

Tentamen Econometrie 2

Dinsdag 1 juli 2003,- 9.30 - 12.30 u , - Zaal B, Gebouw A.

Schrijf op ieder vel je naam, en op het eerste vel ook je registratienummer.

Het tentamen bestaat uit 5 opgaven. Iedere vraag is 20 punten waard en in totaal zijn er 100 punten te verdienen. De waardering staat per vraagdeel in de rechterkantlijn aangegeven.

Met 55 of meer punten heb je een voldoende voor dit onderdeel van het vak behaald. De uitslag van dit tentamen hangt op 16 juni 2003 op het uitslagenbord van de sectie econometrie van de Afdeling Kwantitatieve Economie op de 3e etage.

Eén A4 vel aan beide kanten beschreven met formules is toegestaan.

Licht antwoorden toe.

In elk van deze opgaven dien je ervan uit te gaan dat er voldaan is aan de basisveronderstellingen van het lineaire regressiemodel. Eventuele afwijkingen hierop staan vermeld in de opgave.

1. a.

i) Wat is het verschil tussen het Data Genererend Proces (DGP) en een model. (2)

ii) Wat is het verschil tussen parameters en schatters? (2)

iii) Wat bedoelen we met eigenschappen van schatters? (2)

b. Als een exogene variabele die in een model wordt opgenomen, in werkelijkheid endogeen blijkt te zijn wat is dan het gevolg voor de OLS schatter? (geef ook formules waar mogelijk) (6)

c. Leidt de Instrumentele variabelen (IV) schatter af. Laat zien dat deze consistent is en geef duidelijk de voorwaarden aan waaronder dit geldt. (8)

2. Een actuaaris vermoedt dat de tijd tot een bepaalde gebeurtenis Gamma verdeeld is met verwachting α en vorm parameter β gelijk aan 2. Zij heeft N onafhankelijke waarnemingen y_i (tijd tot de gebeurtenis voor individu i) en verklarende variabelen in een $(k \times 1)$ -vector x_i . De dichtheid van y_i gegeven x_i is

$$\text{pdf}(y_i | \alpha_i) = \exp\{ -y_i/\alpha_i \} \alpha_i^{-2} y_i$$

De verwachting α_i van y_i gegeven x_i wordt gemodelleerd als:

$$\alpha_i = E[y_i | x_i] = \exp\{ x_i \beta \}$$

a. Wat is de loglikelihood voor N onafhankelijke waarnemingen? (5)

b. Leidt de eerste orde voorwaarden voor de maximum likelihood schatter van β af en laat zien dat deze niet expliciet opgelost kunnen worden. (10)

c. Leidt de informatie matrix af. (5)

3. Een onderzoekster analyseert de vraag naar voedsel producten over tijd in de VS en de invloed van prijzen. Zij schat een aantal modellen met OLS op basis van waarnemingen voor de periode van 1959 tot en met 1994. De variabelen zijn als volgt:

Food: uitgaven aan voedsel producten in miljoenen \$/jaar,
 DispIncome: beschikbaar inkomen in miljoenen \$/jaar,
 FoodPI: voedsel prijsindex 1992 = 100,
 CPI: algemene prijsindex 1992 = 100,
 Time: tijd dummy, lineair : 1 in 1959, 36 in 1994.

Model 1

Model 2

Dependent Variable: LOG(FOOD)
 Method: Least Squares
 Sample: 1959 1994
 Included observations: 36

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	(a I)	0.605465	0.810107	0.4239
TIME	(a II)	-0.003123	-1.294114	0.2049
LOG(DISPINCOME)	(a III)	0.709915	0.081261	0.0000
LOG(FOODPI/CPI)		-0.374775	-4.276301	0.0002
R-squared	0.992986	Mean dependent var	6.112169	
Adjusted R-squared	0.992329	S.D. dependent var	0.193428	
S.E. of regression	0.016942	Akaike info criterion	-5.213654	
Sum squared resid	0.009185	Schwarz criterion	-5.037707	
Log likelihood	97.84577	F-statistic	1510.148	
Durbin-Watson stat	0.664819	Prob(F-statistic)	0.000000	

Dependent Variable: LOG(FOOD)
 Method: Least Squares
 Date: 01/25/04 Time: 17:24
 Sample: 1959 1994
 Included observations: 36

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.246659	0.867821	2.588850	0.0145
TIME	0.008364	0.004890	1.710602	0.0971
LOG(DISPINCOME)	0.518511	0.104127	4.979615	0.0000
LOG(FOODPI)	-0.245608	0.094228	-2.606531	0.0139
LOG(CPI)	0.131131	0.122567	1.069875	0.2929
R-squared	0.994270	Mean dependent var	6.112169	
Adjusted R-squared	0.993531	S.D. dependent var	0.193428	
S.E. of regression	0.015558	Akaike info criterion	-5.360292	
Sum squared resid	0.007503	Schwarz criterion	-5.140359	
Log likelihood	101.4853	F-statistic	1344.824	
Durbin-Watson stat	0.779561	Prob(F-statistic)	0.000000	

