Tentamen Econometrie 1, 30 januari 2007, 14.00-16.00 uur

Dit tentamen duurt 2 uur! Toiletbezoek is niet toegstaan.

De uitslag komt uiterlijk na 15 werkdagen op Blackboard. Desgewenst kunt u daarna uw werk inzien bij de docent. Vul bovenaan op ieder antwoordenblad de gevraagde gegevens in. Bij dit tentamen mag u gebruik maken van één handgeschreven formuleblad op A4-formaat. U mag ook de uitgereikte tabellen gebruiken. Inlevering van uw uitwerkingen en van de uitgereikte tabellen is verplicht.

In elk van de opgaven dient u ervan uit te gaan dat er voldaan is aan de basisveronderstellingen van het lineaire regressiemodel. Eventuele afwijkingen blijken uit de opgaven. Tussen haakjes worden onder de regressiecoëfficiënten hun standaardfouten vermeld. Voer eventuele toetsen uit met een significantieniveau van **5%**.

MOTIVEER UW ANTWOORDEN!

Opgave 1 (5+5+5+20+15+15 = 65 punten)

Een onderzoeker wil een consumptiefunctie schatten op basis van gegevens van 20 huishoudens. Voor elk huishouden zijn de consumptieve bestedingen (Y) en het inkomen (X) gegeven (allebei gemeten in duizenden euros). De onderzoeker maakt met Eviews de onderstaande outputs.

Model 1:

Dependent Variable: Y
Method: Least Squares
Date: 01/29/07   Time: 13:05
Sample: 1 20
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.847052	0.703355	1.204302	0.2441
X	0.899325	0.025309	35.53360	0.0000

R-squared	0.985944	Mean dependent var	23.55500
Adjusted R-squared	0.985164	S.D. dependent var	10.78691
S.E. of regression	1.313895	Akaike info criterion	3.478509
Sum squared resid	31.07377	Schwarz criterion	3.578082
Log likelihood	-32.78509	F-statistic	1262.637
Durbin-Watson stat	2.582686	Prob(F-statistic)	0.000000

Model 2:

Dependent Variable: LOG(Y)
Method: Least Squares
Date: 01/25/07 Time: 15:55
Sample: 1 20
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.907839	0.080282	23.76412	0.0000
X	0.044597	0.002889	15.43768	0.0000

R-squared	0.929776	Mean dependent var	3.033911
Adjusted R-squared	0.925874	S.D. dependent var	0.550836
S.E. of regression	0.149971	Akaike info criterion	-0.862115
Sum squared resid	0.404841	Schwarz criterion	-0.762542
Log likelihood	10.62115	F-statistic	238.3220
Durbin-Watson stat	1.275716	Prob(F-statistic)	0.000000

Model 3:

Dependent Variable: LOG(Y)
Method: Least Squares
Date: 01/25/07 Time: 15:56
Sample: 1 20
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.075672	0.057393	1.318496	0.2039
LOG(X)	0.956186	0.018255	52.38022	0.0000

R-squared	0.993482	Mean dependent var	3.033911
Adjusted R-squared	0.993120	S.D. dependent var	0.550836
S.E. of regression	0.045689	Akaike info criterion	-3.239278
Sum squared resid	0.037575	Schwarz criterion	-3.139705
Log likelihood	34.39278	F-statistic	2743.688
Durbin-Watson stat	2.166013	Prob(F-statistic)	0.000000

Model 1 toegepast op twee deelsteekproeven:

Dependent Variable: Y
Method: Least Squares
Date: 01/29/07 Time: 20:24
Sample: 1 20 IF X<=25
Included observations: 10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.053316	0.616164	1.709474	0.1257
X	0.876016	0.037939	23.09013	0.0000

R-squared	0.985217	Mean dependent var	14.36000
Adjusted R-squared	0.983369	S.D. dependent var	5.346692
S.E. of regression	0.689519	Akaike info criterion	2.271210
Sum squared resid	3.803487	Schwarz criterion	2.331727
Log likelihood	-9.356051	F-statistic	533.1539
Durbin-Watson stat	0.103592	Prob(F-statistic)	0.000000

