Klausur: Einführung in die Ökonometrie (Veranstaltungs-Nr. 940, 4 GHP)

Prüfer:
Prof. Dr. Karl-Heinz Paqué
Dr. Ludwig v. Auer

Semester: SS 2000

Als Hilfsmittel sind zugelassen: nicht-programmierbarer Taschenrechner

Diese Klausur besteht aus drei Aufgaben, die in der zur Verfügung stehenden Zeit (2 Std.) zu bearbeiten sind. Runden Sie Ihre Berechnungen auf drei Stellen nach dem Komma.

Aufgabenstellung:

Aufgabe 1:

Ein in der Magdeburger Börde ansässiger landwirtschaftlicher Betrieb hat sich auf den Anbau von Weizen spezialisiert. Der jährliche Weizenertrag y_t (in dt/ha) wurde durch den jährlichen Niederschlag x_t (in mm) bestimmt:

$$y_t = \alpha + \beta x_t + u_t \ .$$

Die Auswertung der Erntejahre 1991 bis 1999 liefert die folgenden Zwischenergebnisse: $S_{xx} = 78368$, $S_{yy} = 1836$, $S_{xy} = 10001$ sowie $\overline{x} = 506,333$ und $\overline{y} = 66,667$.

a) (*10 Punkte*) Berechnen Sie die Werte der KQ-Schätzer $\widehat{\alpha}$ und $\widehat{\beta}$ und einen Schätzwert für die Varianz der Störgrößen u_t.

b) (*4 Punkte*) Wie groß ist das Bestimmtheitsmaß R^2 Ihrer Schätzung? Was sagt dieses Maß hier aus?

c) (*8 Punkte*) Sie vermuten, daß in der regenarmen Magdeburger Börde der Niederschlag einen *positiven* Einfluß auf den Weizenertrag ausübt. Versuchen Sie, diese Vermutung durch einen geeigneten t–Test zu untermauern (Signifikanzniveau 5%).

d) (*6 Punkte*) Betrachten Sie nochmals Ihr Testergebnis aus Aufgabenteil c). Fällt der entsprechende p-Wert Ihres Tests kleiner als 10%, 5% oder sogar 1% aus? Illustrieren Sie in einer einfachen Grafik wie t-Wert und p-Wert Ihres Tests zusammenhängen.

e) (*4 Punkte*) Das Erntejahr 2000 brachte Niederschläge in Höhe von 504 mm. Der Betrieb hat seinen Weizenertrag für das Erntejahr 2000 noch nicht veröffentlicht. Welchen Ertrag würden Sie erwarten?

f) (*4 Punkte*) Unterstellen Sie einmal, der Betrieb habe aus steuerlichen Gründen seinen jährlichen Weizenertrag um jeweils 5 dt/ha zu niedrig angegeben. Zeigen Sie *grafisch*, inwiefern die Aussagekraft der KQ-Schätzer $\widehat{\alpha}$ und $\widehat{\beta}$ aus Aufgabenteil a) von diesen Falschangaben betroffen wäre.

g) (*4 Punkte*) Unterstellen Sie nun, der Weizenertrag sei jeweils um 10% zu niedrig gemessen worden. Zeigen Sie *analytisch oder grafisch,* inwiefern in diesem Fall die Aussagekraft der KQ-Schätzer $\widehat{\alpha}$ und $\widehat{\beta}$ aus Aufgabenteil a) betroffen wäre.

Aufgabe 2:

Pindyck und Rubinfeld (1991) schätzen auf Basis einer Stichprobe von T=3120 Haushalten die US–amerikanische „Nachfrage nach Wohnleistungen (z.B. in Quadratmetern)“, d.h. Nachfrage nach Mietraum, Haus- und Grundstückserwerb, und ähnliches. Die Schätzung liefert die folgenden Ergebnisse:

$$\widehat{\log q_t} = 4,17 - 0,247 \log p_t + 0,960 \log m_t \ , \tag{1}$$

wobei

q_t =Nachfrage des Haushalts t nach Wohnleistungseinheiten
p_t =Preis einer Wohnleistungseinheit in der Wohngegend des Haushalts t
m_t =Einkommen des Haushalts t

a) (*4 Punkte*) Was sagt der Parameterwert $\widehat{\beta}_2 = 0,960$ der Schätzgleichung (1) aus?

