Tentamen Econometrie 1, 30 maart 2004, 9.30-11.30 uur

Dit tentamen duurt 2 uur!

De uitslag komt uiterlijk vanaf 20 april 2004 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. 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 (a=8, b=6, c=6, d=10, totaal 30)

Loterijen zijn een belangrijke bron van inkomsten voor overheden. Vaak worden deze loterijen bekritiseerd, omdat deze zouden werken als een belasting voor de armen en laag opgeleiden. Om dit te onderzoeken wordt een steekproef genomen van 100 volwassenen en voor ieder van hen werden de volgende variabelen gemeten:

Y = bestedingen aan loterijen als percentage van het inkomen van het huishouden
X_1 = aantal jaren scholing
X_2 = leeftijd
X_3 = aantal kinderen
X_4 = persoonlijk inkomen (in duizenden euro's)

Met de data werden de volgende modellen geschat:

Model 1:
$$\underset{(1.79)}{\hat{Y}=11.9}-\underset{(0.13)}{0.43X_1}+\underset{(0.025)}{0.029X_2}+\underset{(0.22)}{0.093X_3}-\underset{(0.028)}{0.074X_4} \qquad R^2=0.433,\ RSS=804$$

Model 2:
$$\underset{(1.19)}{\hat{Y}=13.6}-\underset{(0.13)}{0.46X_1}-\underset{(0.027)}{0.070X_4} \qquad R^2=0.424,\ RSS=818$$

a) Interpreteer en toets de coëfficiënten van X_1 en X_2 in model 1.
b) Toets of de variabelen X_2 en X_3 gezamenlijk uit model 1 weggelaten mogen worden.
c) Toets of de variabelen X_1 en X_4 gezamenlijk uit model 2 weggelaten mogen worden.
d) Stel dat model 2 correct is, maar dat de standaardafwijking van de storingen proportioneel is met $1/X_4$. Welke (nadelige) consequenties heeft dit dan voor de kwaliteit van de bovenstaande schattingen en voor de geldigheid van de uitgevoerde toetsen? Geef precies aan hoe deze nadelige consequenties verholpen kunnen worden.

Opgave 2 (totaal 10)

Beschouw twee alternatieve modellen (in matrixnotatie):

Model 1: $y = X_1 b_1 + e_1$
Model 2: $y = X_1 b_1 + X_2 b_2 + e_2$

waarin X_1 een n x k_1 matrix is en X_2 een n x k_2 matrix. Ga er van uit dat model 2 in elk geval correct is.
Leid formules af (conditioneel op X_1 en X_2) voor de verwachting van b_1^1 (= de OLS-schatter van b_1 gebaseerd op model 1) en van b_1^2 (= de OLS-schatter van b_1 gebaseerd op model 2). Onder welke voorwaarden zijn beide schatters zuiver?

Opgave 3 (a=5, b=8, c=12, d=5, e=12, f=18, totaal 60)

Beschouw het model $y = Xb + e$ (in matrixnotatie), waarin X een n x k matrix is met n=20 en k=4 en de eerste kolom van X uit énen bestaat. Schrijf $X = \begin{pmatrix} i & X_B \end{pmatrix}$, waarin $i = (1 \ldots 1)'$ en maak soortgelijk de decompositie $b = \begin{pmatrix} b_A \\ b_B \end{pmatrix}$ (dus $b_A = b_1$ en $b_B = \begin{pmatrix} b_2 & \ldots & b_k \end{pmatrix}'$). Stel dat M_i de matrix operator is die variabelen afwijking op het gemiddelde brengt, dan kunnen we definiëren
$\tilde{y} = M_i y, \tilde{X}_B = M_i X_B, \tilde{e} = M_i e$ enz.. Nu is gegeven:

$$(\tilde{X}_B'\tilde{X}_B)^{-1} = \begin{pmatrix} 0.5 & -0.7 & 0 \\ -0.7 & 1.2 & 0 \\ 0 & 0 & 0.1 \end{pmatrix}, \ \tilde{X}_B'\tilde{y} = \begin{pmatrix} 6 \\ -2 \\ 4 \end{pmatrix}, \ e'e = 80, \ \bar{y} = 7, \ \tfrac{1}{n} i' X_B = \begin{pmatrix} 3 & 2 & 5 \end{pmatrix}.$$

