Stockholm University
Department of Statistics
Per Gösta Andersson

Econometrics I

WRITTEN EXAMINATION
Tuesday December 4, 2018, 10 am - 3 pm

Allowed tools: Pocket calculator

Passing rate: 50% of overall total, which is 100 points. For detailed grading criteria, see the course description.

For the maximum number of points on each problem detailed and clear solutions are required.

If not indicated otherwise, the disturbance terms u_i in the models are supposed to fulfill the usual requirements of normality, homoscedasticity and independence.

1. (21p) Följande datamaterial har blivit klassiskt: genomsnittspriset Y i dollar för begagnade bilar 1957 i USA, där X = ålder (år).

Ålder (X)	Genomsnittspris (Y)
1	2651
2	1943
3	1494
4	1087
5	765
6	538
7	484
8	290
9	226
10	204

Följande summor är givna: $\sum_{i=1}^{10} X_i = 55$, $\sum_{i=1}^{10} X_i^2 = 385$, $\sum_{i=1}^{10} Y_i = 9\,682$, $\sum_{i=1}^{10} Y_i^2 = 15\,502\,372$, $\sum_{i=1}^{10} X_iY_i = 32\,202$

(a) Skatta β-parameterna för modellen

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

Är du nöjd med anpassningen av skattade modellen till data?
Vad är förklaringsgraden R^2?
Vad skulle residualernas värden kunna tyda på?

(b) Ett alternativ till modellen i (a) är att (naturliga) logaritmen för Y modelleras enligt den linjära modellen

$$lnY_i = \beta_1' + \beta_2' X_i + u_i'$$

Skatta β_1' och β_2'.
Lite räknehjälp: $\sum_{i=1}^{10} lnY_i = 65.2734$, $\sum_{i=1}^{10} (lnY_i)^2 = 433.428$, $\sum_{i=1}^{10} X_i lnY_i = 334.445$
Vad är förklaringsgraden R^2? Är det lämpligt att direkt jämföra den med förklaringsgraden för modellen i(a)?
Hur ser uttrycket ut för $Y_i|X_i$?
Skatta $E(Y_i|X_i)$ för de givna X-värdena och jämför med motsvarande skattningar för modellen i (a). Resultat? Åskådliggör skillnaderna i lämplig plott.

2. (18p) Antag att vi har brukat mäta värdet på någon kvantitet X med hjälp av en viss metod A som ger värdet Y. Nu har en ny mätmetod B introducerats. I följande tabell har vi nio mätningar med mätmetod A och en med B.

	X	Y
A	4.0	3.7
A	8.0	7.8
A	12.5	12.1
A	16.0	15.6
A	20.0	19.8
A	25.0	24.5
B	31.0	31.1
A	36.0	35.5
A	40.0	39.4
A	40.0	39.5

(a) Om vi skattar en enkel linjär regressionsmodell utifrån dessa data, går det då att hävda att mätningen med metod B verkar avvika från de övriga? (Genomför alltså först regressionsanalysen och skatta β-parameterna.)

(b) Oavsett resultatet i (a) ange (utan beräkning) hur man skulle formellt kunna testa huruvida mätmetod B avviker från de övriga.

3. (15p)

   (a) Visa analytiskt (räknemässigt) att

$$\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^{n}(X_i - \bar{X})Y_i$$

   (b) Antag att $(X_1, Y_1), \ldots, (X_n, Y_n)$ används för att OLS-skatta β-parameterna i en enkel linjär regressionsmodell. Hur förändras skattningarna (om de alls förändras) om vi i stället använder oss av $(X_1, Y_1 - c), \ldots, (X_n, Y_n - c)$ (för någon känd konstant c)? Motivera både analytiskt och grafiskt.

4. (15p) Vi vill studera hur inkomst (X) påverkar konsumtion (Y) och även studera om vi har säsongseffekter (kvartalsvis). Följande skattade modell erhölls för åren 2012-2017:

$$\hat{Y}_i = 1.20 + 2.31X_i - 0.71D_{2i} - 0.77D_{3i} - 0.03D_{4i},$$

   där D_{2i}, D_{3i} och D_{4i} är indikatorvariabler för kvartal 2, 3 och 4. Antag $X = 10$.

