Tentamen Econometrie 1, 4 juli 2006, 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 (50 punten)

Op basis van jaarlijkse data over de periode 1970-1999 wordt een model geschat, waarbij de volgende variabelen zijn gedefinieerd:

P = consumenten prijsindex,
PB = benzineprijs,
Y = log(uitgaven aan benzine / PB),
X2 = log(nominaal beschikbaar inkomen / P),
X3 = log(PB / P),
X4 = log(prijsindex van openbaar vervoer / P),
X5 = log(prijsindex van nieuwe auto's / P),
X6 = log(prijsindex van tweede hands auto's / P),

Een lineair model wordt geschat en enkele aanvullende outputs worden gemaakt (zie onder).

a. Interpreteer en toets de coëfficiënten van X2 en X3.
b. Onderzoek multicollineariteit en bereken daarbij onder meer de "Variance Inflation Factors".
c. Bepaal een 95% betrouwbaarheidsellipsoïde voor (β_2, β_3), d.w.z. de coëfficiënten van X2 en X3. Ligt het punt (1, –0.5) daarbinnen? Wat betekent dat?
d. Stel dat X4 een overbodige verklarende variabele is. Welke consequenties heeft dit voor de kwaliteit van de LS-schatters van de coëfficiënten? Onderbouw uw antwoord.
e. Is het aannemelijk dat het model juist wel of juist niet onderhevig is aan heteroscedasticiteit gerelateerd aan de schaalgrootte (het inkomen)?
f. Omschrijf een situatie waarin het zinvol is een Chow toets op dit model toe te passen en geef aan welke Chow toets u in dat geval zou toepassen.

Model:

Dependent Variable: Y
Method: Least Squares
Date: 07/04/06 Time: 12:15
Sample: 1970 1999
Included observations: 30

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.992342	0.552295	7.228642	0.0000
X2	0.866051	0.163916	5.283505	0.0000
X3	-0.439604	0.062818	-6.998054	0.0000
X4	0.033513	0.078636	0.426181	0.6738
X5	0.390288	0.247657	1.575924	0.1281
X6	-0.094666	0.059347	-1.595125	0.1238

R-squared	0.989532	Mean dependent var	6.921244
Adjusted R-squared	0.987352	S.D. dependent var	0.203019
S.E. of regression	0.022832	Akaike info criterion	-4.544418
Sum squared resid	0.012512	Schwarz criterion	-4.264178
Log likelihood	74.16626	F-statistic	453.7607
Durbin-Watson stat	1.208064	Prob(F-statistic)	0.000000

Geschatte covariantie matrix van de coëfficiëntenschatters:

	C	X2	X3	X4	X5	X6
C	0.305030	-0.090523	-0.030547	-0.017898	-0.133080	0.006845
X2	-0.090523	0.026868	0.009070	0.005281	0.039506	-0.001942
X3	-0.030547	0.009070	0.003946	0.003188	0.013921	-0.000920
X4	-0.017898	0.005281	0.003188	0.006184	0.010608	-0.001686
X5	-0.133080	0.039506	0.013921	0.010608	0.061334	-0.002208
X6	0.006845	-0.001942	-0.000920	-0.001686	-0.002208	0.003522

Correlatie matrix:

	Y	X2	X3	X4	X5	X6
Y	1.000000	0.892179	-0.852616	0.828438	-0.744759	0.491734
X2	0.892179	1.000000	-0.544652	0.843419	-0.958944	0.630719
X3	-0.852616	-0.544652	1.000000	-0.625851	0.319910	-0.258033
X4	0.828438	0.843419	-0.625851	1.000000	-0.800456	0.641193
X5	-0.744759	-0.958944	0.319910	-0.800456	1.000000	-0.636252
X6	0.491734	0.630719	-0.258033	0.641193	-0.636252	1.000000

R-kwadraat bij regressie van een verklarende variabele op de overige verklarende variabelen:

X2	X3	X4	X5	X6
0.985582	0.886815	0.844633	0.984447	0.487287

Opgave 2 (20 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 eerste kolom van de matrix X bestaat uit louter énen. Definieer $\iota = (1 \quad ... \quad 1)'$. De LS-schatter van β wordt weergegeven door b en de LS-residuenvector wordt weergegeven door $e = (e_1,...,e_n)$.

