Tentamen Econometrie 1, 23 juni 2003, 9.30-11.30 uur

De uitslag komt uiterlijk vanaf 11 juli 2003 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 een zelfgemaakt formuleblad op A4-formaat. U mag ook de uitgereikte tabellen gebruiken (graag na afloop weer inleveren, samen met uw uitwerkingen).

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!   DIT TENTAMEN DUURT 2 UUR!

Opgave 1 (40 = 8 + 7 + 7 + 5 + 8 + 5)

Een onderzoeker heeft 15 jaarlijkse waarnemingen (1946-1960) van de omzet van een bedrijf (S), het advertentiebudget (A) en het nationale inkomen (X). Op basis van deze data worden de volgende schattingsresultaten verkregen:

$$\underset{}{\hat{S}_t = \underset{(143.0)}{34.3} + \underset{(0.171)}{1.678} A_t} \qquad R^2 = 0.881$$

$$\hat{S}_t = \underset{(106.43)}{2365.1} - \underset{(0.412)}{3.744} X_t \qquad R^2 = 0.864$$

$$\hat{S}_t = \underset{(295.7)}{1113.8} + \underset{(0.220)}{0.958} A_t - \underset{(0.496)}{1.923} X_t \qquad R^2 = 0.947$$

Tevens werden de volgende correlatiecoëfficiënten berekend:

	S	A	X
S	1.000000	0.938868	-0.929659
A	0.938868	1.000000	-0.843225
X	-0.929659	-0.843225	1.000000

a) Aan welk model geeft u de voorkeur en waarom?
b) Leg uit (aan de hand van formules) waarom de standaardfout van de coëfficiënt van A_t toeneemt, wanneer X_t wordt toegevoegd als verklarende variabele.
c) Leg uit (aan de hand van formules) waarom de coëfficiënt van X_t naar beneden gaat, wanneer A_t wordt geschrapt als verklarende variabele.

d) Leg uit waarom de R^2 bij het eerste en tweede model maar weinig lager is dan in het derde model. (Een heel precies antwoord is vereist! Hint: kijk naar de waarden van de coëfficiënten.)

Met de residuen van het derde model (e_t) werd de volgende hulpregressie uitgevoerd:

$$\widehat{|e_t|} = -47.97 + 0.325A_t \qquad R^2 = 0.326$$
$$(65.2) \quad (0.0649)$$

e) Toets het derde model op heteroscedasticiteit. Wat zegt de uitkomst van de toets over de kwaliteit van de OLS-schatters van het derde model en over de standaardfouten van de OLS-schatters?

f) Ga uit van het derde model en veronderstel dat de standaarddeviatie van de storingen proportioneel is met A_t. Formuleer nu een model, zodanig dat de OLS-schatters van dit model "BLUE" zijn.

Opgave 2 (50 = 3 + 12 + 4 + 15 + 10 + 6)

Gegeven is het model $y_i = b_1 + b_2 x_{i2} + b_3 x_{i3} + b_4 x_{i4} + e_i$ met de waarnemingen

$$y = \begin{bmatrix} 4 \\ 5 \\ 7 \\ 5 \\ 5 \end{bmatrix}, \; X = \begin{bmatrix} 1 & 4 & 1 & 0 \\ 1 & 3 & 1 & 0 \\ 1 & 9 & 0 & 1 \\ 1 & 6 & 0 & 0 \\ 1 & 8 & 0 & 0 \end{bmatrix}. \text{ Hierbij geldt: } (X'X)^{-1} = \begin{pmatrix} 20.1 & -2.8 & -10.3 & 5.1 \\ -2.8 & 0.4 & 1.4 & -0.8 \\ -10.3 & 1.4 & 5.9 & -2.3 \\ 5.1 & -0.8 & -2.3 & 3.1 \end{pmatrix}.$$

De OLS-schatters worden weergegeven door $b_1,...,b_4$. De residuele kwadratensom bij OLS is gelijk aan 0.4. Maak bij het beantwoorden van de onderstaande vragen handig gebruik van de gegevens, zodat u geen onnodige berekeningen hoeft te maken.

a) Wat voor soort variabelen zijn x_2 en x_3?
b) Bepaal de OLS-schattingen van de coëfficiënten.
c) Kunt u zonder berekeningen te maken beredeneren welke waarde het residu bij de derde waarneming aanneemt?
d) Toets of $b_1 = b_2 = 0$.
e) Stel $w = (b_1 + 3)(b_1 + b_2)$. Bewijs dat $E[w] = V[b_1] + C[b_1, b_2] + b_1^2 + b_1 b_2 + 3b_1 + 3b_2$.
f) Bereken de verwachte waarde van w, onder de aanname dat $V[e_i^2] = s^2 = 0.5, \quad b_1 = 0.4, \quad b_2 = 0.6$.