   (a) Vad är skattade väntevärdet för Y i kvartal 1?
   (b) Hur tolkas skattade koefficienten -0.71 för kvartal 2?
   (c) Om vi i stället skulle ha haft en modell med indikatorvariabler för alla fyra kvartalen, hur skulle vi då generellt tolka koefficienten framför D_{2i}?

5. (16p) Sant eller falskt? Motivering krävs.

   (a) Autokorrelation kan uppkomma om en viktig regressor inte är inkluderad i en linjär regressionsmodell.
   (b) Om förutsättningarna är giltiga (tillräckliga) för Runs-testet att användas är de också giltiga för användning av Durbin-Watsons test.
   (c) Orsaken till förkastande av nollhypotsen i White-testet kan enbart bero på heteroskedasticitet.
   (d) Antag att vi har två starkt korrelerade X-variabler i en linjär regressionsmodell. Ett F-test kan då genomföras, men t-test kan bli missvisande.

6. (15p) Fyll i de saknade värdena i följande tabell:

Predictor	Coef	SE Coef	T	P
Constant	(a)	0.3501	0.59	0.568
X1	1.8615	(b)	2.31	0.040
X2	2.7241	0.7124	(c)	0.002

S = (d)   R-Sq = 83.4%   R-Sq(adj) = 80.6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	(e)	(f)	(g)	(h)	0.000
Residual Error	12	17.28	1.44		
Total	(i)	104.09			

Formula sheet, Econometrics I, Fall 2018

Under the simple linear model $Y_i = \beta_1 + \beta_2 X_i + u_i$, where $u_i \sim N(0, \sigma^2)$ and given independent pairs of observations $(Y_1, X_1), \ldots, (Y_n, X_n)$, the OLS estimators are:

$$\begin{aligned}
\widehat{\beta}_1 &= \overline{Y} - \widehat{\beta}_2 \overline{X} \\
\widehat{\beta}_2 &= \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})^2} \\
\widehat{\sigma}^2 &= \frac{RSS}{n-2} = \frac{\sum \left(Y_i - \widehat{Y}_i\right)^2}{n-2}
\end{aligned}$$

where $\widehat{Y}_i = \widehat{\beta}_1 + \widehat{\beta}_2 X_i$ and where $E(\widehat{\beta_1}) = \beta_1$, $E(\widehat{\beta_2}) = \beta_2$ and $E(\widehat{\sigma}^2) = \sigma^2$ and further

$$\begin{aligned}
V(\widehat{\beta}_1) &= \frac{\sum X_i^2}{n \sum (X_i - \overline{X})^2} \sigma^2 \\
V(\widehat{\beta}_2) &= \frac{\sigma^2}{\sum (X_i - \overline{X})^2} \\
V(\widehat{Y}_0) &= \sigma^2 \left(\frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum (X_i - \overline{X})^2} \right) \\
V(Y_0 - \widehat{Y}_0) &= \sigma^2 \left(1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum (X_i - \overline{X})^2} \right)
\end{aligned}$$

Distributional results:

$$\begin{aligned}
\frac{\widehat{\beta}_i - \beta_i}{se(\widehat{\beta}_i)} &\sim t(n-2), \quad i = 1, 2 \\
\frac{\widehat{\sigma}^2 (n-2)}{\sigma^2} &\sim \chi^2(n-2)
\end{aligned}$$

Coefficient of determination:

$$r^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum \left(Y_i - \widehat{Y}_i\right)^2}{\sum (Y_i - \overline{Y})^2}$$

Coefficient of correlation:

$$r = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum (X_i - \overline{X})^2 \sum (Y_i - \overline{Y})^2}}$$

where $r = \pm\sqrt{r^2}$

If we let $Y_i^* = w_1 Y_i$ and $X_i^* = w_2 X_2$, then

$$\widehat{\beta}_1^* = w_1\widehat{\beta}_1, \quad \widehat{\beta}_2^* = \left(\frac{w_1}{w_2}\right)\widehat{\beta}_2, \quad \widehat{\sigma}^{*2} = w_1^2\widehat{\sigma}^2$$

