02 -0.0904  0.928014
youngdrivers  6.0036e-01  1.1497e+00  0.5222  0.601887
jailyes       1.5476e-01  6.7210e-02  2.3027  0.021928 *
serviceyes    1.9235e-01  7.6612e-02  2.5107  0.012534 *
income       -1.0137e-04  1.5518e-05 -6.5320 2.493e-10 ***
miles         1.3787e-04  1.7802e-05  7.7448 1.220e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

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

(ii) Begründen Sie, weshalb es auch hier zu Problemen hinsichtlich vergessener Variablen kommen kann.

(c) Für den folgenden Output wurde das Modell um *fixed effects* hinsichtlich der Untersuchungseinheiten ergänzt[1]:

```
t test of coefficients:

                 Estimate   Std. Error     t value     Pr(>|t|)
(Intercept)   2.573631e+00 6.148766e-01  4.18560558 3.815614e-05
beertax      -4.986281e-01 1.855190e-01 -2.68774648 7.625852e-03
unemp        -3.953735e-02 1.042833e-02 -3.79133891 1.836662e-04
drinkage     -1.246578e-02 1.928733e-02 -0.64631985 5.186035e-01
youngdrivers  3.804018e+00 6.997006e-01  5.43663703 1.185887e-07
jailyes      -2.130489e-03 1.342194e-01 -0.01587318 9.873469e-01
serviceyes   -1.730157e-02 1.551231e-01 -0.11153448 9.112727e-01
income        4.016953e-05 2.197847e-05  1.82767606 6.866613e-02
miles         4.784775e-06 9.797334e-06  0.48837519 6.256677e-01
```

[1] Die 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 hinsichtlich der interessierenden Variable `beertax`?

(iii) Welches Problem bleibt weiterhin relevant und wie könnten Sie es in den Griff bekommen?
(Anm.: Denken Sie an die Relevanz zeitlicher Veränderungen.)

(d) Schließlich wurden noch *fixed effects* bezüglich der Zeit in das Modell aufgenommen.[2]

(i) Skizzieren Sie eine entsprechende Regressionsgleichung bei der sowohl *fixed effects* bezüglich der Untersuchungseinheiten als auch der Zeit berücksichtigt werden.

(ii) Wie viele Parameter müssen bei diesem Modellansatz insgesamt geschätzt werden?

(iii) Geben Sie kurz eine Möglichkeit an, wie Sie im Modellansatz aus (d) die Anzahl der zu schätzenden Parameter reduzieren können.

[2]Der Output dieser Schätzung wird für die folgenden Aufgabenteile nicht benötigt und deshalb an dieser Stelle auch nicht angegeben.

Aufgabe 4 [12 + 10 + 8 = 30 Punkte]

(a) Erklären Sie die Idee der *Instrumentalvariablenschätzung*. Gehen Sie dabei insbesondere auf die folgenden Punkte ein:

(i) Mögliche Ursachen für $E(u_i|X_i) \neq 0$.

(ii) Was ist eine Instrumentalvariable?

(iii) Wofür stehen in diesem Zusammenhang die Begriffe *endogene* und *exogene* Variable?

(iv) Was versteht man unter *Relevanz* und *Exogenität* eines Instruments Z?

(b) Gegeben sei ein einfaches lineares Modell der Art

$$Y_i = \beta_0 + \beta_1 X_i + u_i \quad \text{mit} \quad \mathrm{Cov}(X_i, u_i) = 0 \quad ,$$

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

(i) Warum ist X ein valides Instrument?

(ii) Ist X ein starkes Instrument? Wenn ja, warum?

(iii) Warum ist Y kein valides Instrument?

(iv) Begründen Sie, warum der IV-Schätzer $\hat{\beta}^{(2SLS)} = \frac{s_{ZY}}{s_{ZX}}$ in diesem Fall identisch ist mit dem OLS-Schätzer.

(c) Gegeben sei folgendes Regressionsmodell ohne Konstante:

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

Ferner existiere für jede Beobachtung i eine weitere Variable Z_i mit

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

Auf der ersten Stufe einer zweistufigen Kleinst-Quadrate-Schätzung soll das Modell

$$X_i = \gamma Z_i + e_i$$

mit dem Instrument Z geschätzt werden, also gilt:

$$\hat{\gamma} = \frac{\sum\limits_{i=1}^{T} Z_i X_i}{\sum\limits_{i=1}^{T} Z_i^2} \quad , \quad \text{mit } T = \text{Stichprobenumfang}$$

(i) Geben Sie die Instrumentalvariable $\hat{X}_i$ für X_i an.

(ii) Zeigen Sie, dass für $\hat{\beta}^{(2SLS)}$ folgende Beziehung besteht:

$$\hat{\beta}^{(2SLS)} = \frac{\sum\limits_{i=1}^{T} \hat{X}_i Y_i}{\sum\limits_{i=1}^{T} \hat{X}_i^2} = \frac{\sum\limits_{i=1}^{T} Z_i Y_i}{\sum\limits_{i=1}^{T} Z_i X_i}$$

Viel Erfolg!