- a. Welke waarden/getallen horen in de vakken (a I) (a II) en (a III) voor Model 1? (4)
- b. Geef een economische interpretatie van Model 1. Hebben coëfficiënten het verwachte teken? (4)
- c. Wat is het economische verschil tussen Model 2 en Model 1? Hebben de coëfficiënten in Model het verwachte teken? Toets de restrictie dat de coëfficiënten van log(FoodPI) en log(CPI) gelijk zijn, maar met tegengesteld teken. (4)

De onderzoeker vermoed een fundamentele verandering na de eerste olie crisis en in 1973 is voor het eerst een reële daling in uitgaven aan de categorie voedsel. Zij doet daarom twee regressies voor de subperiodes van '59-'73 en van '74-'94.

Dependent Variable: LOG(FOOD)
 Method: Least Squares
 Sample: 1974 1994
 Included observations: 21

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.810586	1.293366	2.173079	0.0451
TIME	0.013929	0.004708	2.958755	0.0092
LOG(DISPINCOME)	0.452250	0.178120	2.539016	0.0219
LOG(FOODPI)	0.036447	0.214645	0.169799	0.8673
LOG(CPI)	-0.192872	0.228264	-0.844949	0.4106
R-squared	0.988239	Mean dependent var	6.244594	
Adjusted R-squared	0.985298	S.D. dependent var	0.105369	
S.E. of regression	0.012776	Akaike info criterion	-5.678217	
Sum squared resid	0.002612	Schwarz criterion	-5.429521	
Log likelihood	64.62128	F-statistic	336.0936	
Durbin-Watson stat	1.096311	Prob(F-statistic)	0.000000	

Dependent Variable: LOG(FOOD)
 Method: Least Squares
 Sample: 1959 1973
 Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-6.465340	3.002690	-2.153183	0.0568
TIME	-0.043229	0.017774	-2.432174	0.0353
LOG(DISPINCOME)	1.536333	0.370334	4.148505	0.0020
LOG(FOODPI)	-0.985172	0.333567	-2.953444	0.0144
LOG(CPI)	1.265421	0.453804	2.788476	0.0192
R-squared	0.988435	Mean dependent var	5.926774	
Adjusted R-squared	0.983810	S.D. dependent var	0.120597	
S.E. of regression	0.015345	Akaike info criterion	-5.254864	
Sum squared resid	0.002355	Schwarz criterion	-5.018848	
Log likelihood	44.41148	F-statistic	213.6778	
Durbin-Watson stat	1.851944	Prob(F-statistic)	0.000000	

- d. Ervan uitgaande dat aan alle klassieke veronderstelling is voldaan toets de nul-hypothese dat er geen structurele verandering heeft plaatsgevonden in 1973. (4)
- e. Geef aan op welke manier de specificatie van het model verbeterd zou kunnen worden. (4)

4. Econometrische analyse wordt soms bemoeilijkt door zelfselectie en “censuur” waardoor simple schattings technieken niet toepasselijk zijn.
- Wat is het verschil tussen censureren (*censoring*) en zelfselectie bij economische data en geef economische voorbeelden van elk. (5)
 - Leg voor beide gevallen nauwkeurig uit waarom gewone kleinste kwadraten (OLS) schatters tot misleidende resultaten kan leiden. (5)
 - Geef een redelijk gedetailleerde beschrijving van twee econometrische modellen voor het analyseren van data met (I) censurering en (II) zelfselectie. (10)

5. Beschouw het dynamische tijdreeks model

$$y_t = \alpha + \delta t + \rho y_{t-1} + \varepsilon_t, \quad t = 1, \dots, T$$

$$y_0 = 0.$$

ε_t onafhankelijk en identiek verdeeld met verwachting 0 en variantie σ^2 .

a. Leidt de verwachting en variantie af van y_t als

- $\rho = 0$.
- $\rho = 1, \delta = 0$

Wat zegt dit over de stationariteit van y_t in beide gevallen? (6)

b. Bepaal of de logaritme van de DAX, de Duitse beurs-index, een unit-root heeft:

- op basis van regressie (I).
- op basis van van de F(0,1).