Under the multiple linear regression model $Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$, where $u_i \sim N(0, \sigma^2)$ and given independent vectors of observations $(Y_1, X_{21}, \ldots, X_{k1}), \ldots, (Y_n, X_{2n}, \ldots, X_{kn})$, the following holds for the OLS estimators:

$$\widehat{\sigma}^2 = \frac{RSS}{n-k} = \frac{\sum\left(Y_i - \widehat{Y}_i\right)^2}{n-k}$$

$$\frac{\widehat{\beta}_i - \beta_i}{se(\widehat{\beta}_i)} \sim t(n-k), \quad i = 1, \ldots, k$$

$$\frac{\widehat{\sigma}^2(n-k)}{\sigma^2} \sim \chi^2(n-k)$$

The multiple coefficient of determination:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum\left(Y_i - \widehat{Y}_i\right)^2}{\sum\left(Y_i - \overline{Y}\right)^2}$$

Adjusted:

$$\overline{R}^2 = 1 - \frac{RSS/(n-k)}{TSS/(n-1)}$$

Testing $H_0 : \beta_2 = \ldots \beta_k = 0$:

$$F = \frac{ESS/(k-1)}{RSS/(n-k)} = \frac{\sum\left(\widehat{Y}_i - \overline{Y}\right)^2/(k-1)}{\sum\left(Y_i - \widehat{Y}_i\right)^2/(n-k)}$$

Comparing an "old" model with a "new" (larger):

$$\begin{aligned} F &= \frac{(ESS_{new} - ESS_{old})/\text{number of new regressors}}{RSS_{new}/(n - \text{number of parameters in the new model})} \\ &= \frac{(R^2_{new} - R^2_{old})/\text{number of new regressors}}{(1 - R^2_{new})/(n - \text{number of parameters in the new model})} \end{aligned}$$

Comparing an "unrestricted" model with a "restricted":

$$F = \frac{(RSS_R - RSS_{UR})/m}{RSS_{UR}/(n-k)} = \frac{(R^2_{UR} - R^2_R)/m}{(1 - R^2_{UR})/(n-k)}$$

where m is the number of linear constraints and k is the number of parameters in the unrestricted model.

Variance inflation factor:

$$VIF_j = \frac{1}{1 - R_j^2}$$

Auxiliary regression:

$$F_j = \frac{R_j^2/(k-2)}{(1 - R_j^2)/(n-k+1)}$$

where $R_j^2 = R^2$ in the regression of x_j on the remaining $(k-2)$ regressors.

Tests of heteroscedasticity: (all test statistics are evaluated under the null hypothesis of no heteroscedasticity)

White's test: Regress $\hat{u}_i^2$ against the $k-1$ regressors and the squares of these.
Test statistic: $n\,R^2 \overset{\text{appr}}{\sim} \chi^2(2(k-1))$

Glejser test: Regress $|\hat{u}_i|$ against the regressor X_j (one regressor at a time)
Test statistic: t-test of the slope

Park test: Regress $ln\hat{u}_i^2$ against the regressor lnX_j, (one regressor at a time)
Test statistic: t-test of the slope

Goldefeld Quandt test of equal variances in two separate regressions:
Test statistic: $\frac{S_1^2}{S_2^2} \sim F(n_1 - k_1, n_2 - k_2)$

Tests of autocorrelation:

The Runs test: For R = number of runs, where $N = N_1 + N_2$ total number of observations:

$$E(R) = \frac{2N_1N_2}{N} + 1$$
$$V(R) = \frac{2N_1N_2(2N_1N_2 - N)}{N^2(N-1)}$$

The Durbin Watson d statistic:

$$d = \frac{\sum_{t=2}^{n}(\widehat{u}_t - \widehat{u}_{t-1})^2}{\sum_{t=1}^{n}\widehat{u}_t^2}$$

