

Department of Statistics

Correction sheet

Date: 28/11/2017

Room: Brunnsvikssalen

Course: Econometrics (eng)

Exam: Econometrics I (eng)

Anonymous code:

EKI-HEC-XKZ

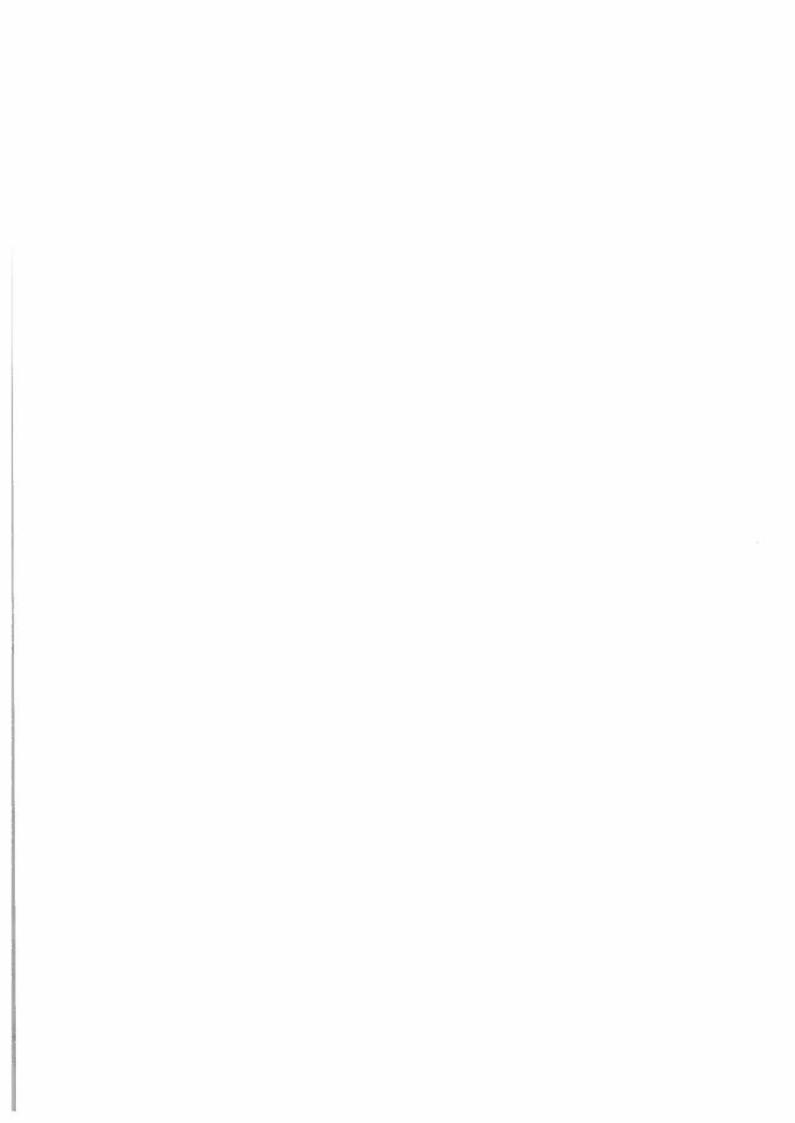
I authorise the anonymous posting of my exam, in whole or in part, on the department homepage as a sample student answer.

NOTE! ALSO WRITE ON THE BACK OF THE ANSWER SHEET

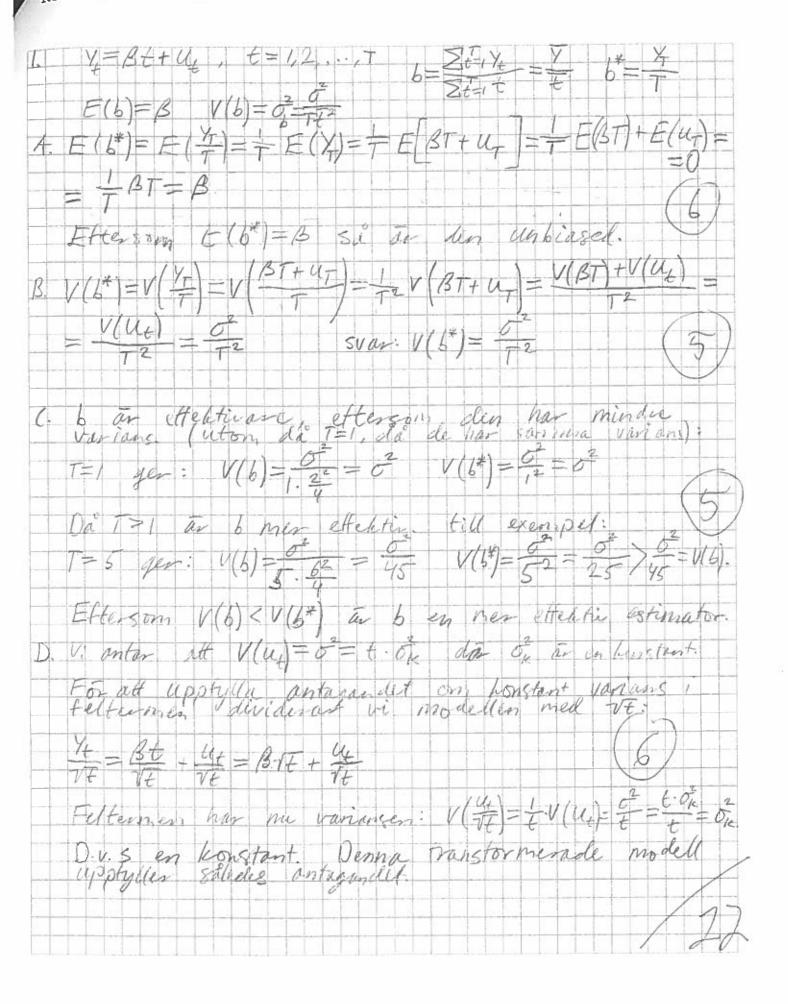
Mark answered questions

1	2	3	4	5	6	7	8	9	Total number of pages
X	X	×	×	×					5 .
Teacher's notes	32	6	12	26					