Geef duidelijk aan wat de nul-hypothese is en waar in de tabel de kritieke waarde van de toets staat.

iii) Moet naar jouw mening de toets met of zonder trend uitgevoerd worden? Motiveer. (7)

(I)

Dependent Variable: D(LNDAX)
Method: Least Squares
Sample: 3/03/1998 8/01/2000
Included observations: 631

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.035894	0.033779	1.062620	0.2884
LNDAX(-1)	-0.004080	0.003908	-1.043961	0.2969
R-squared	0.001730	Mean dependent var		0.000637
Adjusted R-squared	0.000143	S.D. dependent var		0.016107
S.E. of regression	0.016106	Akaike info criterion		-5.416108
Sum squared resid	0.163161	Schwarz criterion		-5.402012
Log likelihood	1710.782	F-statistic		1.089854
Durbin-Watson stat	1.863900	Prob(F-statistic)		0.296905

(II)

Dependent Variable: D(LNDAX)
Method: Least Squares
Sample: 3/03/1998 8/01/2000
Included observations: 631

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.079860	0.044592	1.790908	0.0738
T	7.77E-06	5.15E-06	1.508250	0.1320
LNDAX(-1)	-0.010388	0.005722	-1.815651	0.0699
R-squared	0.005333	Mean dependent var		0.000637
Adjusted R-squared	0.002165	S.D. dependent var		0.016107
S.E. of regression	0.016090	Akaike info criterion		-5.416554
Sum squared resid	0.162572	Schwarz criterion		-5.395410
Log likelihood	1711.923	F-statistic		1.683440
Durbin-Watson stat	1.858895	Prob(F-statistic)		0.186571

Student-t Kritieke waarden. Elementen z zijn zo dat $P[Z < z] = 100(1-\alpha)\%$,
 b.v. als Z een t-verdeling heeft met 30 vrijheidsgraden, dan $P[Z < 2.04227] = 97.5\%$

d.o.f.	90%	95%	97.5%	99.5%
1	3.07766	6.31371	12.7062	63.656
2	1.88562	2.91999	4.30265	9.92482
3	1.63774	2.35336	3.18243	5.84089
4	1.53321	2.13185	2.77644	4.60393
5	1.47588	2.01505	2.57058	4.03212
10	1.37218	1.81246	2.22814	3.16922
30	1.31042	1.69726	2.04227	2.74999
100	1.29007	1.66023	1.98397	2.62589
∞	1.28156	1.64487	1.95999	2.57584

Critical Values for Unit Root Tests

Sample Size	K-Test			t-Test			F(0,1)- Test		
	1 %	5%	10%	1 %	5%	10%	1 %	5%	10%
CASE 1	no constant no trend								
25	-11.9	-7.3	-5.3	-2.66	-1.95	-1.60			
50	-12.9	-7.7	-5.5	-2.62	-1.95	-1.61			
100	-13.3	-7.9	-5.6	-2.60	-1.95	-1.61			
250	-13.6	-8.0	-5.7	-2.58	-1.95	-1.62			
500	-13.7	-8.0	-5.7	-2.58	-1.95	-1.62			
∞	-13.8	-8.1	-5.7	-2.58	-1.95	-1.62			
CASE 2	with constant no trend								
25	-17.2	-12.5	-10.2	-3.75	-3.00	-2.63			
50	-18.9	-13.3	-10.7	-3.58	-2.93	-2.60			
100	-19.8	-13.7	-11.0	-3.51	-2.89	-2.58			
250	-20.3	-14.0	-11.2	-3.46	-2.88	-2.57			
500	-20.5	-14.0	-11.2	-3.44	-2.87	-2.57			
∞	-20.7	-14.1	-11.3	-3.43	-2.86	-2.57			
CASE 4	with constant and trend								
25	-22.5	-17.9	-15.6	-4.38	-3.60	-3.24	10.61	7.24	5.91
50	-25.7	-19.8	-16.8	-4.15	-3.50	-3.18	9.31	6.73	5.61
100	-27.4	-20.7	-17.5	-4.04	-3.45	-3.15	8.73	6.49	5.47
250	-28.4	-21.3	-18.0	-3.99	-3.43	-3.13	8.43	6.34	5.39
500	-28.9	-21.5	-18.1	-3.98	-3.42	-3.13	8.34	6.30	5.36
∞	-29.5	-21.8	-18.3	-3.96	-3.41	-3.12	8.27	6.25	5.34

$$\Delta Y_t = \alpha_1 Y_{t-1} + \sum_{j=1}^k \gamma_j \Delta Y_{t-1} + \varepsilon_t, \quad \text{no constant, no trend}$$

$$\Delta Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \sum_{j=1}^k \gamma_j \Delta Y_{t-1} + \varepsilon_t, \quad \text{constant, no trend}$$

$$\Delta Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 t + \sum_{j=1}^k \gamma_j \Delta Y_{t-1} + \varepsilon_t, \quad \text{constant, trend}$$

$$K = T \hat{\alpha}_1,$$

$$t = \hat{\alpha}_1 / SE(\hat{\alpha}_1)$$

F-test is for $\alpha_2 = 0$ and $\alpha_1 = 0$