Breusch Godfrey test: Nullypothesis: H_0: $\rho_1 = \rho_2 = \ldots = \rho_K = 0$
Test statistic: nR^2 from the regression of $\hat{u}_t$ on the regressors which have produced $\hat{u}_t$ plus lagged $\hat{u}_t$ up to lag K.
n = the number of observations used in this regression.
The test statistic is approximately $\chi^2(K)$

Akaike's information criterion:

$$AIC = \frac{e^{2k/n} RSS}{n}$$

Schwartz's information citerion:

$$SIC = \frac{n^{k/n} RSS}{n}$$

Mallow's C_p criterion:

$$C_p = \frac{RSS_p}{\hat{\sigma}^2} - (n - 2p)$$

Logistic regression (logit model):

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_1 + \beta_2 X_2 + \cdots + \beta_k X_k)}}, \quad ln(\frac{P(Y = 1)}{1 - P(Y = 1)}) = \beta_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

Department of Statistics

Stockholms universitet

Correction sheet

Date: 04/12/2018

Room: Brunnsvikssalen

Course: Econometrics (eng)

Exam: Econometrics I (eng)

Anonymous code: 0002-KTE

☒ I authorise the anonymous posting of my exam, in whole or in part, on the department homepage as a sample student answer.

NOTE! ALSO WRITE ON THE BACK OF THE ANSWER SHEET

Mark answered questions

1	2	3	4	5	6	7	8	9	Total number of pages
X	X	X	X	X	X				3 3C
Teacher's notes 21	13	14	15	16	15				

Points	Grade	Teacher's sign.
94	A	PGd

SU, DEPARTMENT OF STATISTICS

Room: Brunnsvikssalen Anonymous code: 0002-KTE Sheet number: 1

$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad n = 10$

1\. a) OLS:

$$\hat{\beta}_2 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{32202 - 10 \cdot \frac{55}{10} \cdot \frac{9682}{10}}{385 - 10\left(\frac{55}{10}\right)^2} \approx -255{,}14$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = \frac{9682}{10} - \left(-255{,}14 \cdot \left(\frac{55}{10}\right)\right) \approx 2371{,}47$$

dvs. $\boxed{\hat{Y}_i = 2371{,}47 - 255{,}14 X_i}$

inte nöjd då denna avviker ganska kraftigt från obs. värden

För enkel linjär regression
$R^2 = r^2 = (-0{,}9361)^2 \approx 0{,}8763$

$$r = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum x_i^2 - n\bar{x}^2)(\sum y_i^2 - n\bar{y}^2)}} = \frac{32202 - 10 \cdot \frac{55}{10} \cdot \frac{9682}{10}}{\sqrt{\left(385 - 10 \cdot \left(\frac{55}{10}\right)^2\right)\left(15502372 - 10\left(\frac{9682}{10}\right)^2\right)}}$$

$\approx -0{,}9361$

$0{,}8763 = R^2 = 1 - \frac{RSS}{TSS}$

$TSS = \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - n\bar{Y}^2 =$
$= 15502372 - 10\left(\frac{9682}{10}\right)^2 =$
$= 6128259{,}6$

$\frac{RSS}{TSS} = 0{,}1237$

$\sum \hat{u}_i^2 = RSS = 0{,}1237 \cdot 6128259{,}6 =$
$= 758065{,}7125$

X	$\hat{u}_i$	+/-	Runs
1	534,67	+	1
2	81,81	+	
3	-112,05	-	1 — 3 runs
4	-263,91	-	
5	-350,77	-	
6	-302,63	-	
7	-101,49	-	
8	-40,35	-	
9	150,79	+	1
10	363,93	+	

Svar:
Residualernas värden till höger indikerar tänkbart autokorrelation i modellen. Detta skulle kunna testas med ett Runs test, där få runs (i vårt fall) indikerar positiv autokorrelation. P.g.a detta är jag inte nöjd med ~~den skattade modellen~~ till data. Plus att residualerna är relativt stora värden.