Dependent Variable: Y
Method: Least Squares
Date: 01/29/07 Time: 20:25
Sample: 1 20 IF X>25
Included observations: 10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.278963	3.443383	0.952250	0.3689
X	0.834637	0.096213	8.674885	0.0000

R-squared	0.903908	Mean dependent var	32.75000
Adjusted R-squared	0.891897	S.D. dependent var	5.401080
S.E. of regression	1.775825	Akaike info criterion	4.163264
Sum squared resid	25.22845	Schwarz criterion	4.223781
Log likelihood	-18.81632	F-statistic	75.25363
Durbin-Watson stat	2.298533	Prob(F-statistic)	0.000024

a. Laat zien dat het geometrisch (of meetkundig) gemiddelde van *Y*, aan te geven als *G*, gelijk is aan 20.778.

b. Toon aan dat $e^* = \frac{1}{G} e$, waarin *e* de vector van OLS-residuen van model 1 voorstelt en e^* de vector van OLS-residuen van het model $Y/G = \beta_1 + \beta_2 X + \varepsilon$.

c. Toon aan dat $e^* = e$, waarin *e* de vector van OLS-residuen van model 3 voorstelt en e^* de vector van OLS-residuen van het model $\log(Y/G) = \beta_1 + \beta_2 \log X + \varepsilon$.

d. Bepaal uw voorkeur tussen modellen 1, 2 en 3 (afgezien van de vraag of deze modellen aan de Gauss-Markov condities voldoen). Is één model significant beter dan elk van de andere?

e. Toets model 1 op parameter stabiliteit. Formuleer het model, de hypothesen ten aanzien van de model parameters, de toetsingsgrootheid, het kritieke gebied met de waarde van α en de conclusie.
f. Toets of in model 1 heteroscedasticiteit optreedt. Formuleer het model, de hypothesen ten aanzien van de model parameters, de toetsingsgrootheid, het kritieke gebied met de waarde van α en de conclusie.

Opgave 2 (10 punten)

Gegeven is het lineaire regressiemodel in zogenaamde rijnotatie

$$y_i = x_i^{'}\beta + \varepsilon_i \quad (i = 1,\dots,n)$$

met afhankelijke variabele y_i, storingsterm ε_i, k-dimensionale vector van onafhankelijke variabelen x_i en k-dimensionale vector van regressiecoëfficiënten β. Leid door het principe der kleinste kwadraten toe te passen op het gegeven model in rijnotaie de formule van de LS-schatter in rijnotatie $b = \left(\sum_{i=1}^{n} x_i x_i^{'}\right)^{-1}\left(\sum_{i=1}^{n} x_i y_i\right)$ af, zonder gebruik te maken van enig symbool dat een matrix voorstelt.

Opgave 3 (4+4+4+13=25 punten)

Beschouw het lineaire regressie model $y = X\beta + \varepsilon$. Hierin is X een deterministische $n \times k$ matrix, y en ε zijn $n \times 1$ en β is $k \times 1$. De OLS-schatter van β wordt weergegeven door b en de OLS-residuenvector wordt weergegeven door $e = (e_1,\dots,e_n)'$. Ga er vanuit dat de standaard veronderstellingen gelden, behalve die van de sferische storingen. In plaats daarvan wordt verondersteld dat $V[\varepsilon] = \sigma^2\Omega$, waarin Ω een *onbekende* diagonaalmatrix is met hoofddiagonaalelementen $(\omega_1,\dots,\omega_n)$ zodanig dat $\omega_1 + \dots + \omega_n = 1$. Verder wordt aangenomen dat $\lim \frac{1}{n} X'X = C$ en $\lim \frac{1}{n}\sigma^2 X'\Omega X = D$, waarin C en D symmetrische positief definiete matrices zijn.

a. Toon aan dat de OLS-schatter b zuiver is.
b. Toon aan dat $V[b] = (X'X)^{-1}(\sigma^2 X'\Omega X)(X'X)^{-1}$.
c. Toon aan dat de OLS-schatter b consistent is.
d. Stel dat de restrictie $H_0 : R\beta = q$ moet worden getoetst, waarbij R een bekende $r\times k$ matrix is en q een bekende r dimensionale vector. Vanwege de heteroscedasticiteit is de op OLS gebaseerde F-toets ongeldig. Leidt een toetsingsgrootheid af die wel toepasbaar is, met de asymptotische verdeling daarvan (onder H_0). Hierbij mag u uitgaan van asymptotische normaliteit van b, d.w.z. $\sqrt{n}(b-\beta) \xrightarrow{d} N(0, C^{-1}DC^{-1})$, en u kunt gebruik maken van de consistente “White-schatter” $S = \frac{1}{n}\sum_{i=1}^{n} e_i^2 x_i x_i^{'}$ van de matrix D, waarin $x_i^{'}$ de i-de rij van X is.