Points Grade Teacher's sign

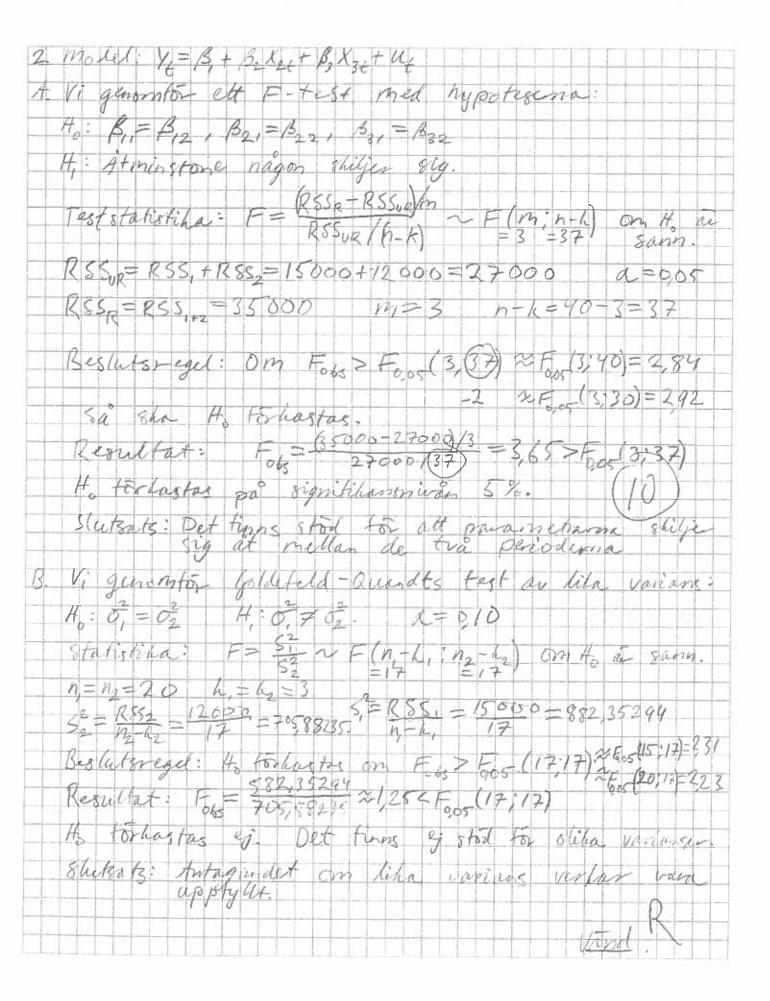


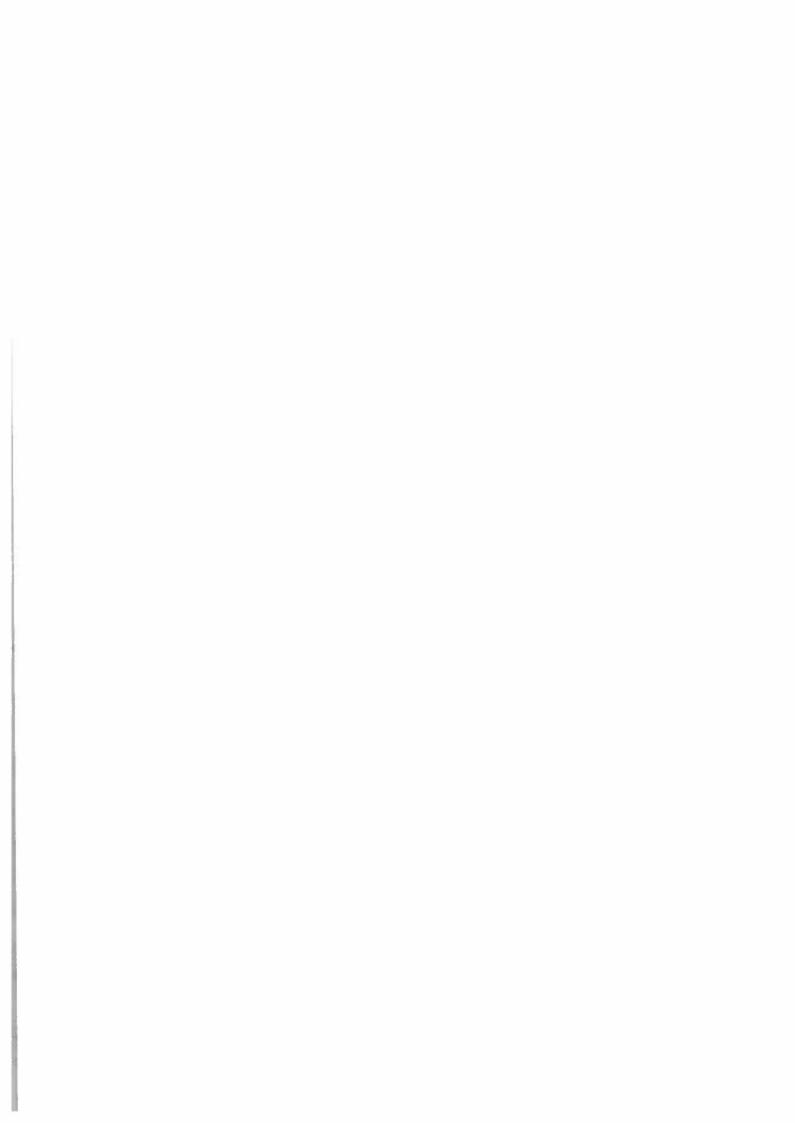
Room: Brunsviks alen Anonymous code: EKHEC-XKZ Sheet number:





Room: Brunsvilsgalen Anonymous code: EKI-HEC-XKZSheet number:



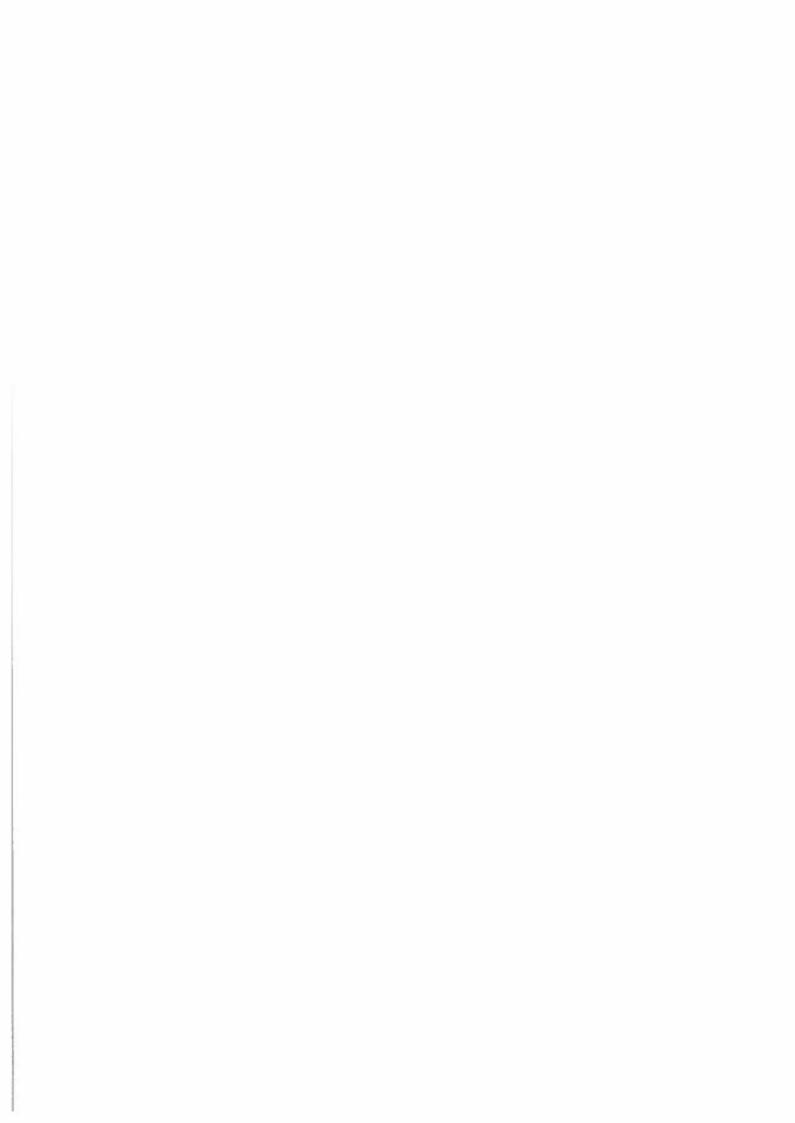


Denna modell kommer att ha samma TES som
TES = 109000 efterson alla ses at med men den
kommer att ha RSS = RSS + RSS eftersom Den genom
De kommer att bilda clika etrationer för period i rap
26 RSS = RSS + RSS = 27000. K=6 n=40
Vi genomfor ett F-test på help 120 deller: H: B=B=B=B=B=D H: Sa an of facul V=005. Statistika: F = \frac{F(S)}{885/34} \rangle F(S; 34) om H. in 3ann.