Opgave 3 (10)

Bewijs dat bij het lineaire model $y = Xb + e$ voor de OLS-residuen e geldt $X'e = 0$.

Uitwerking tentamen Econometrie 1, 23 juni 2003

Opgave 1

a) In het derde model zijn de t-waarden van A en X allebei significant, nl. 0.958 / 0.220 = 4.355 en –1.923 / 0.496 = –3.877, terwijl $t_{2.5\%,12}$ = 2.179. Dit suggereert dat beide verklarende variabelen van belang zijn, ondanks de multicollineariteit in het derde model, veroorzaakt door de sterke correlatie tussen A en X. Echter, de coëfficiënt van X is negatief, hetgeen niet acceptabel is (tenzij het een "Giffen good" betreft). Waarschijnlijk fungeert het inkomen X, dat trendmatig stijgt, als een proxy voor de trend. M.a.w. er is een trendmatige daling van de verkopen. Het model dat de voorkeur zou verdienen met A en de trend als verklarende variabelen is niet gegeven. Uit de beschikbare modellen kunnen we het derde model prefereren, omdat we beter een proxy voor de ontbrekende trend kunnen gebruiken, dan dit effect helemaal weglaten. Dus voorkeur voor het derde model.

b) Bij het eerste model is de standaardfout $se(b_A) = \sqrt{\dfrac{s_u^2}{nVar(A)}}$ en bij het derde model is de standaardfout $se(b_A) = \sqrt{\dfrac{s_u^2}{nVar(A)}\dfrac{1}{1-r_{AX}^2}}$. Het verschil is dus de factor $\sqrt{\dfrac{1}{1-r_{AX}^2}} = \sqrt{\dfrac{1}{1-(-0.843225)^2}} = \sqrt{3.46} = 1.86$ en een verschil in de waarde van s_u^2. In dit geval is s_u^2 kleiner bij het derde model en overheerst dus het effect van de factor 1.86, die leidt tot een verhoging van de standaardfout.

c) We nemen aan dat A een relevante verklarende variabele is, gezien de significante t-waarde en het correcte teken van de coëfficiënt. Indien deze wordt weggelaten, dan neemt de overblijvende verklarende variabele X voor een deel de rol over van A omdat X en A gecorreleerd zijn. Dat kunnen we zien aan de vertekening (aangenomen dat A en X deterministisch zijn).

$$E[b_X] = E\left[\frac{Cov(X,S)}{Var(X)}\right] = E\left[\frac{Cov(X, b_1 + b_2 A + b_3 X + u)}{Var(X)}\right] =$$

$$E\left[\frac{b_2 Cov(X,A) + b_3 Cov(X,X) + Cov(X,u)}{Var(X)}\right] = b_2\frac{Cov(X,A)}{Cov(X)} + b_3$$

dus de vertekening is $b_2\dfrac{Cov(A,X)}{Var(X)}$. Omdat Cov(X,A) < 0 en $b_2 > 0$ (nemen we aan) is de vertekening negatief en dit betekent dat we bij het tweede model een lagere coëfficiënt van X mogen verwachten. Dat zien we inderdaad terug in de schattingen.

d) We zien bij het weglaten van een verklarende variabele in het eerste en tweede model dat de overblijvende regressiecoëfficiënt een vertekening van 0 vandaan krijgt. Dit is uiteraard gunstig voor de verklaring, hetgeen zich uit in een relatief hoge t-waarde van de verklarende variabele en een relatief hoge R^2 in het eerste en tweede model. NB. Bij een vertekening naar 0 toe doet het omgekeerde zich voor, dus dan krijg je juist een relatief lage t-waarde en een relatief lage R^2 bij enkelvoudige regressie.