a) Hoe ziet M_i er precies uit? Leidt de exacte gedaante af.
b) Toon aan dat geldt $\tilde{y} = \tilde{X}_B b_B + \tilde{e}$.
c) Bereken de OLS-schatter $b = \begin{pmatrix} b_A \\ b_B \end{pmatrix}$ van b en de standaardfout van de regressie.
d) Definieer $e = y - Xb$ en $\tilde{e} = \tilde{y} - \tilde{X}_B b_B$. Toon aan dat geldt $\tilde{e} = e$.
e) Veronderstel dat het model $y = Xb + e$ voldoet aan de Gauss-Markov aannamen. Onderzoek in hoeverre het model $\tilde{y} = \tilde{X}_B b_B + \tilde{e}$ dan ook voldoet aan de Gauss-Markov aannamen.
f) Stel u wilt kleinste kwadraten toepassen op het model bij onderdeel b) onder de restrictie $Rb_B = q$. (1) Stel de Lagrange functie $L^*(b_B^*, l^*)$ van dit probleem op, waarin b_B^* de schatter van b_B
Vermenigvuldig de vergelijking $\partial L^* / \partial b_B^* = 0$ voor met $R(\tilde{X}_B'\tilde{X}_B)^{-1}$ en bepaal de oplossing van l^*. (4) Substitueer de oplossing van l^* in de vergelijking $\partial L^* / \partial b_B^* = 0$ en bepaal de oplossing van b_B^*. (5) Bereken deze schatter, gegeven de restrictie $2b_2 - b_3 = 5$.

Uitwerking tentamen Econometrie 1, 30 maart 2004

1.a) De coëfficiënt van X_1 is –0.43, d.w.z. een extra jaar scholing leidt tot 0.43 procentpunten minder uitgaven aan loterijen. Dit effect is significant bij 5%, want t = –0.43/0.13 = –3.3 < –1.99 = $-t_{0.025,95}$.

De coëfficiënt van X_2 is 0.029, d.w.z. een jaar hogere leeftijd leidt tot 0.029 procentpunten meer uitgaven aan loterijen. Dit effect is niet significant bij 5%, want t = 0.029/0.025 = 1.16 < 1.99 = $t_{0.025,95}$.

b) Partiële F-toets van model 1 tegen model 2: $F = \dfrac{(818-804)/2}{804/(100-5)} = 0.827 < 3.1 = F_{0.05,2,95}$ dus niet significant bij 5%, d.w.z. X_2 en X_3 mogen gezamenlijk uit model 1 worden weggelaten.

c) Goodness-of-fit toets van model 2: $F = \dfrac{0.424/2}{(1-0.424)/(100-3)} = 35.70 > 3.1 = F_{0.05,2,97}$ dus significant bij 5%, d.w.z. X_1 en X_4 mogen niet gezamenlijk uit model 2 worden weggelaten.

d) Onder heteroscedasticiteit zijn de OLS-schatters nog wel zuiver en consistent, maar niet meer BLUE, d.w.z. niet meer efficiënt. Bovendien zijn de gegeven standaard fouten dan ongeldig en dus ook de uitgevoerde de t-toetsen. Dit kan verholpen worden door gewogen kleinste kwadraten toe te passen (WLS=weighted least squares). Dit houdt in dat het model getransformeerd wordt tot $X_4Y = b_1X_4 + b_2X_4X_1 + b_3X_4X_4 + X_4U$, waarin U de storingsterm van model 2 voorstelt. Het getransformeerde model voldoet aan de Gauss-Markov condities, zodat de OLS-schatters zuiver, consistent en efficiënt (BLUE) zijn en de standaardfouten en t-toetsen zijn dan geldig.

2. $E[b_1^1 \mid X] = E[(X_1'X_1)^{-1}X_1'y \mid X] = E[(X_1'X_1)^{-1}X_1'(X_1b_1 + X_2b_2 + e) \mid X]$

$= E[b_1 + (X_1'X_1)^{-1}X_1'X_2b_2 + (X_1'X_1)^{-1}X_1'e) \mid X] = b_1 + (X_1'X_1)^{-1}X_1'X_2b_2$

Zuiver als $(X_1'X_1)^{-1}X_1'X_2b_2 = 0$. Dit treedt op als $b_2 = 0$ of $X_1'X_2 = 0$.

Schrijf model 2 als $y = Xb + e$ met $X = \begin{pmatrix} X_1 & X_2 \end{pmatrix}$, $b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$.