Best treget: H. torsartan om F(S; 34) = \frac{F(S; 30)}{800} = \frac{2}{53}

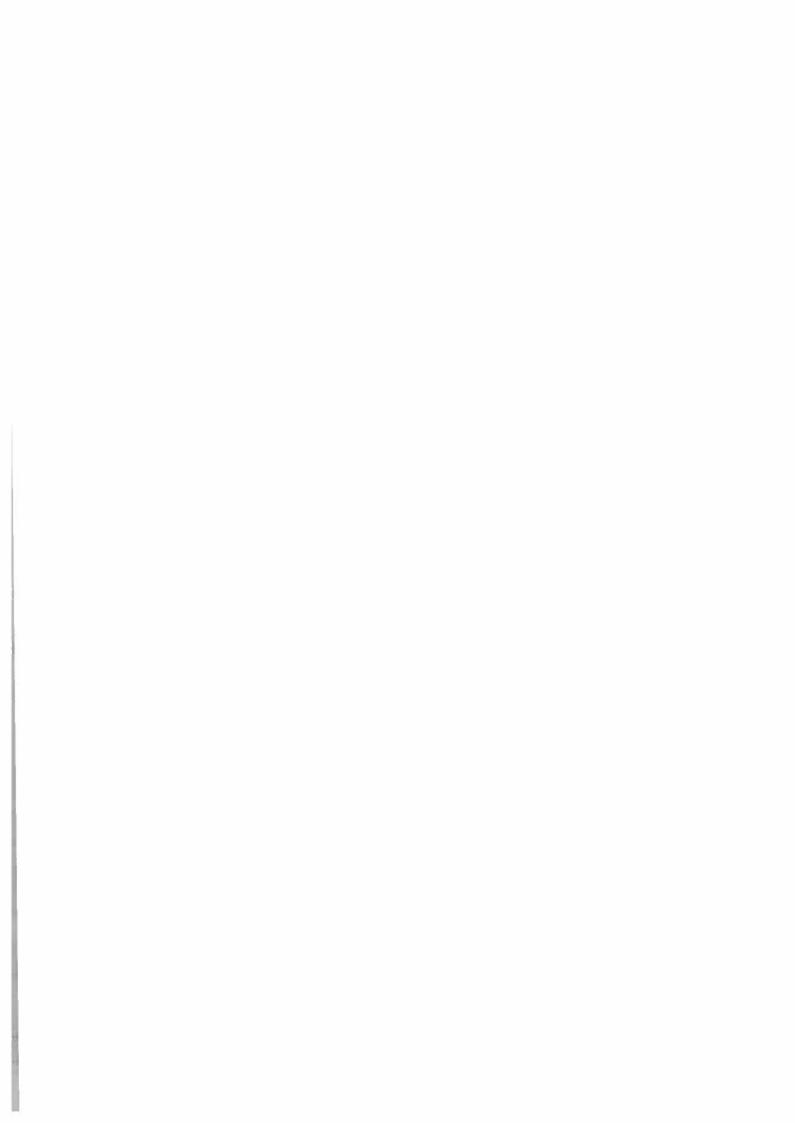
Resultat: For \frac{2}{27000/34} = \frac{20}{40} = \frac{5}{600} = \frac{5}{34}.

H. torsartas Par \$ \frac{5}{600} = \frac{2}{1000/34} = \frac{20}{1000/34} = \frac{20}{10000/34} = \frac{20}{10000/34} = \frac{20}{10000/34 Slutterts: Denna modell forllarar atminstre en



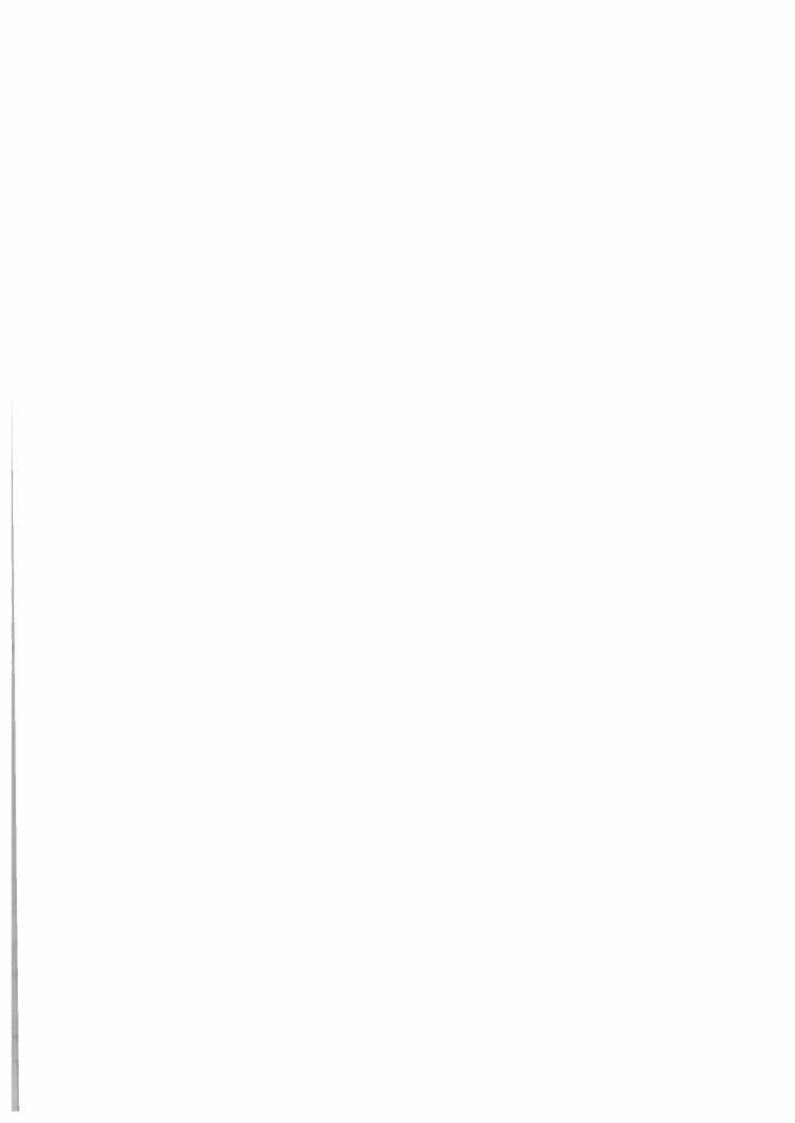
Room: Brunharitagales Anonymous code: FK1-HEC-XKZ Sheet number: 3-

3. Det falska	pastaendet an 6) Eice = 0 (altri).
Detta efters	on vi inte han vota wi salt
Vi har bara	eligien ster sant som i salt eligien etterson shutte det (dans Zu;=0).
8 var: 6)	(etaral 20; -0):