e) In de hulpregressie is de t-waarde van A gelijk aan 0.325/0.0649 = 5.008 en dit is significant ($t_{2.5\%,13}$ = 2.160), dus er is significante heteroscedasticiteit. Dit betekent dat de OLS-schatters nog wel zuiver en consistent zijn, maar niet meer efficiënt. Bovendien zijn de berekende standaardfouten (en de t-toetsen) ongeldig.

f) Aanname: $s_{u_t} = l\,A_t$. Getransformeerd model: $\frac{S_t}{A_t} = b_1\frac{1}{A_t} + b_2 + b_3\frac{X_t}{A_t} + \frac{u_t}{A_t}$. Hierbij wordt aangenomen dat $E\left[\frac{u_t}{A_t}\right] = 0$, $V\left[\frac{u_t}{A_t}\right] = l^2$ (dit is de gegeven aanname), $C\left[\frac{u_t}{A_t}, \frac{u_s}{A_s}\right] = 0$ voor $t \neq s$ en $\frac{u_t}{A_t}$ onafhankelijk van $\frac{1}{A_t}$ en $\frac{X_t}{A_t}$. Omdat de Gauss-Markov aannames gelden, zijn de OLS-schatters BLUE (de beste lineaire zuivere schatters).

Opgave 2

a) x_2 is een kwantitatieve verklarende variabele en x_3 is een verklarende dummy variabele.

b) $b = (X'X)^{-1}X'y$. $X'y = \begin{pmatrix} 26 \\ 164 \\ 9 \\ 7 \end{pmatrix}$, $b = \begin{pmatrix} 6.4 \\ -0.2 \\ -1.2 \\ 2.4 \end{pmatrix}$

c) Omdat er een unieke dummy is voor de derde waarneming zal deze ervoor zorgen dat het residu bij de derde waarneming precies gelijk aan nul wordt.

d) Wald toets $H_0: b_1 = b_2 = 0$ tegen $H_1: b_1 \neq 0$ en/of $b_2 \neq 0$.

$J = 2,\quad R = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix},\quad q = \begin{pmatrix} 0 \\ 0 \end{pmatrix},\quad n = 5,\quad k = 4,\quad e'e = \frac{0.4}{5-4} = 0.4$, zodat

$$F = \frac{\left((Rb-q)'\{R(X'X)^{-1}R'\}^{-1}(Rb-q)\right)/J}{eé/(n-k)} = \frac{(6.4 \quad -0.2)\begin{pmatrix} 20.1 & -2.8 \\ -2.8 & 0.4 \end{pmatrix}^{-1}\begin{pmatrix} 6.4 \\ -0.2 \end{pmatrix}/2}{0.4/1}$$

$$= \frac{(6.4 \quad -0.2)\frac{1}{0.2}\begin{pmatrix} 0.4 & 2.8 \\ 2.8 & 20.1 \end{pmatrix}\begin{pmatrix} 6.4 \\ -0.2 \end{pmatrix}/2}{0.4} = \frac{50.1/2}{0.4} = 62.625 < 199.5 = F_{5\%;2;1}$$

De nulhypothese $b_1 = b_2 = 0$ wordt dus niet verworpen.

e) Bewijs:

$$E[w] = E[(b_1+3)(b_1+b_2)] = E[b_1^2 + b_1b_2 + 3b_1 + 3b_2] = E[b_1^2] + E[b_1b_2] + 3E[b_1] + 3E[b_2]$$
$$= \{V[b_1] + E[b_1]^2\} + \{C[b_1,b_2] + E[b_1]E[b_2]\} + 3b_1 + 3b_2$$
$$= V[b_1] + C[b_1,b_2] + b_1^2 + b_1b_2 + 3b_1 + 3b_2$$

f) $V[b_1]$ = 0.5*20.1 = 10.05, $C[b_1,b_2]$ = 0.5*-2.8 = -1.4, dus
$E[w]$ = 10.05 – 1.4 + 0.4^2 + 0.4*0.6 + 3*0.4 + 3*0.6 = 12.05.

Opgave 3

$e = y - Xb = y - X(X'X)^{-1}X'y = [I - X(X'X)^{-1}X']y$, dus

$X'e = X'[I - X(X'X)^{-1}X']y = [X' - X']y = 0$