Residualerna kan också indikera att modellen är felspecificerad, eller att vi inte tagit med en [illegible]

b) OLS: $\ln Y_i = \beta_1' + \beta_2' X_i + u_i'$

$$\hat{\beta}_2' = \frac{\Sigma (\ln(Y_i))(X_i) - n \cdot \overline{\ln Y} \cdot \overline{X}}{\Sigma X_i^2 - n\overline{X}^2} = \frac{334{,}445 - 10 \cdot \left(\frac{65{,}2734}{10}\right)\left(\frac{55}{10}\right)}{385 - 10\left(\frac{55}{10}\right)^2} \approx$$

$$\approx -0{,}2977$$

$$\hat{\beta}_1' = \overline{\ln Y} - \hat{\beta}_2' \overline{X} = \frac{65{,}2734}{10} - (-0{,}2977)\frac{55}{10} \approx 8{,}1646$$

$$\boxed{\ln(\hat{Y}) = 8{,}1646 - 0{,}2977 X_i}$$

Det är inte lämpligt att direkt jämföra R^2 för dessa två modeller ty de står i olika form (modell 1: Y_i och modell 2: $\ln Y_i$). Därav bör man <u>inte</u> jämföra dessa värden.

$$\boxed{R^2 = r^2 = (-0{,}9962)^2 \approx 0{,}9926}$$ ← förklaringsgrad modell 2

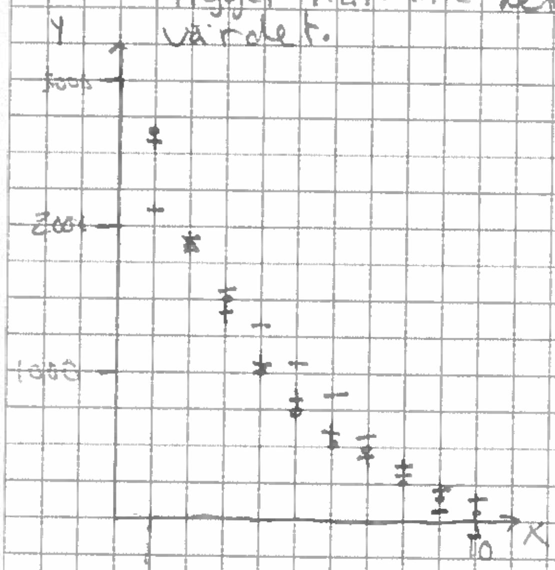
$$r = \frac{\Sigma(\ln(Y_i))(X_i) - n \cdot \overline{\ln Y} \cdot \overline{X}}{\sqrt{(\Sigma X_i^2 - n\overline{X}^2)(\Sigma (\ln Y_i)^2 - n \cdot \overline{\ln Y}^2)}} = \frac{334{,}445 - 10\left(\frac{65{,}2734}{10}\right)\left(\frac{55}{10}\right)}{\sqrt{\left(385 - 10\left(\frac{55}{10}\right)^2\right)\left(433{,}429 - 10\left(\frac{65{,}2734}{10}\right)^2\right)}} \approx$$

$$\approx -0{,}9963$$

$$e^{\ln Y_i} = Y_i = e^{\beta_1' + \beta_2' X_i + u_i'}$$

Svar: Om vi jämför värdena i tabellen till höger ser vi att Modell 2 ligger närmare det observerade värdet.

X	Modell 1: $\hat{Y}_i$	Modell 2: $\hat{Y}$
1	2116,33	2609,46
2	1861,19	1937,59
3	1606,05	1438,71
4	1350,91	1068,27
5	1095,77	793,22
6	840,63	588,98
7	585,49	437,34
8	330,35	324,73
9	75,21	241,12
10	−179,93	179,03



Modell 1: – Modell 2: +

Obs. värden: •

← Se tabell till vänster

Som vi ser i tabellen följer + och • varandra bättre än – och •.