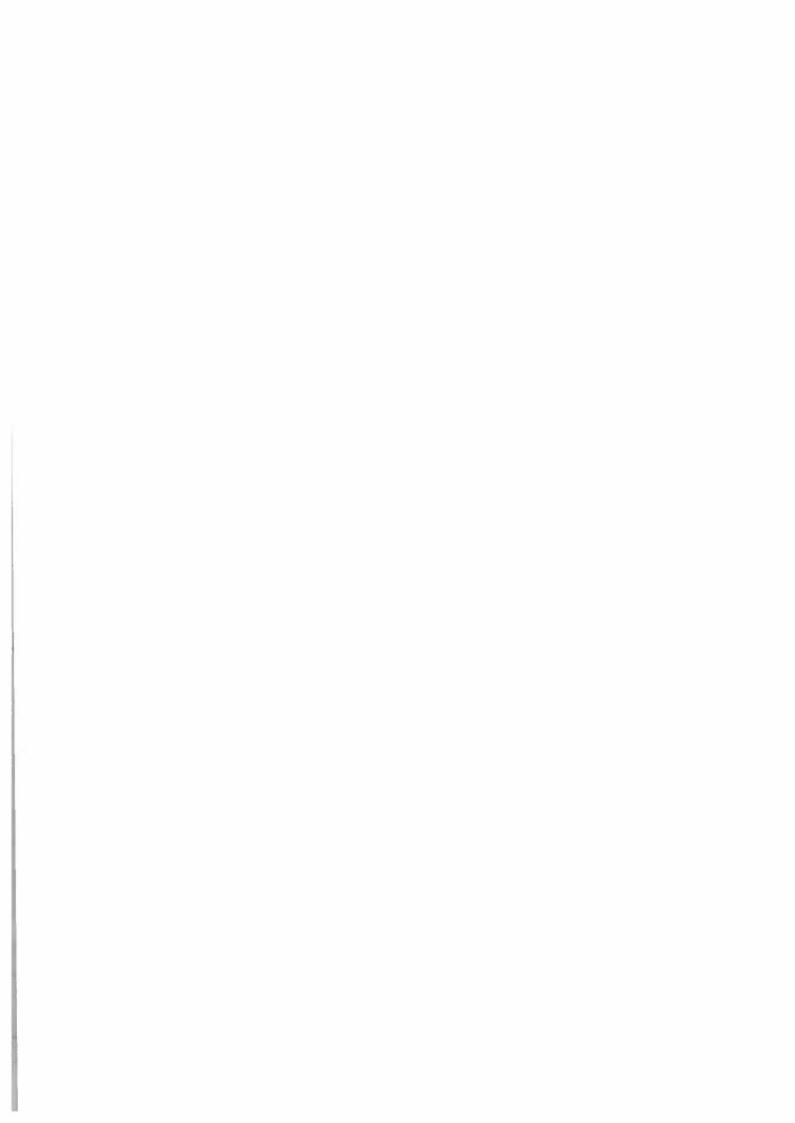
Room: Brannsvikssalen Anonymous code FKI-HEC-XKZ Sheet number: 4.

7.0% 17.	a ned 1?
(1 P) X=1+1 = 2,40+030(4+	(-1)
(1 P) X=1+1 = 2,40+030(4+	(-1)
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	
$= \frac{(1-\beta) X=A }{2} = \frac{0.30}{5} = \frac{1.35}{5} = \frac{0.005}{5} = \frac{1.35}{5}$	t the alltgå 35% stone
b) at sant.	2 hørt talas om varmanhet 90 for vorja dag. 8 m: b).
(1-0,5)	$=0=-2,40\pm0.30x$
50%. Person att ha	len shattarle sarun let heten hent talas on vara nachet



Room: Burnsvikscalen Anonymous code: EKI-HEC-XKZSheet number: 5.

	On det inte finns nagon satornes Vigintion sa a. Ba=B=B=O i modell 2. For att se our sa	
1	By= 3= B=0 i modell 2. For att se om so fallet gos vi ett F-text dan vi janto den ent	ā-
	4: B=B=5=0 H: Sa a c; fallet. A=00	
	Teststatistica: F=(RSSp-RSSon)/m 2 F(m)/n-1) on the al-	Exhair
	RSS_R=RSS_=4110,19301 RSS_R=RSS_=1302,75351	
	$m = 3$ $n - k = 48 - 5 = 43$ $\approx E(3; 10) =$	284
	Besluteregel: Ho to hastas on For (3:43) ac (3:69=	2,76
	Resultat: F= (4110,1936,1-1802 7535)/3 30,894>F05(3,43)	
	1071025103 Pa 3 10 51711- 110-a.	
	Slutzatz : Det finns stok for sasonasvariation.	
)	Modellen i Durbin-Watson-testet a:	
	$u = \rho u + e_t$, due ρ as leone lationen mellan $u = \frac{2\rho_2(\hat{u} - \hat{u}_t)^2}{2\rho_2(\hat{u} - \hat{u}_t)^2}$	ya.
	mellas o och 4 dan vanden nava o indike positi autohorrelation och varden mara 4 nega D-W testan for torsta ordningens autohom	in.
	Modeller i Breusch-Grafreys-testet at:	relation
_	4=P, UL, +P, UL, 2+P, UL, 2+ +P, UL,	
		d
	Ho and p = p = p3 = = p = 6 och testas me statistican nR2 (tran regressionen an ar pr regressioner na som givit residentemper och a, a, a, a, a, a, at approx ×2(k)-tordelad B g testar for autokondation upp till ordning k	0
	By testar for autohoralation upp till ordning K	
	1 lette CN Singe in at still for forthe ording	E41. C
	Darenot visa B- G att det Finne Stid (a=0.05)	15
	(p-value = 0,0347).	/



Stockholm University

Department of Statistics

Econometrics I

WRITTEN EXAMINATION

Tuesday November 28 2017, 10 am - 15 pm

Tools allowed: Pocket calculator

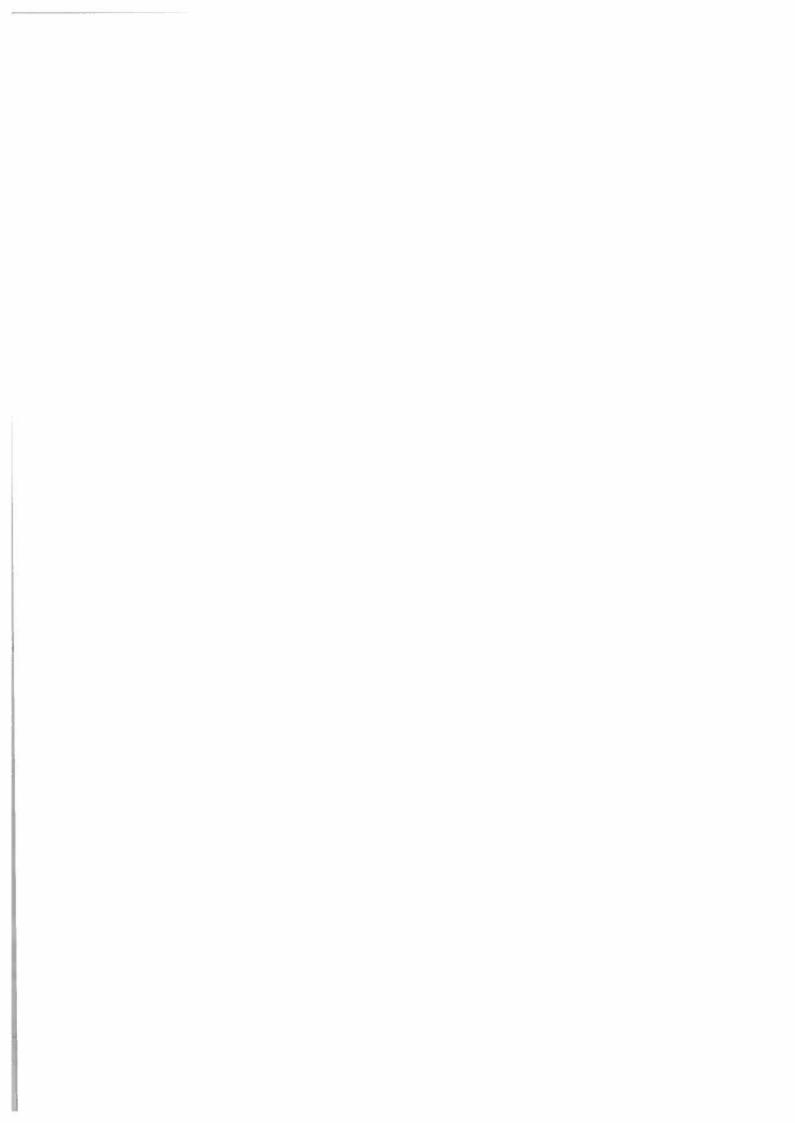
Passing rate. 50% of overall total, which 100 points. For detailed grading

Criteria, see the course description.