Uitwerking tentamen Econometrie 1, 30 januari 2007

Opgave 1

a. $\log G = \log\left(Y_1 \times ... \times Y_n\right)^{1/n} = \dfrac{1}{n}\sum \log Y_i = \overline{\log Y} = 3.033911 \Rightarrow G = e^{3.033911} = 20.778$.

b. Definieer $y = \begin{bmatrix} Y_1 \\ ... \\ Y_n \end{bmatrix}, X = \begin{bmatrix} 1 & X_1 \\ ... & ... \\ 1 & X_n \end{bmatrix}, M = I - X(X'X)^{-1}X'$. Dan geldt $e^* = M\left(\dfrac{1}{G}y\right) = \dfrac{1}{G}My = \dfrac{1}{G}e$.

c. Definieer $y = \begin{bmatrix} \log Y_1 \\ ... \\ \log Y_n \end{bmatrix}, X = \begin{bmatrix} 1 & \log X_1 \\ ... & ... \\ 1 & \log X_n \end{bmatrix}, M = I - X(X'X)^{-1}X', \iota = \begin{bmatrix} 1 \\ ... \\ 1 \end{bmatrix}$. Dan geldt

$e^* = M\left(y - [\log G]\iota\right) = My - [\log G]M\iota = e - 0 = e$.

d. Model 1 is niet zomaar vergelijkbaar met modellen 2 en 3, omdat de afhankelijke variabele verschillend is. Vergelijken we model 2 met model 3 dan geven we de voorkeur aan model 3, vanwege de hogere R^2. Om model 3 met model 1 te kunnen vergelijken passen we de Zarembka schaling toe. Dit houdt in dat Y geschaald wordt tot Y/G, dus dan krijgen we de residuenvectoren die in onderdelen b en c met e^* zijn aangeduid. We krijgen nu met het geschaalde model 1 de residuele kwadratensom $e^{*\prime}e^* = \dfrac{1}{G^2}e'e = \dfrac{1}{20.778^2}31.07377 = 0.071976$ en met het geschaalde model 3 krijgen we $e^{*\prime}e^* = e'e = 0.037575$. De laatste is het kleinste en daarom gaat de voorkeur uit naar model 3. De toetsingsgrootheid van de Box Cox test is $\chi^2 = \dfrac{n}{2}\log\dfrac{RSS_1^*}{RSS_3^*} = \dfrac{20}{2}\log\dfrac{0.071976}{0.037575} = 6.50$ en deze heeft een chi-kwadraat verdeling met 1 vrijheidsgraad. De uitkomst ligt in het 5%-kritieke gebied $\chi^2 \geq$ 3.841, dus model 3 is significant beter dan model 1. Uiteraard is model 3 ook significant beter dan model 2, want bij die toets krijgen we in de teller een nog grotere residuele kwadratensom, nl. 0.404841.

e. Chow toets. Model: $Y_i = \beta_1 + \beta_2 X_i + \beta_3 D_i + \beta_4 D_i X_i + \varepsilon$, waarin $D_i = 1$ als $X_i>25$ en $D_i = 0$ elders. $H_0 : (\beta_3, \beta_4) = (0,0)$ tegen $H_1 : (\beta_3, \beta_4) \neq (0,0)$. De toetsingsgrootheid is $F = \dfrac{(RSS - RSS_1 - RSS_2)}{(RSS_1 + RSS_2)/(n-2k)} = \dfrac{(31.07377 - 3.803487 - 25.22845)/2}{(3.803487 + 25.22845)/(20-4)} = 0.563$ met een F-verdeling met (2,16) vrijheidsgraden. Het kritieke gebied bij 5% is $F \geq 3.63$. De uitkomst ligt hier niet in, dus er is geen significante parameter instabiliteit.