/2

SU, DEPARTMENT OF STATISTICS

Room: Brunnsvikssalen **Anonymous code:** 0002-KTE **Sheet number:** 2

$Y_i = \beta_1 + \beta_2 X_i$

2. a) OLS

$$\hat{\beta}_2 = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n\bar{X}^2} = \frac{8846{,}5 - 10 \cdot \frac{232{,}5}{10} \cdot \frac{229}{10}}{6974{,}25 - 10 \cdot \left(\frac{232{,}5}{10}\right)^2} \approx 0{,}9948$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = \frac{229}{10} - 0{,}9948 \cdot \frac{232{,}5}{10} \approx -0{,}2281$$

$$\boxed{\hat{Y}_i = -0{,}2281 + 0{,}9948 X_i}$$

Ja, man kan hävda det, ty resultatet av den skattade modellen visar att värdet på $\hat{Y}_i$ inte kan vara större än X_i-värdet, vilket gäller för alla observerade värden på Y_i med metod A. Däremot för mätmetod B har vi observerat ett Y_i-värde större än X_i-värdet vilket inte är förenligt med modellen. Man bör dock tänka på att vi endast har få observationer och där endast en observation med mätmetod B, men utifrån resultaten anser jag att man kan hävda att mätningen med metod B avviker från de övriga.

Prediktioner?

b) Man skulle kunna beräkna ett prediktionsintervall för en enskild Y-observation på önskad signifikansnivå (t.ex $\alpha = 0{,}05$). Täcker det beräknade prediktionsintervallet för $Y \mid X_i = 31$ det observerade värdet med mätmetod B kan vi inte förkasta H_0, att mätmetod avviker från de övriga, med 95% konfidens. Däremot om det uträknade prediktionsintervallet för $Y_i \mid X_i = 31$ ej täcker det observerade värdet (31,1) kan vi förkasta H_0 med 95% konfidens och säga att mätmetod B skiljer sig från de övriga.

$Y \mid X_i = 31 \rightarrow$

$n = 10$
$k = 2$

$t_{0{,}025}(n-k) = t_{0{,}025}(8)$

95% P.I. för $Y \mid X_i = 31$

$\hat{Y}_0 \pm t_{\frac{\alpha}{2}}(8) \cdot \sqrt{V(Y - \hat{Y}_0)}$

← används 5 m ene pc!

/13

3\. a) $\sum(X_i-\bar{X})(Y_i-\bar{Y}) = \sum(X_i-\bar{X})Y_i - \sum(X_i-\bar{X})\bar{Y} =$ konstant $\sum X_i$

$= \sum(X_i-\bar{X})Y_i - \bar{Y}\left(\sum(X_i-\bar{X})\right) = \sum(X_i-\bar{X})Y_i - \bar{Y}\left(\sum X_i - n\bar{X}\right) =$

$= \sum(X_i-\bar{X})Y_i - \bar{Y}\cdot 0 = \sum(X_i-\bar{X})Y_i$

V.S.V.

b) $$\hat{\beta}_2' = \frac{\sum(X_i-\bar{X})\left((Y_i-c)-(\bar{Y}-c)\right)}{\sum(X_i-\bar{X})^2} = \frac{\sum(X_i-\bar{X})(Y_i-c-\bar{Y}+c)}{\sum(X_i-\bar{X})^2} =$$

$$= \frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sum(X_i-\bar{X})^2}$$

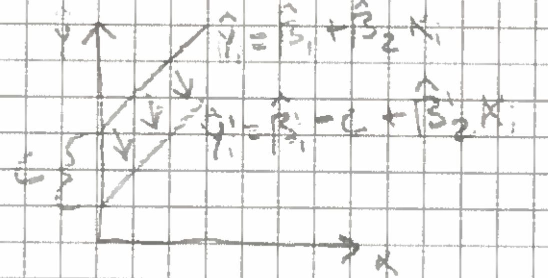
dvs $\hat{\beta}_2'$ förändras inte utan $\hat{\beta}_2' = \hat{\beta}_2$

$$\hat{\beta}_1' = (\bar{Y}-c) - \hat{\beta}_2\bar{X} = (\bar{Y} - \beta_2\bar{X}) - c$$

urspr. skattning av $\hat{\beta}_1$ — konstant

dvs $\hat{\beta}_1'$ kommer att förändras i form av att intercepet förflyttas med $-c$

$\sum X_i$



Detta visas grafiskt i tabellen till vänster, intercepet har flyttats ned med c enheter med lutningen är densamma.

