Klausur: Einführung in die Ökonometrie (11036)

Wintersemester 2009/2010

Prüfer: Prof. Dr. Karl-Heinz Paqué, Dr. Bettina Büttner

Als Hilfsmittel sind zugelassen: Taschenrechner, fremdsprachige Wörterbücher

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. Wichtige Hinweise: **Runden** Sie sowohl bei Zwischen- als auch bei Endergebnissen auf 3 Stellen nach dem Komma. Die Funktion log bezeichnet den natürlichen Logarithmus (ln).

Aufgabe 1: Ein Freund zeigt Ihnen eine von ihm erstellte Studie, die zu dem Ergebnis kommt, dass sich die Luftverschmutzung negativ auf den Kaufpreis eines Hauses auswirkt. Die Studie umfasst 506 Beobachtungen aus verschiedenen Regionen. Ihr Freund hat dabei folgendes Modell geschätzt:

$$\log(preis_t) = \alpha + \beta_1 \log(NOX_t) + u_t' \tag{1}$$

wobei

preis, = durchschnittlicher Kaufpreis eines Hauses in Region t in \$US

 NOX_t = Luftverschmutzung in Region t (in Stickstoffoxidmolekülen pro Mill. Luftmoleküle) Sie vermuten allerdings, dass der Kaufpreis auch durch die Kriminalität in dieser Region beeinflusst wird:

$$\log(preis_t) = \alpha + \beta_1 \log(NOX_t) + \beta_2 verbrechen_t + u_t$$
 (2)

wobei verbrechen, = Anzahl der begangenen Verbrechen pro Kopf in Region t bezeichnet.

- a) (8 Punkte) Angenommen, Modell (2) ist das korrekte Modell und Sie schätzen fälschlicherweise Model (1). Welche Konsequenzen ergeben sich für die Punktschätzer, Intervallschätzer und Hypothesentests? Beschreiben Sie zudem, welche Konsequenzen sich für die Punktschätzer, Intervallschätzer und Hypothesentests im umgekehrten Fall ergeben, d.h. wenn Modell (1) das korrekte Modell ist, jedoch Modell (2) geschätzt wird.
- b) (7 *Punkte*) Für Modell (1) ergibt sich ein Wert für das Bestimmtheitsmaß von $R^2 = 0,264$ und für Modell (2) ein Wert von $R^2 = 0,379$. Verwenden Sie diese Information, um sich anhand eines aussagekräftigen Kriteriums für eines der beiden Modelle zu entscheiden.
- c) (13 Punkte) Angenommen, es steht Ihnen ein weiteres Modell zur Verfügung. Dieses unterscheidet sich von Modell (2) darin, dass die Variable verbrechen, durch die Variable zimmer, ersetzt wurde:

$$\log(preis_t) = \alpha + \beta_1 \log(NOX_t) + \beta_3 zimmer_t + u_t''$$
(3)

wobei

zimmer, = durchschnittliche Anzahl an Zimmern in einem Haus in Region t bezeichnet

Beschreiben Sie ein geeignetes Testverfahren, das eine Auswahl zwischen Modell (2) und Modell (3) ermöglicht und diskutieren Sie die möglichen Entscheidungen!

d) (2 Punkte) Wie müssten die Testergebnisse in Aufgabenteil c) lauten, wenn Sie sich letztendlich für folgendes Modell entscheiden:

$$\log(preis_t) = \alpha + \beta_1 \log(NOX_t) + \beta_2 verbrechen + \beta_3 zimmer_t + u_t'''$$
 (4)

- e) (4 Punkte) Die Schätzung des Modells (4) in Teilaufgabe d) liefert folgende Ergebnisse für die Steigungsparameter: $\hat{\beta}_1 = -0.455$; $\hat{\beta}_2 = -0.015$; $\hat{\beta}_3 = 0.288$. Wie interpretieren Sie $\hat{\beta}_1$ und $\hat{\beta}_3$?
- f) (8 Punkte) Weshalb könnten die Variablen NOX_t und $zimmer_t$ negativ miteinander korreliert sein? Wenn dies der Fall ist, würde man im Modell (1) den Schätzer $\hat{\beta}_1$ über- oder unterschätzen?
- g) (6 Punkte) Erläutern Sie kurz die unterschiedlichen Vorgehensweisen der Steinmetzund der Maurermethodologie.

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

$$mathe_t = \alpha + \beta_1 \log(ausgaben_t) + \beta_2 \log(gehalt_t) + \beta_3 schüler_t + u_t$$
 (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 nichtlogarithmierter Form vorliegen.
- b) (7 Punkte) STATA liefert Ihnen folgenden Output:

Source	SS	df	MS		Number of obs	
Model Residual	2894.26967 41922.9108		964.756556 103.769581	R-squared =		= 0.0000 = 0.0646
Total	44817.1805	407	110.115923		Adj R-squared Root MSE	= 0.0576 = 10.187
mathe	Coef.	Std. Er	r, t	P> t	[95% Conf.	Interval]
logausgaben loggehalt schüler _cons	2.136837 17.40375 0002416 -173.3405	4.0755 4.484 .00021 39.039	21 3.88 41 -1.13	0.600 0.000 0.260 0.000	-5.875022 8.588453 0006625 -250.0865	10.1487 26.21905 .0001793 -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 $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ individuell signifikant sind (Signifikanzniveau 1 %).

d) (13 Punkte) Angenommen, Sie möchten die gemeinsame Signifikanz der Variablen "log(ausgaben_t)" und "schüler_t" 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 ergänzt:

```
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 $y_t = \alpha + \beta x_t + u_t$. Erläutern Sie anhand einer geeigneten Grafik die Größe der Verzerrungen, 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 erhalten. 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) (4 Punkte) Illustrieren Sie in einer einfachen Grafik den Unterschied zwischen der Störgröße u_t und dem Residuum \hat{u}_t .