The exam will be handed back: not decided

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

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



Task 1 (22 points)

Assume the model $Y_t = \beta t + u_t$, where t = 1, 2, 3, ..., T and that you want to estimate β . You are suggested the following estimators:

$$b = \frac{\sum_{1}^{T} Y_{t}}{\sum_{1}^{T} t} = \frac{\bar{Y}}{\bar{t}}$$
 and $b^{*} = \frac{Y_{T}}{T}$.

The estimator b is known to be unbiased with variance: $\sigma_b^2 = \frac{\sigma^2}{T\bar{t}^2} = \frac{\sigma^2}{T\frac{(T+1)^2}{4}}$.

- A. Show that the estimator b^* also is unbiased.
- B. Derive the variance of b^* assuming constant error variance (σ^2)
- C. Which estimator of the two is the most efficient one? Demonstrate for T=5.
- D. The estimator b is the GLS-estimator when the error variance σ^2 is proportional to t. Present under this assumption the transformed model for which the assumption of homoscedastic error term is fulfilled. This task is independent of task A-C.

Task 2. (34 points)

From a regression analysis we collect the following information.

Model:
$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

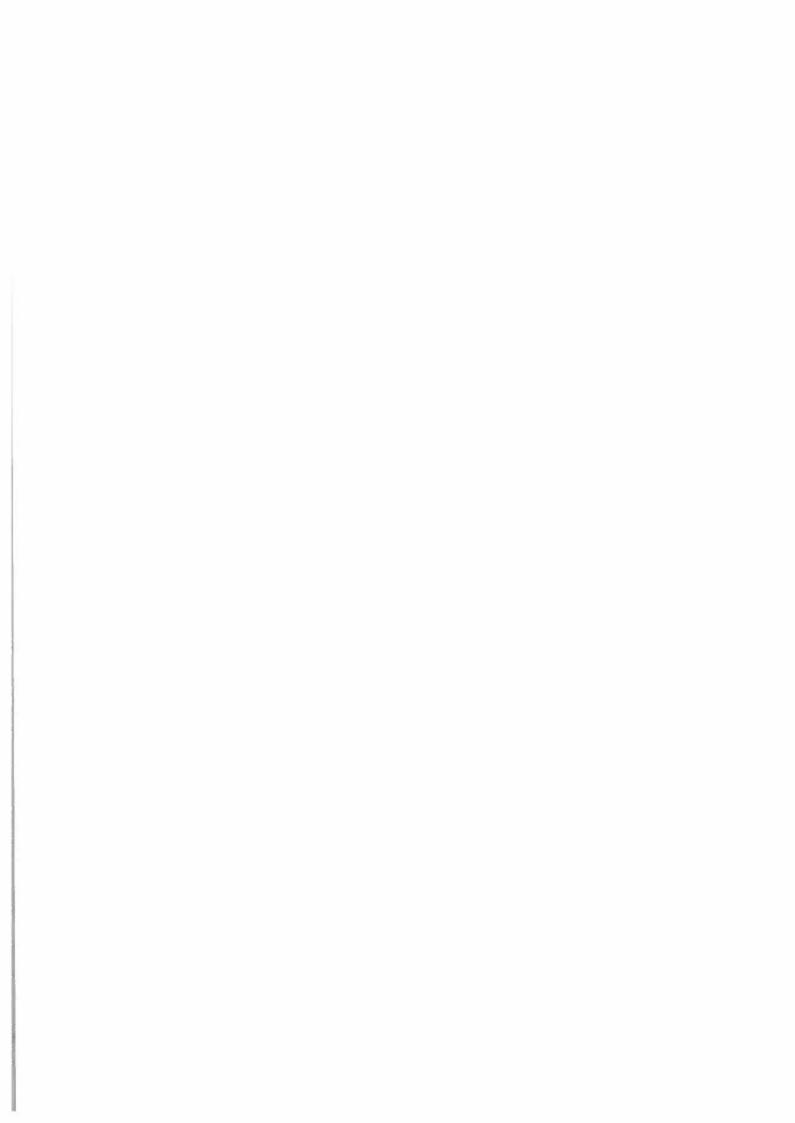
	Period 1	Period 2	Period 1+2
TSS	46000	42000	108000
RSS	15000	12000	35000
N	20	20	40

The estimation results for the period 1 and period 2 are based on two independent regressions. The results from period 1+2 are obtained using all 40 observations.

- A. Are the parameters of the model the same for the two periods? Perform a formal test. Your solution should include null- and alternative hypothesis in terms of the beta parameters, test statistic, its distribution with specification of degrees of freedom, decision rule, results and a conclusion.
- B. Is the assumption of equal disturbance variance fulfilled in task A? Perform a formal test. Your solution should include null- and alternative hypothesis, test statistic, its distribution with specification of the degrees of freedom, decision rule, results and a conclusion.
- C. Suppose we estimate the Model:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 D_t + \beta_5 X_{2t} D_t + \beta_6 X_{3t} D_t + u_t$$
 on the same data, where $D_t = 0$ for period 1 and equal to 1 for period 2. Make a formal test of whether this model explain at least some of the variation in Y_t .

The tasks A, B and C can be solved independently of each other.



Task 3 (6 points)

Which of the following statements is not correct for a simple linear regression model with an intercept:

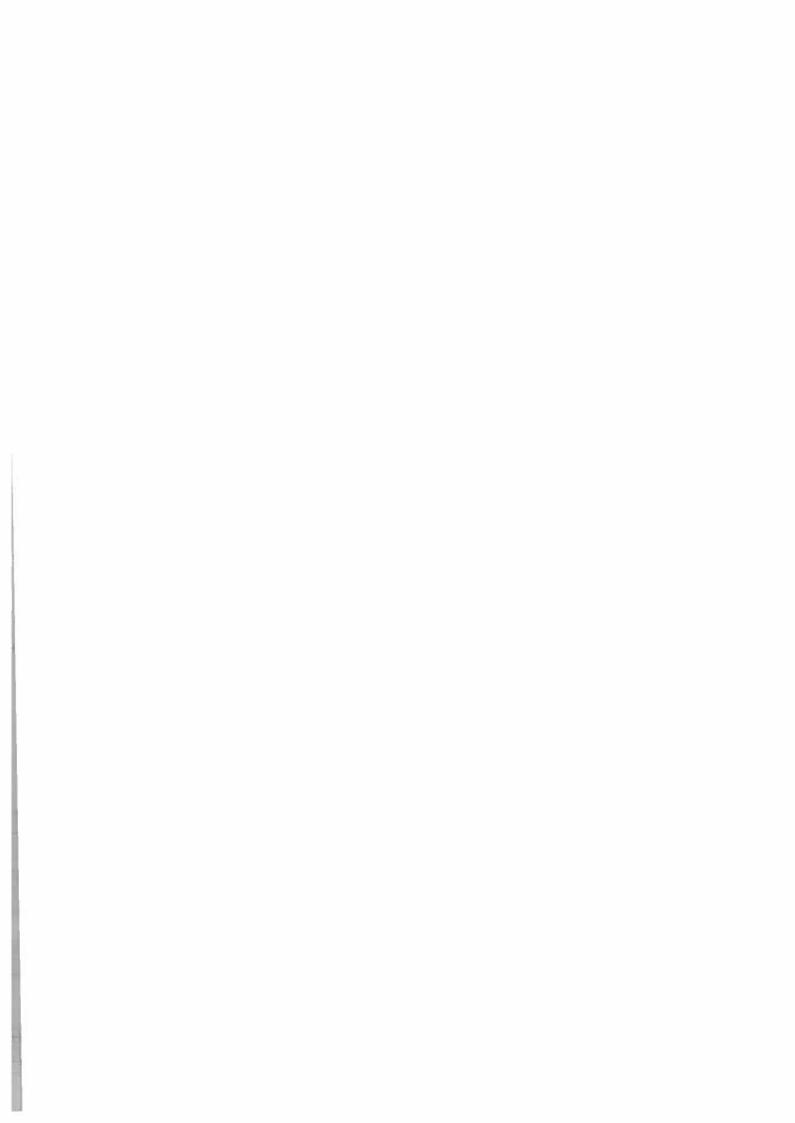
- a) $\sum \hat{u}_i =$ is always equal to 0
- b) $\sum u_i$ = is always equal to 0
- c) $\sum \hat{Y}_i = \sum Y_i$ (always)
- d) $\sum X_i \hat{u}_i = \text{is always equal to } 0.$
- e) $\sum \hat{Y}_i \hat{u}_i =$ is always equal to 0.