b) (*6 Punkte*) Unter den Haushalten der Stichprobe befinden sich auch zahlreiche „farbige Familien“. Pindyck und Rubinfeld wollen überprüfen, ob „farbige Familien“ eine andere Nachfragefunktion nach Wohnleistungen besitzen als „nicht–farbige Familien“. Formulieren Sie ein geeignetes Modell mit dessen Hilfe eine solche Überprüfung mittels F–Test erfolgen kann. Interpretieren Sie die neu hinzugekommenen Parameter.

c) (*7 Punkte*) Wie lautet die Nullhypothese des F–Tests aus Aufgabenteil b)? Eine Schätzung des entsprechenden unrestringierten Modells ergibt $S_{\widehat{u}\widehat{u}} = 13640$ und für das restringierte Modell erhält man $S^0_{\widehat{u}\widehat{u}} = 13838$. Zu welcher Entscheidung kommen Sie in Ihrem F–Test (Signifikanzniveau 5%)?

d) (*3 Punkte*) Begründen Sie kurz, warum bei der Schätzgleichung (1) hohe Multikollinearität auftreten könnte.

e) (*8 Punkte*) „Hohe Multikollinearität hat keine Rückwirkung auf die Streuung der KQ–Schätzer $\widehat{\beta}_k$“. Kommentieren Sie diese Aussage unter Rückgriff auf die geeignete Formel der Formelsammlung. Gehen Sie dabei auch auf den Begriff der „autonomen Variation“ ein.

f) (*4 Punkte*) „Multikollinearität läßt sich dadurch beseitigen, daß Variablen mit sehr kleinen t–Werten (z.B. $t < 0,5$) aus dem ökonometrischen Modell eliminiert werden“. Begründen Sie, warum dieser Empfehlung nicht gefolgt werden sollte.

g) (*8 Punkte*) Um zu überprüfen, ob Modell (1) fehlspezifiziert ist, kann das RESET–Verfahren eingesetzt werden. Beschreiben Sie die Arbeitsschritte dieses Verfahrens. Begründen Sie kurz, warum dieses Verfahren geeignet ist, Fehlspezifikation anzuzeigen.

Aufgabe 3: Kurzfragen

a) (*4 Punkte*) Inwiefern beeinflußt das Signifikanzniveau den Intervallschätzer eines Steigungsparameters?

b) (*6 Punkte*) Erläutern Sie, inwiefern der Wert des standardisierten Bestimmtheitsmaßes sowohl von der Schätzgenauigkeit als auch von den Risiken einer verzerrten Schätzung bestimmt wird. Wie geschieht dieses beim Akaike Informationskriterium?

c) (*8 Punkte*) Wie lautet die Entscheidungsregel eines Durbin–Watson Tests der Nullhypothese $H_0 : \rho \geq 0$? Illustrieren Sie diese Regel anhand einer geeigneten Grafik.

d) (*4 Punkte*) Das Modell adaptiver Erwartungen lautet $y_t = \alpha_0 + \beta_0 x_t + \lambda y_{t-1} + u_t$, wobei $u_t = v_t - \lambda v_{t-1}$. Warum kommt für die Schätzung dieses Modells die KQ–Methode nicht in Betracht?

e) (*8 Punkte*) Um eine Konsumfunktion zu schätzen, wird im nachfolgenden RATS–Programm der logarithmierte reale Konsum c auf den realen Zins r und das logarithmierte reale Einkommen y regressiert.

```
cal 1974 1 4
all 1994:4
open data consumpt.rat
data(format=rat,org=obs) / c r y
linreg c / ud
# constant r y
```

RATS wirft folgenden verkürzten Output aus:

Variable	Coeff	Std Error	T-Stat	Signif
**********				
1. Constant	-0.085	0.111	-0.766	0.446
2. R	0.003	0.003	1.12	0.268
3. Y	0.989	0.020	50.45	0.000

Stimmen die Vorzeichen der Steigungsparameter mit dem überein, was Sie erwartet hätten? Testen Sie, ob die Einkommenselastizität der Konsumnachfrage gleich 1 ist (Signifikanzniveau 5%).

f) (*10 Punkte*) Das RATS–Programm aus Aufgabenteil e) wird durch folgenden Programmteil ergänzt:

```
set ud2 = ud**2
set r2 = r**2
set y2 = y**2
set ry = r*y
linreg ud2 /
# constant r y r2 y2 ry
compute Z = %rsquared*%nobs
cdf chisquared Z 5
```

Welcher Test wird in diesem Programmteil durchgeführt? Skizzieren Sie kurz die ökonometrische Idee des Tests und beschreiben Sie die notwendigen Arbeitsschritte anhand des Programmteils. Entscheiden Sie den Test anhand des folgenden RATS–Outputs (Signifikanzniveau 5%):

Chi-Squared(5) = 5.260384 with Significance Level 0.385.