Bewijs het volgende:

a. $\overline{e} = 0$ (Hint: Begin bij de meetkundige interpretatie).
b. $e_i = \varepsilon_i - x_i'(b - \beta)$.
c. $e_i = (\varepsilon_i - \overline{\varepsilon}) - (x_i - \overline{x})'(b - \beta)$, waarin x_i' de i-de rij van X is en $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.
d. $p\lim(e_i - \varepsilon_i) = 0$. Neem hierbij aan dat x_i een vector van constanten is.

Opgave 3 (30 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 eerste kolom van de matrix X bestaat uit louter énen. Definieer $\iota = (1 \quad ... \quad 1)'$.

Partitioneer $X = \begin{bmatrix} \iota & X_B \end{bmatrix}$ en soortgelijk $\beta = \begin{pmatrix} \beta_A \\ \beta_B \end{pmatrix}$.

Neem aan dat de standaard veronderstellingen gelden, behalve dat $E[\varepsilon_i] = \mu_i$ in de plaats van de standaard veronderstelling $E[\varepsilon_i] = 0$ komt. Een tilde (~) boven een variabele of matrix betekent dat deze variabele respectievelijk de kolommen van de matrix in afwijking van het gemiddelde zijn gebracht.

a. Bewijs dat $b_B = \beta_B + (\tilde{X}_B'\tilde{X}_B)^{-1}\tilde{X}_B'\tilde{\varepsilon}$.
b. Veronderstel dat $\mu_i = c$ (dus constant). Bewijs dat de LS-schatter b_B van β_B in dit geval zuiver is.
c. Veronderstel dat k=2 en $\mu_i = 3x_{i2}$, waarin x_{i2} element (i,2) van de matrix X voorstelt. Leid voor dit geval de vertekening van b_2 af.
d. Kan de standaard veronderstelling $E[\varepsilon_i] = 0$ worden gecontroleerd door naar een grafiek van de LS-residuen e_i te kijken? Beschouw in het bijzonder de gevallen onder onderdelen b en c.
e. Veronderstel dat $E[\varepsilon_i] = 0$. Bewijs dat $V(b_B) = \sigma^2(\tilde{X}_B'\tilde{X}_B)^{-1}$. Bewijs tevens dat voor k=2 geldt $V(b_2) = \dfrac{\sigma^2}{nVar(x_2)}$ (hierin stelt x_2 de verklarende variabele voor).

Uitwerking tentamen Econometrie 1, 4 juli 2006

Opgave 1

a. X2: De inkomenselasticiteit van de benzineconsumptie wordt geschat op 0.866. Deze elasticiteit wijkt significant af van 0 (p-waarde = 0.000 < 0.05).
X3: De prijselasticiteit van de benzineconsumptie wordt geschat op -0.440. Deze elasticiteit wijkt significant af van 0 (p-waarde = 0.000 < 0.05).

b. Oorzaak van multicollineariteit:

Verklarende variabele	X2	X3	X4	X5	X6
R-kwadraat (is gegeven)	0.985582	0.886815	0.844633	0.984447	0.487287
Variance Inflation Factor (VIF)	69.357747	8.835093	6.436373	64.29628	1.950409

De VIF is erg groot voor X2 en X5 en ook aanzienlijk voor X3 en X4.
Kijken we naar de correlaties tussen de verklarende variabelen, dan zien we een zeer hoge correlatie tussen X2 en X5 en ook vrij hoge correlaties tussen X2 en X4 en tussen X5 en X4.
Dit alles wijst erop dat X2 en X5 bijzonder sterk onderhevig zijn aan multicollineariteit en X3 en X4 in wat mindere mate, maar ook nog altijd aanzienlijk.
Gevolgen van multicollineariteit:
De coëfficiënten van X2 en X5 hebben grote standaardfouten. Dat de coëfficiënt van X2 toch nog significant is te danken aan de hoge waarde van de coëfficiënt. De coëfficiënten van X3 en X4 (en ook X6) hebben kleinere standaardfouten, maar ze zijn ook weer niet heel klein. Hier zijn de gevolgen van multicollineariteit minder groot. Desondanks is de coëfficiënt van X4 (en die van X6) niet significant.

c. 95%-betrouwbaarheidsellipsoïde voor $R\beta = \begin{bmatrix} \beta_2 \\ \beta_3 \end{bmatrix}$ met $R = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$:

$(Rb - R\beta)'[R\{s^2(X'X)^{-1}\}R']^{-1}(Rb - R\beta)/J \le F_\alpha[J, n-k]$, waarin

$Rb = \begin{bmatrix} 0.866051 \\ -0.439604 \end{bmatrix}$ (zie output), $R\{s^2(X'X)^{-1}\}R' = \begin{bmatrix} 0.026868 & 0.009070 \\ 0.009070 & 0.003946 \end{bmatrix}$ (zie geschatte covariantie matrix), J=2, $F_\alpha[J, n-k] = F_{0.05}[2,24] = 3.40$. Dit wordt dan

$$\begin{bmatrix} 0.866051 - \beta_2 \\ -0.439604 - \beta_3 \end{bmatrix}' \frac{1}{0.000023756} \begin{bmatrix} 0.003946 & -0.009070 \\ -0.009070 & 0.026868 \end{bmatrix} \begin{bmatrix} 0.866051 - \beta_2 \\ -0.439604 - \beta_3 \end{bmatrix} / 2 \le 3.40 \text{ oftewel}$$

$$\begin{bmatrix} 0.866051 - \beta_2 \\ -0.439604 - \beta_3 \end{bmatrix}' \begin{bmatrix} 24.427 & -56.146 \\ -56.146 & 166.322 \end{bmatrix} \begin{bmatrix} 0.866051 - \beta_2 \\ -0.439604 - \beta_3 \end{bmatrix} \le 1 \text{ oftewel}$$

$$24.427(0.866051 - \beta_2)^2 + 166.322(-0.439604 - \beta_3)^2 - 112.292(0.866051 - \beta_2)(-0.439604 - \beta_3) \le 1$$

Dit is een ellips met middelpunt (0.866051, –0.439604).
Invullen van (1, –0.5) geeft 1.95 aan de linkerzijde dus dit voldoet niet. Dit betekent dat het punt $(\beta_2, \beta_3) = (1, -0.5)$ niet betrouwbaar is en kan worden verworpen bij een toets met $\alpha = 5\%$.

d. De LS-schatters blijven zuiver en consistent, want het model is wel geldig (met $\beta_4 = 0$ en daarom niet efficiënt geformuleerd). Daardoor blijven de afleidingen $E[b] = \ldots = \beta$ en $p\lim b = \ldots = \beta$ gewoon geldig. Omdat er een efficiëntere formulering bestaat, nl. zonder X4, is er een alternatieve schatter, nl. de LS-schatter op basis van het model zonder X4, zeg *b**. Nu is $V[b] - V[b^*] = \sigma^2(X'X)^{-1} - \sigma^2(X^{*\prime}X^*)^{-1}$, waarin *X** gelijk is aan de regressorenmatrix *X* na verwijdering van X4, positief semi definiet en dit betekent dat *b** efficiënter is dan *b*.

e. De afhankelijke variabele is de logarithme van de consumptie en dit is al te beschouwen als een transformatie die heteroscedasticiteit tegen gaat. Daarom is het aannemelijk dat het model niet meer onderhevig is aan heteroscedasticiteit (maar het kan ook niet geheel worden uitgesloten).

f. Als men vermoedt dat er sprake kan zijn van een structurele breuk binnen de tijdreeks, waardoor er in feite twee deelsteekproeven zijn waarop verschillende parameterwaarden van toepassing zijn, dan is het zinnig een Chow toets uit te voeren. Bijv. als vanaf een bepaald jaar nieuwe wetgeving is ingegaan die de mobiliteit beïnvloedt. Als beide deelsteekproeven minstens 7 waarnemingen bevatten, dan kan de Chow toets op structurele verandering toegepast worden en anders de Chow voorspeltoets.

Opgave 2

a. e is het orthogonale complement bij projectie op de kolommenruimte van X, dus $X'e = 0$. Omdat de eerste kolom van X uit enen bestaat, volgt dat $\sum e_i = 0,$ dus $\overline{e} = 0$.

b. $e_i = y - x_i'b = x_i'\beta + \varepsilon_i - x_i'b = \varepsilon_i - x_i'(b-\beta)$

c. Uit a en b volgt $\overline{e} = \overline{\varepsilon} - \overline{x}'(b-\beta) = 0$, zodat
$e_i = e_i - \overline{e} = \varepsilon_i - x_i'(b-\beta) - \overline{\varepsilon} + \overline{x}'(b-\beta) = (\varepsilon_i - \overline{\varepsilon}) - (x_i - \overline{x})'(b-\beta)$

d. Gebruik makend van onderdeel b, de regel van Slutzky en consistentie van b volgt:
$p\lim[e_i - \varepsilon_i] = p\lim[-x_i(b-\beta)] = -x_i(p\lim b - \beta) = -x_i(\beta - \beta) = 0$.