Task 4 (12 points)

A new brand is introduced on a market. This is followed up by an advertising campaign. During a period after the introduction the market penetration is measured in the target group. Let Y=1 if a person have heard of the brand and let Y=0 if not. Let X be the number of days after the introduction. On the basis of collected data a logistic regression were performed with the following result:

$$\ln\left(\frac{\hat{P}}{1-\hat{P}}\right) = -2.40 + 0.30X$$
 $P = P(Y=1)$

- A. Which of the following statements is correct?
- a) The ODDS that a person has heard of the brand is estimated to increase with 0.30 units for each additional day since the introduction.
- b) The ODDS that a person has heard of the brand is estimated to increase with 35% for each additional day since the introduction.
- c) The probability that a person has heard of the brand is estimated to increase with 0.30 units for each additional day since the introduction.
- d) The probability that a person has heard of the brand increases with 0.30% for each additional day since the introduction.
- e) The probability that a person has heard of the brand is estimated to increase with 35% for each additional day since the introduction.
- B. According to the estimated model, after how many days is the probability that a person has heard of the brand 50%?

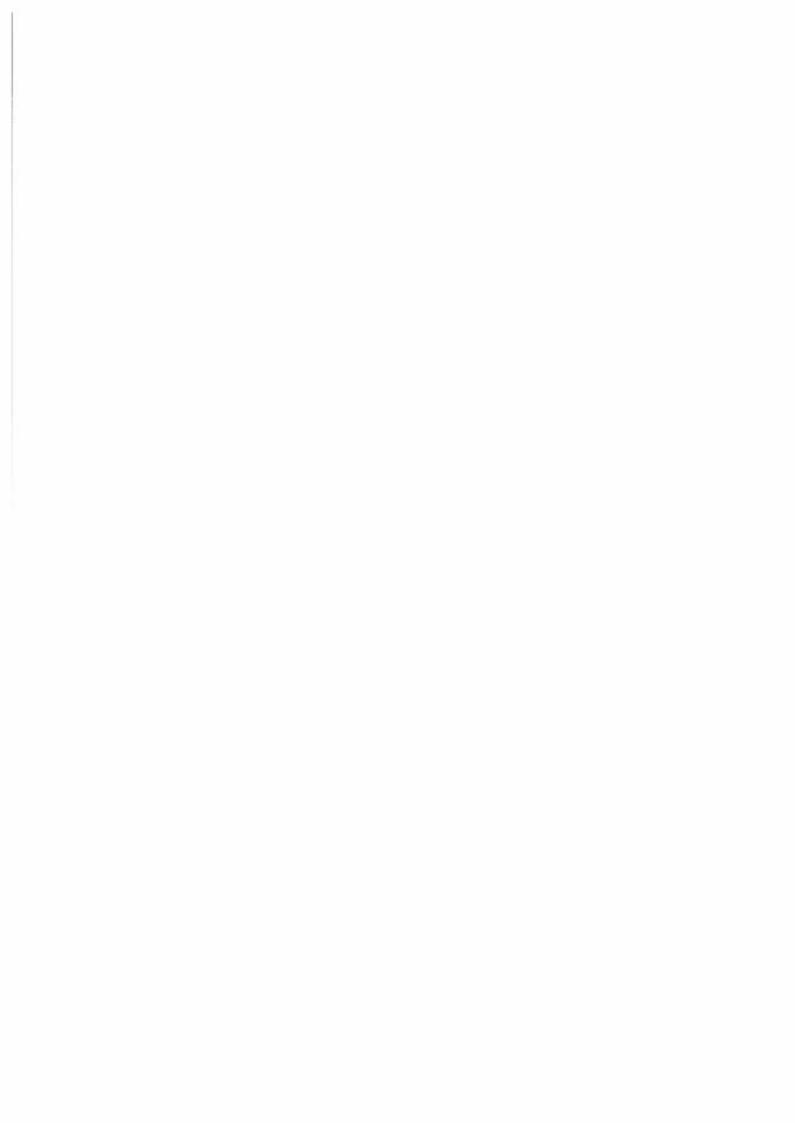


Task 5 (26 points)

Given: The quarterly data below and the regression outputs on the next page

time	Dlt	D2t	D3t	D4t	Yt
1	1	0	0	0	110.2
	0	i	0	0	134.9
2 3	0	0	i	ő	128.4
4	0	0	ò	ĭ	117.7
-					117.7
5	1	0	0	0	117.5
6	0	1	0	0	137.9
7	0	0	1	0	144.4
8	0	0	0	1	126.9
9	1	0	0	0	115.7
10	0	1	0	0	142.5
11	0	0	1	0	143.8
12	0	0	0	1	125.1
13	ì	0	0	0	138.6
14	0	Ĭ	0	ő	135.3
15	0	0	1	0	149.9
16	0	0	0	1	138.2
17	1	0	0	0	143
18	0	1	0	0	150.8
19	0	0	1	0	153.7
20	0	0	0	1	147.6
21	1	0	0	0	145
22	0	1	0	0	161.4
23	0	0	Ĭ	0	152.1
24	0	0	ò	ĭ	142.8
25	1	0	0	0	144.9
					163.4
26	0	1	0	0	
27	0	0	1	0	167.7
28	0	0	0	1	160.4
29	1	0	0	0	157.9
30	0	1	0	0	164.6
31	0	0	1	0	167.2
32	0	0	0	1	154.2
33	1	0	0	0	156.7
34	0	1	0	0	174
35	0	0	Ĭ	Ö	190.5
36	0	0	0	ĭ	169.1
37	1	0	0	0	151.8
38	0	1	0	0	174.6
39	0	0	1	0	179.1
40	0	0	0	1	166.1
41	1	0	0	0	170.4
42	0	1	0	0	192.5
43	0	0	1	0	187.3
44	0	0	0	1	170.3
45	Ĭ	0	Ö	0	171.2
46	0	ĭ	0	ő	194.2
47	0	0	1	0	198.8
48	0	0	0	1	184
40	U	υ	v	1	104

- A. Perform a formal test of whether there is any seasonal variation in Y_t.
- B. In connection with the estimation results you find information about potential autocorrelation. Present the models for ut for the Durbin Watson test and the Breusch Godfrey test and the corresponding null hypothesis (in terms of the parameters of the models) for the two test. Comment shortly (at most two sentences per test) on the results of the tests.