Lösungs-Skizze

Aufgabe 1

a) Beobachtungsumfang $T = 9$. Berechnung der KQ-Schätzer:

$$\widehat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{10001}{78368} = 0,128$$
$$\widehat{\alpha} = \overline{y} - \widehat{\beta} \cdot \overline{x} = 66,667 - 0,128 \cdot 506,333 = 1,856$$

Schätzer für die Störgrößenvarianz:

$$\begin{aligned} S_{\widehat{u}\widehat{u}} &= S_{yy} - S_{\widehat{y}\widehat{y}} = S_{yy} - \widehat{\beta} S_{xy} \\ &= 1836 - 0,128 \cdot 10001 = 555,872 \\ \widehat{\sigma}^2 &= S_{\widehat{u}\widehat{u}}/T - 2 = 555,872/7 = 79,410 \end{aligned}$$

b) Bestimmtheitsmaß:

$$R^2 = \frac{S_{\widehat{y}\widehat{y}}}{S_{yy}} = \frac{S_{yy} - S_{\widehat{u}\widehat{u}}}{S_{yy}} = \frac{S_{xy}^2}{S_{xx} S_{yy}} = 0,695$$

$69,5\%$ der Variation des Ernteertrages wird durch die Werte des Niederschlags erklärt. 30,5% der Variation bleiben damit unerklärt.

c) einseitiger t-Test

1. Getestet wird die Nullhypothese: $H_0 : \beta \leq 0$ gegen die Alternativhypothese $H_1 : \beta > 0$ mit Signifikanzniveau $a = 5\%$.
2. Berechnung von $\widehat{se}(\widehat{\beta})$:
$$\widehat{se}(\widehat{\beta}) = \sqrt{\widehat{\sigma}^2 / S_{xx}} = \sqrt{79,410/78368} = 0,032$$
3. Berechnung des t-Wertes:
$$t = \frac{0,128 - 0}{0,032} = 4$$
4. Kritischer Wert:
$$t_{0,05} = 1,8946 \text{ (7 Freiheitsgrade)}$$
5. Da $t > t_{0,05}$ wird H_0 abgelehnt, d.h. der Niederschlag besitzt einen positiven signifikanten Einfluß auf den Weizenertrag.

d) Da $t_{0,01} = 2,9980$ und $t > t_{0,01}$ fällt der entsprechende p–Wert des Tests kleiner als 1% aus.

e) Die Prognose beträgt:

$$\widehat{y}_0 = \widehat{\alpha} + \widehat{\beta} x_0 = 1,856 + 0,128 \cdot 504 = 66,368$$

f) Der Niveauparameter wird im Durchschnitt um 5 dt/ha unterschätzt, der Steigungsparameter bleibt hingegen unverzerrt (vgl. Abbildung 16.1 auf Seite 255 im Buch)

g) analytisch:

$$\begin{aligned}
\text{wahrer Zusammenhang}: & \; y_t^* = \alpha + \beta x_t + u_t \\
\text{beobachtet wurde aber}: & \; y_t = y_t^* \cdot 0,9 \\
& \Rightarrow y_t/0,9 = \alpha + \beta x_t + u_t \\
& \Rightarrow y_t = 0,9\alpha + 0,9\beta x_t + 0,9u_t \\
\text{oder einfacher}: & \; y_t = \alpha' + \beta' x_t + u_t'
\end{aligned}$$

wobei $\alpha' = 0,9\alpha$, $\beta' = 0,9\beta$ und $u_t' = 0,9u_t$. Da dieses Modell die A-, B- und C-Annahmen erfüllt, werden α' und β' unverzerrt geschätzt. Das bedeutet aber, daß α und β um 10% unterschätzt werden.

Aufgabe 2

a) Steigt das Einkommen eines Haushalts um 1%, dann steigen die Ausgaben dieses Haushaltes für Wohnleistungen um 0,960% (Elastizität).

b) Allgemeines Dummy-Variablen-Modell:

$$\log q_t = \alpha + D_t\gamma + \beta_1 \log p_t + \delta_1 \log m_t + D_t\delta_2 \log m_t + u_t$$

wobei

$$\begin{aligned}
D_t &= 0, \text{ wenn Familie } t \text{ „nicht-farbig"} \\
D_t &= 1, \text{ wenn Familie } t \text{ „farbig"}
\end{aligned}$$

γ gibt die Differenz zwischen den Niveauparametern der „farbigen" und „nicht-farbigen" Familien an.