f. Goldfeld-Quandt toets. Model: $Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$. Veronderstel $V(\varepsilon_i) = \sigma_1^2$ (constant) als $X_i \leq 25$ en $V(\varepsilon_i) = \sigma_2^2$ (constant) als $X_i > 25$. $H_0 : \sigma_1^2 = \sigma_2^2$ tegen $H_0 : \sigma_1^2 \neq \sigma_2^2$. De toetsingsgrootheid is $F = \dfrac{RSS_2/(n_2 - k)}{RSS_1/(n_1 - k)} = \dfrac{25.22845/(10-2)}{3.803487/(10-2)} = 6.633$ met een F-verdeling met (8,8) vrijheidsgraden. Het kritieke gebied bij 10% is $F \geq 3.44$. De uitkomst ligt hier in, dus er is significante heteroscedasticiteit.

Opgave 2

We moeten $RSS = \sum\left(y_i - x_i^{'}b\right)^2$ minimaliseren m.b.t. *b*.

De 1[e] orde voorwaarden zijn $\frac{\partial RSS}{\partial b_j} = \sum 2\left(y_i - x_i^{'}b\right)\left(-x_{ij}\right) = -2\sum\left(y_i - x_i^{'}b\right)x_{ij} = 0$ voor *j*=1,...,*k*.

In één kolom samennemen geeft $\frac{\partial RSS}{\partial b} = -2\sum\left(y_i - x_i^{'}b\right)x_i = 0$.

Dit leidt tot $\sum x_i y_i - \sum x_i x_i^{'}b = 0$ en dus $b = \left(\sum x_i x_i^{'}\right)^{-1}\left(\sum x_i y_i\right)$.

Opgave 3

a. $b = (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + \varepsilon) = \beta + (X'X)^{-1}X'\varepsilon$. $E[b] = \beta + (X'X)^{-1}X'E[\varepsilon] = \beta$.

b. $V[b] = E[(b-\beta)(b-\beta)'] = E[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}] = (X'X)^{-1}X'E[\varepsilon\varepsilon']X(X'X)^{-1}$
$= (X'X)^{-1}(\sigma^2 X'\Omega X)(X'X)^{-1}$.

c. $E[b] = \beta$ en $\lim V[b] = \lim \frac{1}{n}(\frac{1}{n}X'X)^{-1}(\frac{1}{n}\sigma^2 X'\Omega X)(\frac{1}{n}X'X)^{-1} = 0 \times C^{-1}DC^{-1} = 0$.
Dit impliceert dat $p\lim b = \beta$.

d. Uit het gegeven $\sqrt{n}(b-\beta) \xrightarrow{d} N(0, C^{-1}DC^{-1})$ volgt $\sqrt{n}(Rb - R\beta) \xrightarrow{d} N(0, RC^{-1}DC^{-1}R')$.
Omdat *D* symmetrisch en positief definiet is, kunnen we schrijven $D = H'H$ waarin *H* een reguliere $n \times n$ matrix is. We kunnen nu ook schrijven $RC^{-1}DC^{-1}R' = RC^{-1}HH'C^{-1}R' = (RC^{-1}H)(RC^{-1}H)'$, dus deze matrix is ook weer symmetrisch en postief definiet.
Derhalve geldt $n(Rb - R\beta)'\left[RC^{-1}DC^{-1}R'\right]^{-1}(Rb - R\beta) \xrightarrow{d} \chi^2[k]$.
Dankzij $p\lim S = D$ en $\lim \frac{1}{n}X'X = C$, kunnen we de matrices *C* en *D* vervangen en krijgen we

$$n(Rb - R\beta)'\left[R\left(\frac{1}{n}X'X\right)^{-1} S\left(\frac{1}{n}X'X\right)^{-1} R'\right]^{-1}(Rb - R\beta) \xrightarrow{d} \chi^2[k].$$

Onder $H_0 : R\beta = q$ wordt dit $(Rb - q)'\left[nR\left(X'X\right)^{-1} S\left(X'X\right)^{-1} R'\right]^{-1}(Rb - q) \xrightarrow{d} \chi^2[k]$.