. regress Yt time

Source SS	df	MS			obs = 48
Model 18202.118 Residual 4110.1936	36 l 01 46	18202 89.35	20219	Prob > F R-squared Adj R-squa	= 203.71 = 0.0000 = 0.8158 ared = 0.8118 = 9.4526
Yt Coef.	Std. Err.		10.00	[95% Con	_
time 1.405672		14.27 43.30	0.000	1.20743 114.446	1.603914 125.6052

Durbin-Watson d-statistic = 1.950222.

regress Yt time D2t D3t D4t

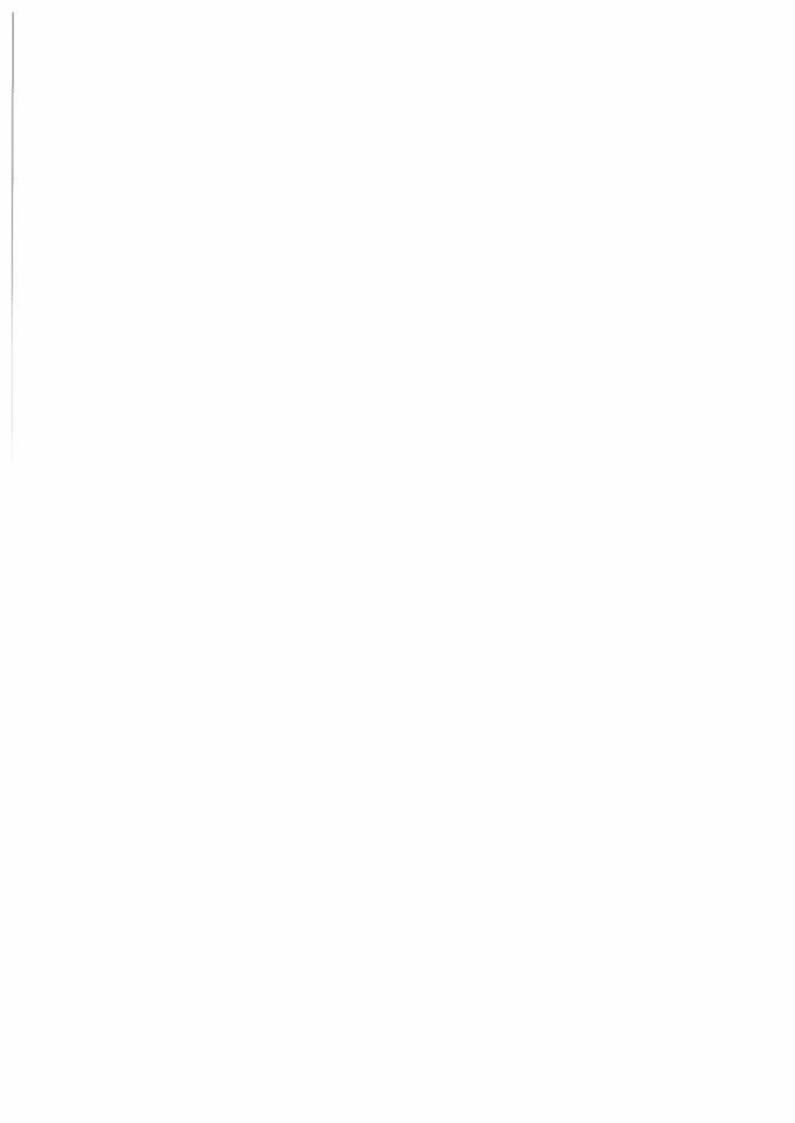
Source	SS	df MS	_	Number of F(4, 43)	
Model 21 Residual 13		52.38953 2965932	Prob > F	= 0.0000 $= 0.9416$	
Total 223	312.3116	47 474	.730035	Root MSE	= 5.5042
Yt Co		l. Err. t	P> t	[95% Conf.	Interval]
D2t 15.		5359 24.3 7831 6.9 0039 7.6	0.000	1.283815 11.0003 12.66267	1.515879 20.06667 21.73794

Durbin-Watson d-statistic = 1.865191

Breusch-Godfrey LM test for autocorrelation

lags(p)	chi2	df	Prob > chi2
K=5	12.007	5	0.0347

H0: no serial correlation



Prel. Solutions to exam in Econometrics I November 2017.

Task 1

A)
$$b^* = \frac{Y_T}{T} = \frac{\beta T + u_T}{T} = \beta + \frac{u_T}{T}$$
 $E(b^*) = \beta + \frac{E(u_T) = 0}{T} = \beta$ Hence unbiased

B)
$$V(b^*) = V\left(\beta + \frac{u_T}{T}\right) = V\left(\frac{u_T}{T}\right) = \frac{\sigma^2}{T^2}$$

C)
$$\sigma_b^2 = \frac{\sigma^2}{T^{\frac{(T+1)^2}{4}}} = \frac{\sigma^2}{5^{\frac{(5+1)^2}{4}}} = \frac{\sigma^2}{45}$$
 $V(b^*) = \frac{\sigma^2}{5^2} = \frac{\sigma^2}{25}$ Of the two estimators b has the

smallest variance and is therefore more efficient than b*.

D)
$$\frac{Y_t}{\sqrt{t}_t} = \beta \sqrt{t} + \frac{u_t}{\sqrt{t}}$$
 $(V(\frac{u_t}{\sqrt{t}}) = \frac{\sigma^2 t}{t} = \sigma^2)$

Task 2

A) Model period 1: $Y_t = \beta_{11} + \beta_{21}X_{2t} + \beta_{31}X_{3t} + u_t$

Model period 2: $Y_t = \beta_{12} + \beta_{22}X_{2t} + \beta_{32}X_{3t} + u_t$

 H_0 : $\beta_{11} = \beta_{12}$, $\beta_{21} = \beta_{22}$, $\beta_{31} = \beta_{32}$

H₁: At least one of the restrictions in H₀ is false Significance level: 5%

Test statistic: $\frac{(RSS_R - RSS_{UR})/k}{RSS_{UR}/(n-2k)}$ F-distributed with k=3 and 40-2*3=34 d.f. given u IN(0, σ^2) for both periods.,

Decision rule: Reject H₀ if $F_{obs} > F_{0.05, 3, 34} = ??$ $F_{3.30} = 2.92$ $F_{3.40} = 2.84$

$$F_{obs} = (35000-27000)/3/(27000/(40-6)) = 3,36$$

The result is significant. All parameters do not seen to be the same for the two periods.

B) Denote the two periods error variances by σ_1^2 and σ_2^2 .

$$H_0$$
: $\sigma_1^2 = \sigma_2^2$ H_1 : $\sigma_1^2 \neq \sigma_2^2$ Significance level: 10%

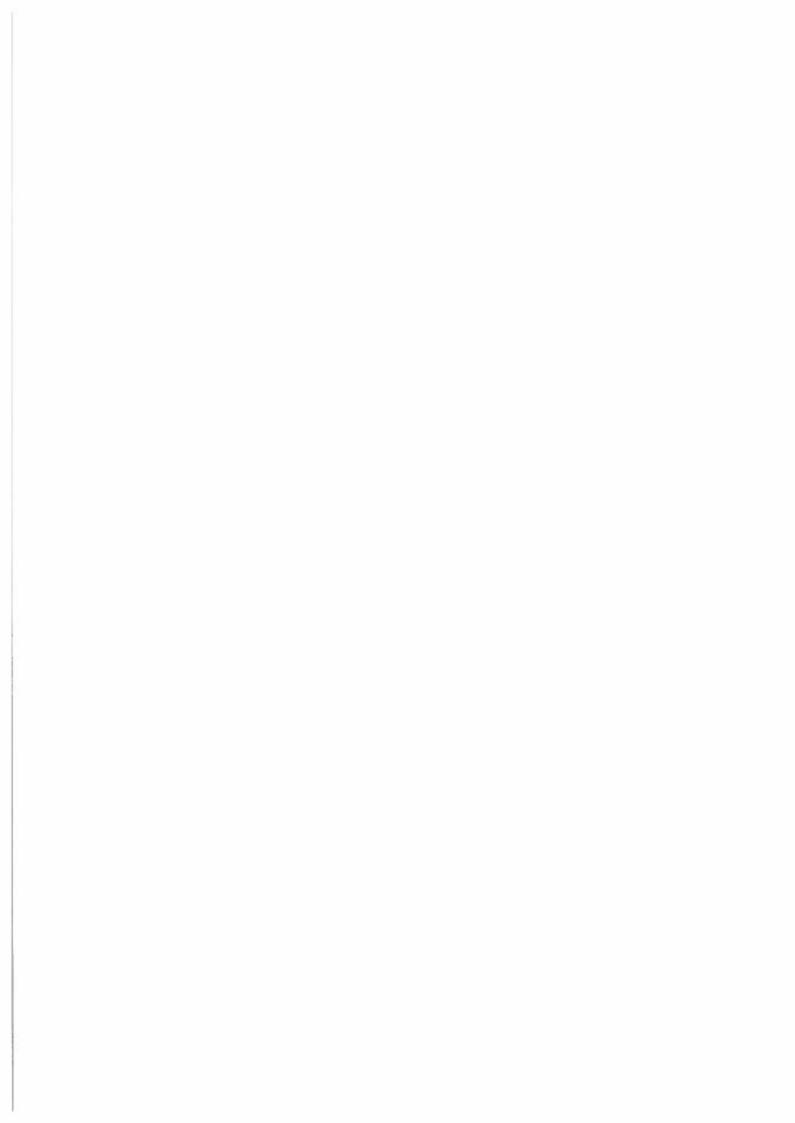
Test statistic: $\frac{s_1^2}{s_2^2}$ F-distributed with n_1 -k=17 and n_2 -k=17 d.f. given u IN(0, σ^2)

