Klausur: Einführung in die Ökonometrie (11036) Sommersemester 2012
Prüfer: Prof. Dr. Karl-Heinz Paqué, Dr. Bettina Büttner
Als Hilfsmittel sind zugelassen: Taschenrechner, Formelsammlung, fremdsprachiges Wörterbuch

Die Aufgabenstellung umfasst **drei** Aufgaben, die alle in der zur Verfügung stehenden Zeit von 120 Minuten zu bearbeiten sind. Sie können maximal 120 Punkte erreichen.
Wichtiger Hinweis: **Runden** Sie sowohl bei Zwischen- als auch bei Endergebnissen auf **3** Stellen nach dem Komma.

Aufgabe 1: Ein Freund zeigt Ihnen eine von ihm erstellte Studie, bei der er annimmt, dass der Preis amerikanischer Immobilien von der Größe des Grundstückes, der Wohnfläche sowie der Anzahl der Schlafräume abhängt. Ihr Freund hat dabei folgendes Modell geschätzt:

$$\log(preis_t) = \alpha + \beta_1 * \log(lotsize_t) + \beta_2 * \log(sqrft_t) + \beta_3 * bdrms_t + u_t \qquad (1)$$

wobei

$preis_t$ = durchschnittlicher Kaufpreis einer Immobilie t in 1000 $US

$lotsize_t$ = Größe des Grundstückes der Immobilie t (in Quadratmetern)

$sqrft_t$ = Wohnfläche der Immobilie t (in Quadratmetern)

$bdrms_t$ = Anzahl der Schlafräume der Immobilie t

bezeichnen.

Als Ergebnis erhält Ihr Freund die folgende Tabelle:

Source	SS	df	MS
Model	5.15504028	3	1.71834676
Residual	2.86256324	84	.034078134
Total	8.01760352	87	.092156362

Number of obs =	88
F(3, 84) =	50.42
Prob > F =	0.0000
R-squared =	0.6430
Adj R-squared =	0.6302
Root MSE =	.1846

lprice	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
llotsize	.1679667	.0382812	4.39	0.000	.0918404	.244093
lsqrft	.7002324	.0928652	7.54	0.000	.5155597	.8849051
bdrms	.0369584	.0275313	1.34	0.183	-.0177906	.0917074
_cons	-1.297042	.6512836	-1.99	0.050	-2.592191	-.001893

a) (*7 Punkte*) Hat die Anzahl der Schlafräume einen signifikanten Einfluss auf den Preis der Immobilie? Was könnten Gründe für dieses Ergebnis sein?

b) (*4 Punkte*) Sie vermuten, dass die Annahme B2 verletzt ist und Heteroskedastizität vorliegt. Wie wirkt sich diese Annahmeverletzung auf die Punkt- und Intervallschätzer sowie die Hypothesentests aus?

c) (*12 Punkte*) Mit Hilfe eines White-Testes überprüfen Sie, ob tatsächlich eine Annahmeverletzung B2 vorliegt. Das dazugehörige Hilfsmodell lautet:

$$\hat{u}_t = \delta_0 + \delta_1 x_{1t} + \delta_2 x_{2t} + \delta_3 x_{3t} + \delta_4 x_{1t}^2 + \delta_5 x_{2t}^2 + \delta_6 x_{3t}^2 + \delta_7 x_{1t} x_{2t} + \delta_8 x_{1t} x_{3t} + \delta_9 x_{2t} x_{3t} + v_t$$

Stellen Sie die Nullhypothese auf und begründen Sie, welche Werte die Parameter und das Bestimmtheitsmaß bei Gültigkeit der Nullhypothese annehmen werden! Das geschätzte R^2=0.0392. Zu welchem Ergebnis kommen Sie beim White-Test? (Hinweis: Aus Gründen der Übersichtlichkeit bezeichnet x_1 die erste exogene Variable log(*lotsize*) , x_2 die zweite exogene Variable log(*sqrft*) und x_3 die dritte exogene Variable *bdrms*.

d) (*4 Punkte*) Nennen Sie zwei plausible exogene Variablen, die in dem Modell (1) als relevante exogene Variablen fehlen könnten und begründen Sie kurz deren Einfluss auf den Kaufpreis!

e) (8 *Punkte*) Erläutern Sie die Bedeutung sowie den Unterschied zwischen dem Bestimmtheitsmaß und dem korrigierten Bestimmtheitsmaß und gehen Sie hierbei auch auf die nummerischen Werte dieser Schätzung ein.