/4

4\. a) $E(Y_i \mid X=10, D_2=0, D_3=0, D_4=0) = \boxed{1{,}20 + 2{,}31\cdot 10 = 24{,}3}$

dvs skattade väntevärdet för kvartal 1 är 24,3 under antagandet att $X=10$.

b) Den skattade koefficienten −0,71 för kvartal 2 tolkas som den förväntade skillnaden i konsumtion jämfört med kvartal 1. Då denna är negativ indikerar detta att vi förväntar oss lägre konsumtion i kv 2 än i kv 1 i genomsnitt, allt annat lika.

allt annat lika

c) Då skulle vi istället tolka koefficienten som intercept för kvartal 2 dvs genomsnittlig (förväntad) konsumtion, givet ingen inkomst ($X_i=0$), i kvartal 2.

/5

SU, DEPARTMENT OF STATISTICS

Room: Brunnsvikssalen **Anonymous code:** 0002-KTE **Sheet number:** 3

5. a) Sant, ta exemplet att en bra modell har formen:
$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$, men vi använder modell
$Y_i' = \beta_1' + \beta_2' X_{2i} + V_i$. Då kommer störningsvariabeln i vår modell ha formen $V_i = \beta_3 X_{3i} + u_i$ och resulterar i att $Cov(V_i, V_j) \neq 0$. Dvs vi har nu autokorrelation, vilket gör påståendet sant.

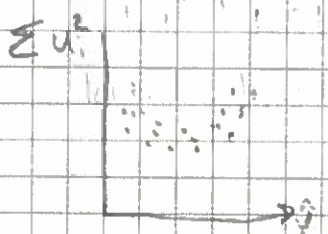
R

b) Falskt. Runs test är ett icke-parametriskt test som inte förutsätter exempelvis normalfördelning, medan Durbin-Watsons test bygger på t.ex. antagande om normalfördelade störningsvariabler. Därmed är påståendet falskt.

R

c) Falskt. Om vi förkastar nollhypotesen i White-testet kan orsaken också vara att vår modell är felspecificerad. Ett exempel ges nedan:

R



I grafen ser vi att residualerna indikerar att vi kanske borde undersöka ett kvadratiskt samband. (inte heteroskedasticitet). Därmed är påståendet falskt, ty jag visat att det kan bero på annan faktor.

d) Sant. Om vi genomför t-test för parametrarna i en multikolinjär modell (som exemplet antyder) kommer resultatet av dessa kunna bli missvisande och inte gå i linje med resultatet från ett övergripande F-test. F-testet däremot kan fortfarande användas då den endast säger oss om minst en av de förklarande variablerna bör tas med i modellen. Dvs F-tester är fortfarande ok att använda, men inte t-test. Påståendet är falskt

R

1/6

6. $a = 0{,}206559$ $\quad \frac{a}{0{,}3501} = 0{,}59 \Rightarrow a = 0{,}3501 \cdot 0{,}59 = 0{,}206559$

$b = 0{,}8015$ $\quad \frac{1{,}8515}{b} = 2{,}31 \Rightarrow b = \frac{1{,}8515}{2{,}31} \approx 0{,}8015$

$c = 3{,}82$ $\quad \frac{2{,}7241}{0{,}7124} = c = 3{,}82$

$d = 1{,}2$ $\quad s = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{RSS}{n-k}} = \sqrt{\frac{17{,}28}{12}} = 1{,}2$

$e = 2$ $\quad e = k - 1 = 3 - 1 = 2$

$f = 86{,}81$ $\quad f = ESS = TSS - RSS = 104{,}09 - 17{,}28 = 86{,}81$

$g = 43{,}405$ $\quad g = \frac{f}{e} = \frac{86{,}81}{2} = 43{,}405$

$h = 30{,}14$ $\quad F = \frac{ESS/(k-1)}{RSS/(n-k)} = \frac{86{,}81/2}{17{,}28/12} \approx 30{,}14$

$i = 14$ $\quad k = 3 \quad n - k = 12 \Rightarrow n = 12 + k = 15$

$i = n - 1 = 15 - 1 = 14$

1/15