Since we have put the largest residual variance in the numerator we only need to study the rejection region to the right.

Reject H₀ if F_{obs} is larger than $F_{0.05, 17, 17} = ?? F_{15, 17} = 2,31 F_{20, 17} = 2,23$

 F_{obs} = 15000*17/(12000*17)=1,25 non sign .

We don't get any support for unequal error variance on 10% significance level.



C. Model: $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 D_t + \beta_5 X_{2t} D_t + \beta_6 X_{3t} D_t + u_t$

 H_0 : $\beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$ H_1 : At least one of β_2 , β_4 to β_6 differs from zero

Significance level: say 5%

Test statistic: $\frac{ESS/(k-1)}{RSS/(n-k)}$ that is F-distributed with k-1=5 and n-k=34 df if H₀ is true and u_t IN(0, σ^2).

Decision rule: Reject H₀ if $F_{obs} > F_{0,05,5,34} = ?? F_{5,30} = 2,53 F_{5,40} = 2,45$.

ESS=TSS-RSS=108000-15000-12000=81000, RSS=15000+12000=27000

 F_{obs} =81000*34/(27000*5)=20,4. The result is significant and strongly indicates that the model has at least some explanatory power.

Task 3 Answer: Alternative b

Task 4 A) Alternative b

B) In order for P(Y=1) must $e^{-(-2,4+0,30X)}=1$, that is -2,4+0,3X=0, which gives X=8.

Task 5

A: Model: $Y_t = \beta_1 + \beta_2 time + \beta_3 D2_t + \beta_4 D3_t + \beta_5 D4_t + u_t$

 H_0 : $\beta_3 = \beta_4 = \beta_5 = 0$ H_1 : At least one of β_3 , β_4 and β_4 differs from zerol

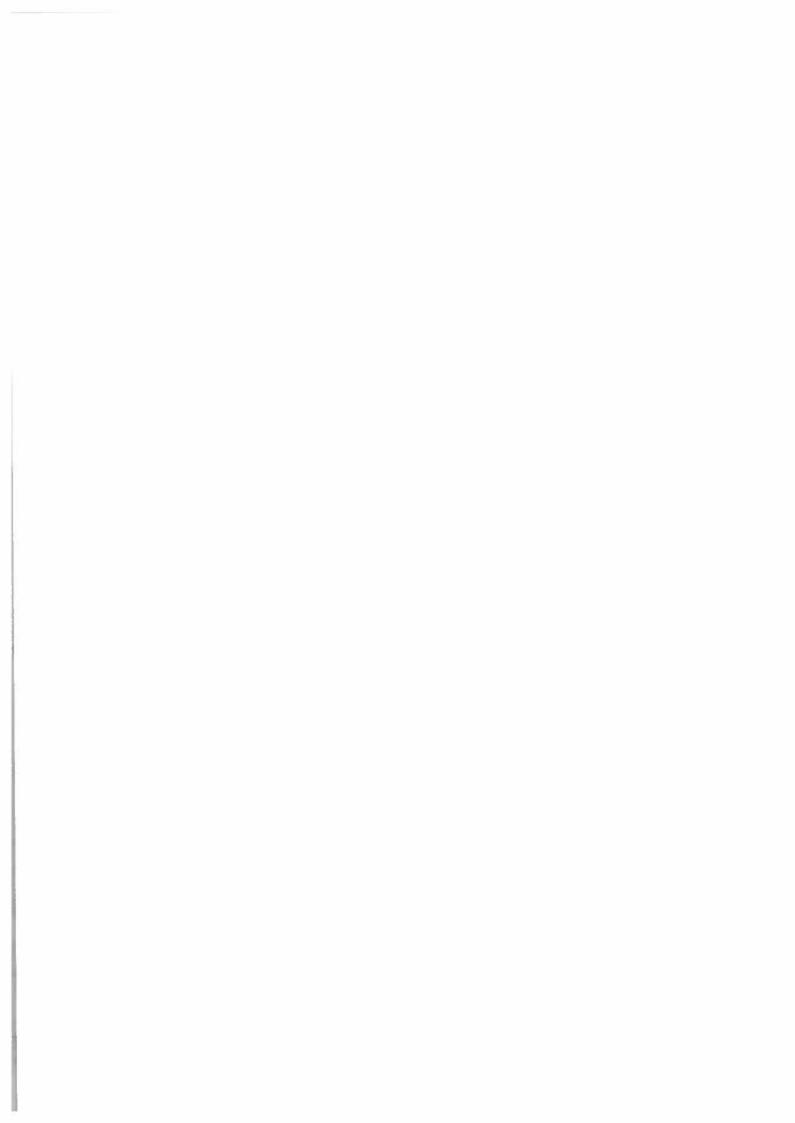
Signifikanslevel: say 5%

Test statistic: $F=((RSS(R) - RSS_{LIR})/m)/s^2_{LIR}$

F-distributed with m=3 och 48-5=43 d.f. given u IN(0, σ^2),

Reject H_0 if $F_{obs} > F_{0,05,3,43} = ??$ $F_{3,40} = 2,84$

Result: $F_{obs} = \frac{(4110.19301 - 1302.75351)/3}{5,5042^2} = 30,9$ Significant result. We can be pretty sure of existence of seasonality in the data.



B) Durbin Watson test:

$$u_t = \rho u_{t-1} + \varepsilon_t$$

 H_0 : $\rho=0$ H_1 : $\rho>0$ If we test for positive autocorrelation of first order.

Reject H₀ if d<d₁ do not reject if d>d_U

Breusch Godfreys test:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \rho_4 u_{t-4} + \rho_5 u_{t-5} + \varepsilon_t$$

 H_0 : $\rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho_5 = 0$ against H_1 : At least one of $\rho_k \neq 0$, k=1, 2,...,5.

Comments on the d-test:

The value on the Durbin Watson test statistic is close to 2 (1,95 and 1,87) and hence is clearly non-significant ($d_{U,k'=1}=1,6$; $d_{U,k'=4}=1,72$. We have not detected any autocorrelation of order 1 in the two models.

Comments on the Breusch Godfreys test:

The p-value of the test is equal to 0.0347, which means significant results on the 5% level. This suggests the existence of higher order autocorrelation up to order 5 and that the model 2 can be improved.