Aufgabe 2: Sie möchten überprüfen, inwiefern die Leistung eines Schülers von den Bildungsausgaben pro Schüler, den Gehältern der Lehrer (als Maßstab für Unterrichtsqualität) sowie von der Anzahl der registrierten Schüler pro Schule abhängig ist. Dafür steht Ihnen ein Datensatz mit 408 Beobachtungen zur Verfügung. Betrachten Sie nun folgendes Modell:

$$mathe_t = \alpha + \beta_1 \log(ausgaben_t) + \beta_2 \log(gehalt_t) + \beta_3 schüler_t + u_t \qquad (5)$$

wobei:

$mathe_t$ = Anteil der Schüler mit bestandener Mathe-Abschlussprüfung (in %) in Schule t

$ausgaben_t$ = Ausgaben pro Schüler (in $US) an Schule t

$gehalt_t$ = durchschnittliches Gehalt eines Lehrers (in $US) an Schule t

$schüler_t$ = Anzahl der Schüler an Schule t

a) (*4 Punkte*) Es sei unterstellt, dass Sie den Datensatz bereits in STATA eingelesen haben. Schreiben Sie nun ein geeignetes STATA-Programm, um das angegebene Modell zu schätzen. Gehen Sie dabei davon aus, dass die Daten nur in nicht-logarithmierter Form vorliegen.

b) (*7 Punkte)* STATA liefert Ihnen folgenden Output:

Source	SS	df	MS
Model	2894.26967	3	964.756556
Residual	41922.9108	404	103.769581
Total	44817.1805	407	110.115923

Number of obs =	408
F(3, 404) =	9.30
Prob > F =	0.0000
R-squared =	0.0646
Adj R-squared =	0.0576
Root MSE =	10.187

mathe	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
logausgaben	2.136837	4.075512	0.52	0.600	-5.875022	10.1487
loggehalt	17.40375	4.48421	3.88	0.000	8.588453	26.21905
schüler	-.0002416	.0002141	-1.13	0.260	-.0006625	.0001793
_cons	-173.3405	39.03954	-4.44	0.000	-250.0865	-96.59446

Wie lautet ihr geschätztes Modell und wie interpretieren Sie die von STATA ausgegebenen Werte der Steigungsparameter $\hat{\beta}_1$ und $\hat{\beta}_3$?

c) (*3 Punkte*) Überprüfen Sie anhand des STATA-Outputs, ob die Steigungsparameter $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ signifikant sind (Signifikanzniveau 5%).

d) (*13 Punkte*) Angenommen, Sie möchten die gemeinsame Signifikanz der Variablen „$\log(ausgaben_i)$" und „$schüler_i$" mit einem geeigneten Test überprüfen (Signifikanzniveau 5%). Skizzieren Sie kurz die ökonomische Idee des Tests und beschreiben Sie die notwendigen Arbeitsschritte. Nennen Sie in diesem Zusammenhang die Nullhypothese und Alternativhypothese ihres Tests. Wie müsste der aus der Übung bekannte STATA-Befehl lauten, um diesen Test durchzuführen?

Bei der richtigen Durchführung des Tests liefert Ihnen STATA einen empirischen Wert von 0,71. Um eine Testentscheidung treffen zu können, möchten Sie sich nun noch den kritischen Wert anzeigen lassen. Wie lautet in diesem Fall der entsprechende Befehl und welchen Wert müsste der kritische Wert ungefähr annehmen? Zu welchem Ergebnis kommen Sie bei Ihrem Test?

e) (*6 Punkte*) Das STATA-Programm aus Aufgabenteil d) wird durch folgenden Programmteil ergänzen:

```
set obs 409
replace logausgaben=8 if _n==409
replace loggehalt=11 if _n==409
replace schüler=5000 if _n==409
predict mathedach if _n==409
list mathedach if _n==409

Der dazugehörige STATA-Output ist:
mathedach=33.986
```

Erläutern Sie unter Verwendung der einzelnen Programmzeilen, welche Operationen hier durchgeführt werden und interpretieren Sie anschließend den STATA-Ouput.

Aufgabe 3: Kurzfragen

a) (*9 Punkte*) Nehmen Sie an, es ereigne sich ein konstanter Messfehler von ϕ in der exogenen Variable des Modells $y_i = \alpha + \beta x_i + u_i$. Erläutern Sie anhand einer geeigneten Grafik die Größe der Verzerrung, die sich in den KQ-Schätzern der Parameter α und β ergeben!

b) (*6 Punkte*) Erläutern Sie kurz die folgenden Eigenschaften von Schätzern: Erwartungstreue, Effizienz und Konsistenz!

c) (*4 Punkte*) Stellen Sie sich vor, Sie hätten einen F-Test der Nullhypothese $H_0 : \beta = 0$ mit 40 Freiheitsgraden durchgeführt und einen F-Wert von 5,8724 erreicht. Welchem t-Wert entspricht dieser F-Wert und was wäre Ihre Testentscheidung (Signifikanzniveau von 5%) bei der Durchführung eines t-Tests?

d) (*10 Punkte*) Beschreiben Sie kurz, bei welcher Annahmeverletzung Sie eine Instrumentvariablenschätzung vornehmen würden und erläutern Sie die Vor- und

Nachteile gegenüber einer KQ-Schätzung. Erläutern Sie auch, welche Eigenschaften die Instrumentvariable erfüllen muss.

e) (*6 Punkte*) Was besagt ein berechneter Durbin-Watson-Wert von d=1,10 in einem Modell mit drei exogenen Variablen und einem Stichprobenumfang von 30?

f) (*5 Punkte*) Illustrieren Sie in einer einfachen Grafik den Unterschied zwischen der Störgröße u_t und dem Residuum $\hat{u}_t$.

g) (*6 Punkte*) Erläutern Sie die Konsequenzen von perfekter und imperfekter Multikollinearität!

h) (*6 Punkte*) Was versteht man unter einer VKQ-Methode, was unter einer GVKQ-Methode?