$E[b^2 \mid X] = E[(X'X)^{-1}X'y \mid X] = E[(X'X)^{-1}X'(Xb + e) \mid X]$

$= E[b + (X'X)^{-1}X'e) \mid X] = b$, dus $E[b_2^2 \mid X] = b_2$.

Zuiver zonder verdere voorwaarden.

3. a) M_i geeft het orthogonale complement bij projectie op i, dus

$$M_i = I_n - i(i'i)^{-1}i' = I_n - \tfrac{1}{n}ii' = \begin{pmatrix} 1-\frac{1}{n} & -\frac{1}{n} & \dots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & \dots & -\frac{1}{n} \\ \dots & \dots & \dots & \dots \\ -\frac{1}{n} & -\frac{1}{n} & \dots & 1-\frac{1}{n} \end{pmatrix}.$$

b) Uit $y = Xb + e = \begin{pmatrix} i & X_B \end{pmatrix}\begin{pmatrix} b_A \\ b_B \end{pmatrix} + M_ie = ib_A + X_Bb_B + e$ volgt

$M_iy = M_iib_A + M_iX_Bb_B + M_ie \quad \Rightarrow \quad \tilde{y} = 0 + \tilde{X}_Bb_B + \tilde{e}$.

c) De OLS-schatter van b_B is volgens de stelling van Frisch-Waugh gelijk aan $b_B = (\tilde{X}_B'\tilde{X}_B)^{-1}\tilde{X}_B'\tilde{y} =$

$\begin{pmatrix} 0.5 & -0.7 & 0 \\ -0.7 & 1.2 & 0 \\ 0 & 0 & 0.1 \end{pmatrix}\begin{pmatrix} 6 \\ -2 \\ 4 \end{pmatrix} = \begin{pmatrix} 4.4 \\ -6.6 \\ 0.4 \end{pmatrix}$. En $b_A = \bar{y} - \begin{pmatrix} \bar{x}_2 & \bar{x}_3 & \bar{x}_4 \end{pmatrix} b_B = 7 - [3\cdot 4.4 - 2\cdot 6.6 + 5\cdot 0.4] = 5$.

De standaardfout van de regressie is $s=\sqrt{\dfrac{e'e}{n-k}}=\sqrt{\dfrac{80}{20-4}}=\sqrt{5}=2.236$.

d) $\tilde{e}=\tilde{y}-\tilde{X}_B b_B = M_i y - M_i X_B b_B = M_i y-(M_i i b_A + M_i X_B b_B) = M_i[y-(ib_A+X_B b_B)] = M_i[y-Xb]$
$= M_i e = M_i M y = My = e$.

e) Veronderstel de Gauss-Markov condities voor $y = Xb + e$, d.w.z. dit is een correct model met rang(X)=k, $E[e \mid X]=0, V[e \mid X]=s^2 I_n$. Hieruit volgt ook $C[X,e]=0$.
Bij b) is aangetoond dat het model $\tilde{y}=\tilde{X}_B b_B+\tilde{e}$ dan ook correct is. De matrix $\tilde{X}_B$ is van volle rang, want als dit niet zo zou zijn, dan zou er een vector $a\neq 0$ bestaan zodanig dat $\tilde{X}_B a=0$, en dan zou ook gelden $M_i X_B a=(I_n-\frac{1}{n}ii')X_B a = X_B a - i(\overline{X}'_B a)=\begin{pmatrix} i & X_B\end{pmatrix}\begin{pmatrix}\overline{X}'_B a\\ a\end{pmatrix}=0$ met $\begin{pmatrix}\overline{X}'_B a\\ a\end{pmatrix}\neq 0$, zodat X niet van volle rang zou zijn. (NB. Volgens de stelling van Frisch-Waugh krijgen we de OLS-schatter van b_B ook door op dit model OLS toe te passen en dit betekent ook dat $\tilde{X}_B$ van volle rang moet zijn). Verder is $E[\tilde{e}\mid X]=E[M_i e\mid X]=M_i E[e\mid X]=0$, dus ook voldaan. Tenslotte is $V[\tilde{e}\mid X]=V[M_i e\mid X]=M_i V[e\mid X]M_i=M_i s^2 I_n M_i = s^2 M_i$. Dit betekent dat $V[e_i\mid X]=(1-\frac{1}{n})s^2$ constant is, dus de storingen zijn homoscedastisch. Echter, $C[e_i,e_j\mid X]=-\frac{1}{n}s^2\neq 0$ voor $i\neq j$, dus de storingen zijn niet ongecorrelleerd. Alleen aan de eis van ongecorrelleerde storingen is niet voldaan.

f) (1) $L^*(b_B^*,l^*)=\left(\tilde{y}-\tilde{X}_B b_B^*\right)'\left(\tilde{y}-\tilde{X}_B b_B^*\right)+2(l^*)'\left(Rb_B^*-q\right)$.