δ_1 und δ_2 geben jeweils den Unterschied zwischen den Steigungsparametern der „farbigen" und „nicht-farbigen" Familien an (δ_1 den unterschiedlichen Einfluß des Preises einer Wohnungseinheit und δ_2 den unterschiedlichen Einfluß des Einkommens auf die Nachfrage nach Wohnungseinheiten).

c)

$$\begin{aligned}
H_0 &: \gamma = \delta_1 = \delta_2 = 0 \text{ mit Signifikanzniveau } a = 5\% \\
F &= \frac{\left(S_{\widehat{u}\widehat{u}}^0 - S_{\widehat{u}\widehat{u}}\right)/L}{S_{\widehat{u}\widehat{u}}/\left(T-K-1\right)} = \frac{66}{13640/3114} = \frac{66}{4,380} = 15,068 \\
F_{0,05} &= 2,839
\end{aligned}$$

Da $F > F_{0,05}$ wird H_0 abgelehnt. Es besteht ein signifikanter Unterschied zwischen der Nachfragefunktion von „farbigen" und „nicht-farbigen" Familien.

d) Haushalte mit höherem Einkommen neigen dazu, in kostspieligere Wohnviertel zu ziehen. Dies führt zu positiver Korrelation zwischen den exogenen Variablen p_t und m_t (bzw. zwischen $\log p_t$ und $\log m_t$).

e)

$$var\left(\widehat{\beta}_1\right) = \frac{\sigma^2}{S_{11}\left(1 - R^2_{1\cdot 2}\right)}$$

Die Aussage ist größtenteils falsch. Ein hohes $R^2_{1\cdot 2}$ erhöht $var\left(\widehat{\beta}_1\right)$ und damit die Schätzgenauigkeit. Hohe Multikollinearität kann jedoch kompensiert werden durch niedrige Störgrößenvarianz (σ^2) und hohe Variation in der exogenen Variable (S_{11}). Der Nenner beschreibt die autonome Variation in x_{1t}. Sie entspricht S_{11} korrigiert um die in $(1 - R^2_{1\cdot 2})$ ausgedrückt Multikollinearität zwischen x_{1t} und x_{2t}.

f) Hohe Multikollinearität bedeutet große $\widehat{se}\left(\widehat{\beta}_k\right)$-Werte und damit tendentiell kleine t-Werte. Der kleine t-Wert ist also auf Multikollinearität und nicht auf Irrelevanz zurückzuführen. Eliminiert man relevante Variablen, so führt dies zu verzerrten Punkt- und Intervallschätzern. Ferner sind Hypothesentests wertlos.

g) RESET-Verfahren:

1. Für das Modell (1) werden die geschätzten Werte der endogenen Variable $\widehat{\log q_t}$ und die Summe der Residuenquadrate $S_{\widehat{u}\widehat{u}}$ berechnet.

2. Das Modell (1) wird um die „exogenen" Variablen $\widehat{\log q_t}^2$ und $\widehat{\log q_t}^3$ erweitert zum folgenden Modell:

$$\log q_t = \alpha + \beta_1 \log p_t + \beta_2 \log m_t + \beta_3 \widehat{\log q_t}^2 + \beta_4 \widehat{\log q_t}^3 + u_t^* \qquad (2)$$

   Das Modell (2) wird mittels KQ-Methode geschätzt, und es wird ebenfalls die Summe der Residuenquadrate $S^*_{\widehat{u}\widehat{u}}$ berechnet.

3. Für Modell (1) wird ein F-Test mit der Nullhypothese $H_0 : \beta_2 = \beta_3 = 0$ durchgeführt.

4. Berechnung des F-Wertes:

$$F_{(L,T-K^*-1)} = \frac{\left(S_{\widehat{u}\widehat{u}} - S^*_{\widehat{u}\widehat{u}}\right)/L}{S^*_{\widehat{u}\widehat{u}}/\left(T - K^* - 1\right)}$$

   K^* : Anzahl der exogenen Variablen im erweiterten Modell (2)

   L : Anzahl der in der Linearkombinationen in der Nullhypothese

5. Wenn der berechnete F-Wert größer ist als der kritische F_a-Wert, dann wird H_0 verworfen und das lineare Modell (1) ist fehlspezifiziert.