Opgave 3

a. Volgens de stelling van Frisch-Waugh, waarbij de variabelen in afwijking van het gemiddelde worden gebracht met $M_\iota = I_n - \iota(\iota'\iota)^{-1}\iota' = I_n - n^{-1}\iota\iota'$, geldt

$$b_B = (X_B'M_\iota X_B)^{-1}X_B'M_\iota y = (X_B'M_\iota X_B)^{-1}X_B'M_\iota(X\beta + \varepsilon) = (X_B'M_\iota X_B)^{-1}X_B'(M_\iota X\beta + M_\iota\varepsilon) =$$
$$= (X_B'M_\iota X_B)^{-1}X_B'(M_\iota X_B\beta_B + M_\iota\varepsilon) = \beta_B + (X_B'M_\iota X_B)^{-1}X_B'M_\iota\varepsilon = \beta_B + (\tilde{X}_B'\tilde{X}_B)^{-1}\tilde{X}_B'\tilde{\varepsilon}.$$

b. $E[b_B] = E[\beta_B + (\tilde{X}_B'\tilde{X}_B)^{-1}\tilde{X}_B'M_\iota\varepsilon = \beta_B + (\tilde{X}_B'\tilde{X}_B)^{-1}\tilde{X}_B'M_\iota E[\varepsilon] = \beta_B + (\tilde{X}_B'\tilde{X}_B)^{-1}\tilde{X}_B'M_\iota\iota c = \beta_B$

c. $E[b_B] = \beta_B + (\tilde{X}_B'\tilde{X}_B)^{-1}\tilde{X}_B'M_\iota E[\varepsilon] = \beta_B + (\tilde{x}_2'\tilde{x}_2)^{-1}\tilde{x}_2'M_\iota 3x_2 = \beta_B + 3(\tilde{x}_2'\tilde{x}_2)^{-1}\tilde{x}_2'\tilde{x}_2 = \beta_B + 3$. De vertekening van b_2 is dus gelijk aan 3.

d. Nee, de standaard veronderstelling $E[\varepsilon_i] = 0$ kan niet worden gecontroleerd door naar een grafiek van de LS-residuen e_i te kijken. De reden is dat eventuele afwijkingen van de standaard veronderstelling zoveel mogelijk door het model worden opgepakt. In geval b zal de constante term zich aanpassen en in geval c zal de coëfficiënt van de verklarende variabele zich aanpassen. Indien $E[\varepsilon_i] = \mu_i$ andere componenten bevat (dus afgezien van een constante en componenten die samenhangen met de verklarende variabelen), dan zullen deze componenten meestal onzichtbaar blijven voor de onderzoeker. Alleen als daar een bepaalde systematiek in zit (bijv. een onderscheid tussen deelsteekproeven), dan zou deze ontdekt kunnen worden.

e. Uit onderdeel a volgt:
$$V[b_B] = E[(\tilde{X}_B'\tilde{X}_B)^{-1}\tilde{X}_B'\tilde{\varepsilon}\tilde{\varepsilon}'\tilde{X}_B(\tilde{X}_B'\tilde{X}_B)^{-1}] = (\tilde{X}_B'\tilde{X}_B)^{-1}\tilde{X}_B'M_\iota E[\varepsilon\varepsilon']M_\iota\tilde{X}_B(\tilde{X}_B'\tilde{X}_B)^{-1} =$$
$$\sigma^2(\tilde{X}_B'\tilde{X}_B)^{-1}\tilde{X}_B'M_\iota M_\iota\tilde{X}_B(\tilde{X}_B'\tilde{X}_B)^{-1} = \sigma^2(\tilde{X}_B'\tilde{X}_B)^{-1}$$

Voor k=2 volgt verder: $V(b_2) = \dfrac{\sigma^2}{\sum(x_{i2} - \overline{x}_2)^2} = \dfrac{\sigma^2}{nVar(x_2)}$.