(2) $\dfrac{\partial L^*}{\partial b_B^*}=-2\tilde{X}'_B\left(\tilde{y}-\tilde{X}_B b_B^*\right)+2R'l^*=0$, $\dfrac{\partial L^*}{\partial l^*}=2(Rb_B^*-q)=0$

(3) $-2R(\tilde{X}'_B\tilde{X}_B)^{-1}\tilde{X}'_B\left(\tilde{y}-\tilde{X}_B b_B^*\right)+2R(\tilde{X}'_B\tilde{X}_B)^{-1}R'l^*=0 \quad\Rightarrow$

$-2R(\tilde{X}'_B\tilde{X}_B)^{-1}\tilde{X}'_B\tilde{y}+2Rb_B^*+2R(\tilde{X}'_B\tilde{X}_B)^{-1}R'l^*=0 \quad\Rightarrow$

$-2Rb_B+2q+2R(\tilde{X}'_B\tilde{X}_B)^{-1}R'l^*=0 \quad\Rightarrow\quad l^*=[R(\tilde{X}'_B\tilde{X}_B)^{-1}R']^{-1}(Rb_B-q)$

(4) $-\tilde{X}'_B\left(\tilde{y}-\tilde{X}_B b_B^*\right)+R'[R(\tilde{X}'_B\tilde{X}_B)^{-1}R']^{-1}(Rb_B-q)=0 \quad\Rightarrow$

$b_B^*=(\tilde{X}'_B\tilde{X}_B)^{-1}\{\tilde{X}'_B\tilde{y}-R'[R(\tilde{X}'_B\tilde{X}_B)^{-1}R']^{-1}(Rb_B-q)\} \quad\Rightarrow$

$b_B^*=b_B-(\tilde{X}'_B\tilde{X}_B)^{-1}R'[R(\tilde{X}'_B\tilde{X}_B)^{-1}R']^{-1}(Rb_B-q)$

(5) Restrictie $\begin{pmatrix}2 & -1 & 0\end{pmatrix}b_B=5$, dus $R=\begin{pmatrix}2 & -1 & 0\end{pmatrix}$, $q=5$.

$$b_B^*=\begin{pmatrix}4.4\\-6.6\\0.4\end{pmatrix}-\begin{pmatrix}0.5 & -0.7 & 0\\-0.7 & 1.2 & 0\\0 & 0 & 0.1\end{pmatrix}\begin{pmatrix}2\\-1\\0\end{pmatrix}\left[\begin{pmatrix}2 & -1 & 0\end{pmatrix}\begin{pmatrix}0.5 & -0.7 & 0\\-0.7 & 1.2 & 0\\0 & 0 & 0.1\end{pmatrix}\begin{pmatrix}2\\-1\\0\end{pmatrix}\right]^{-1}\left[\begin{pmatrix}2 & -1 & 0\end{pmatrix}\begin{pmatrix}4.4\\-6.6\\0.4\end{pmatrix}-5\right]$$

$$\Rightarrow b_B^*=\begin{pmatrix}4.4\\-6.6\\0.4\end{pmatrix}-\begin{pmatrix}1.7\\-2.6\\0\end{pmatrix}\left[\begin{pmatrix}2 & -1 & 0\end{pmatrix}\begin{pmatrix}1.7\\-2.6\\0\end{pmatrix}\right]^{-1}\left[15.4-5\right] \Rightarrow b_B^*=\begin{pmatrix}4.4\\-6.6\\0.4\end{pmatrix}-\begin{pmatrix}1.7\\-2.6\\0\end{pmatrix}[6]^{-1}[10.4]$$

$$\Rightarrow b_B^*=\begin{pmatrix}4.4\\-6.6\\0.4\end{pmatrix}-\begin{pmatrix}2.947\\-4.507\\0\end{pmatrix}=\begin{pmatrix}1.453\\-2.093\\0.4\end{pmatrix}$$