Aufgabe 3

a) Der Intervallschätzer fällt umso schmaler aus, je größer das Signifikanzniveau a.

b) Berechnung des standardisierten Bestimmtheitsmaßes:

$$\overline{R}^2 = 1 - (1 - R^2)\frac{T-1}{T-K-1}.$$

Dabei ist das Bestimmtheitsmaß R^2 ein Indikator für das Verzerrungsrisiko, da zusätzliche exogene Variablen das Verzerrungsrisiko mindern bzw. das Bestimmtheitsmaß erhöhen und somit auch $\overline{R}^2$ tendenziell steigt. Als Indikator für die Schätzvarianz dient die Zahl der Freiheitsgrade $T-K-1$. Zusätzliche exogene Variablen verringern die Zahl der Freiheitsgrade. Somit wird der Nenner kleiner, der gesamte Bruch größer und das standardisierte Beszimmtheitsmaß $\overline{R}^2$ sinkt tendenziell.

Beim Akaike Informations Kriterium ist die Summe der Residuenquadrate $S_{\widehat{u}\widehat{u}}$ der Indikator für das Verzerrungsrisiko. Die Anzahl der zu schätzenden Parameter K dient wieder als Indikator für die Schätzvarianz.

c)

Wenn $d > 4 - d^L_{0,05}$, dann lehne $H_0 : \rho \geq 0$ ab und akzeptiere $H_1 : \rho < 0$.
Wenn $d < 4 - d^H_{0,05}$, dann lehne $H_0 : \rho \geq 0$ nicht ab.
Wenn$4 - d^H_{0,05} \leq d \leq 4 - d^L_{0,05}$, dann treffe keine Entscheidung.

d) Die Störgröße u_t und die „exogene Variable" y_{t-1} sind korreliert, und zwar kontemporär. Grund: Wegen $u_t = v_t - \lambda v_{t-1}$ ist v_{t-1} in u_t und wegen $y_{t-1} = \alpha_0 + \beta x_{t-1} + \lambda y_{t-2} + v_{t-1} - \lambda v_{t-2}$ auch in y_{t-1} enthalten.

e) Für den Steigungsparameter des Zinses wird (gemäß Keynesianischer Theorie) ein negatives Vorzeichen erwartet. Die Schätzung liefert aber einen positiven Wert, der allerdings nicht signifikant ist. Der Steigungsparameter für das reale Einkommen hat erwartungsgemäß ein positives Vorzeichen (Wert ist signifikant).

t-Test der Nullhypothese $H_0 : \beta_2 = 1$ gegen die Alternativhypothese $H_1 : \beta_2 \neq 1$ mit Signifikanzniveau $a = 0,05$

$$\widehat{se}\left(\widehat{\beta}_2\right) = 0,020$$
$$t = \frac{0,989 - 1}{0,020} = -0,55$$
$$t_{0,025} \approx 2,0211 \text{ (84 Freiheitsgrade)}$$

Da $|t| < t_{0,025}$ kann die Nullhypothese nicht abgelehnt werden. Es kann also nicht ausgeschlossen werden, daß die Einkommenselastizität der Konsumnachfrage gleich 1 ist.

f) White-Test: Zuerst werden durch set-Befehle einige Variablen definiert: die Residuenquadrate durch ud2, der quadrierte Realzins durch r2, das quadrierte Realeinkommen durch y2 und das Produkt aus Realzins und Realeinkommen durch ry (Zeile 1-4).

1. Nullhypothese: Störgrößen sind homoskedastisch.

2. KQ-Schätzung des Ausgangsmodells und Abspeichern der Residuen.

3. Hilfsmodell mit Residuenquadraten ud2 als endogene Variable sowie r, y, r2, y2 und ry als exogene Variablen.

   KQ-Schätzung dieses Hilfsmodells (Zeile 5-6).

   Idee: Wenn das Bestimmtheitsmaß des Hilfsmodells nahe 0 ist, dann üben die exogenen Variablen keinen systematischen Einfluß auf die quadrierten Residuen aus und verursachen somit keine Heteroskedastizität.

4. Berechnung von $R^2 \cdot T$ (Zeile 7); $R^2 \cdot T \sim \chi^2_{(v)}$, $v =$ Anzahl der Steigungsparameter des Hilfsmodells.

   $\chi^2_{(5)} = 11,0705$

Da $\chi^2 < \chi^2_{(5)}$, kann H_0 nicht abgelehnt werden. Es kann also von Homoskedastizität ausgegangen werden.