

STOCKHOLM UNIVERSITY
Department of Statistics
Econometrics I, Regression analysis, ST223G
Autumn semester 2019

Written Examination in Econometrics I

Date 2019-12-03
Hour: 10.00-15.00
Examiner: Jörgen Säve-Söderbergh
Allowed tools: 1) Textbook: Wooldridge, J.M. *Introductory Econometrics: A Modern Approach*, Cengage.
2) Pocket calculator
3) Notes written in the text book are allowed.

- Note that no formula sheet is provided.
- Passing rate: 50% of overall total, which is 100 points. For detailed grading criteria, see the course description.
- The maximum number of points for each problem is stated after each question. If not indicated otherwise, to obtain the maximum number of points on each problem, detailed and clear solutions are required. Answers may be given in English or Swedish.

Good luck!

1. This question concerns a data set that includes information on wages, education, parents' education and several other variables for 1230 working men in 1991. We also have a measure of cognitive ability (*abil*) and the costs for entering college education in 2017 and 2018.

(a) The model

$$educ = \beta_0 + \beta_1 motheduc + \beta_2 fatheduc + \beta_3 abil + \beta_4 abil^2 + u$$

was estimated using R:

Coefficients:

	Estimate	Std. Error	t value
(Intercept)	8.240226	0.287410	28.671
motheduc	0.190126	0.028096	6.767
fatheduc	0.108939	0.019601	5.558
abil	0.401462	0.030288	13.255
I(abil^2)	0.050599	0.008304	

Residual standard error: 1.758 on 1225 degrees of freedom

Multiple R-squared: 0.4444, Adjusted R-squared: 0.4425

F-statistic: 244.9 on 4 and 1225 DF, p-value: < 2.2e-16

Test the null hypothesis that *educ* is linearly related to *abil* against the alternative that the relationship is quadratic. (10 p)

- (b) Using the equation in (a), test $H_0 : \beta_1 = \beta_2$ against a two-sided alternative. State the null and alternative clearly. Use 5% significance level. [Hint It is necessary to know that $\widehat{\text{Cov}}(\hat{\beta}_1, \hat{\beta}_2) = -0.0002928221$.] (10 p)
- (c) The costs for entering college education in 2017 and 2018 were added to the model in (a). Do a formal hypothesis testing to determine whether they are jointly statistically significant or not. Use 5% significance level. (10 p)

2. A telephone survey attempted to elicit the demand for a (fictional) "ecologically friendly" apple. Each family was (randomly) presented with a set of prices for regular apples and the eco-labeled apples. They were asked how many pounds of each kind of apple they would buy. We use the following variables from this survey

- *ecolbs* = the amount ecological apples that you would buy at a certain price.
- *ecoprc* = price of ecological apple
- *regprc* = price of regular (not ecological) apple
- *faminc* = family income
- *hsize* = household size

a) Does the variable *ecolbs* seem to have a continuous distribution over strictly positive values? What implications does your answer have for the suitability of a Tobit model for *ecolbs*? The variable *ecolbs* is illustrated in figure 1 below. The number of households that did not demand any amount of ecological apples were 248. (10 p)

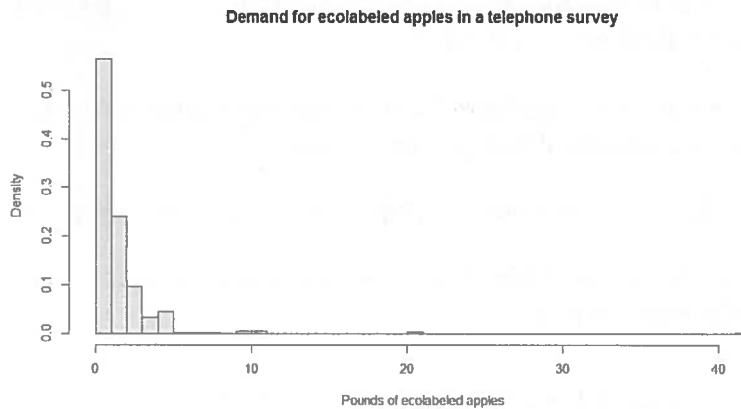


Figure 1: Relative histogram for part a)

b) A Tobit model was estimated using R. Write the estimated equation. Which variables seem to influence the dependent variable? Discuss this question without a formal hypothesis testing framework. (10 p)

Call:

```
tobit(ecolbs ~ ecoprc + regprc + faminc + hhsize)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.002687	0.667972	1.501	0.1333
ecoprc	-5.821484	0.885840	-6.572	4.97e-11 ***
regprc	5.655171	1.064647	5.312	1.09e-07 ***
faminc	0.006638	0.004001	1.659	0.0971 .
hhsize	0.130315	0.095038	1.371	0.1703

Gaussian distribution

Number of Newton-Raphson Iterations: 3

Log-likelihood: -1266 on 6 Df

Wald-statistic: 48.4 on 4 Df, p-value: 7.778e-10

- c) Are the signs of the coefficients on the price variables from (b) what you expect? Explain. (10 p)
3. The following model was suggested for studying expenditures on marketing for Swedish municipalities (*sv. kommuner*).

$$\text{Marketing} = \beta_0 + \beta_1 \text{Population} + \beta_2 \text{Tax} + \beta_3 \text{ProportionYoung} + u$$

From previous studies we know that it is reasonable to assume the following on the error term u

$$E(u | \text{Population}, \text{Tax}, \text{ProportionYoung}) = 0$$

and

$$\text{Var}(u | \text{Population}, \text{Tax}, \text{ProportionYoung}) = \sigma^2 \text{Population}^2$$

Are you satisfied with the model and the assumptions? If not, please explain what the problem consists of and demonstrate how we can solve the problem in this case. (20 p)

4. Let *corn* denote per capita consumption of corn in bushels at the county level, let *price* be the price per bushel of corn, let *income* denote per

capita county income, and let *rainfall* be inches of rainfall during the last corn-growing season. The following simultaneous equations model imposes the equilibrium condition that supply equals demand:

$$\text{corn} = \alpha_1 \text{price} + \beta_1 \text{income} + u_1$$

$$\text{corn} = \alpha_2 \text{price} + \beta_2 \text{rainfall} + \gamma_2 \text{rainfall}^2 + u_2$$

The variables *income*, *rainfall* and *rainfall*² are considered exogenous in the model. Is the model identified, that is, are both equations identified? (20 p)

STOCKHOLM UNIVERSITY
Department of Statistics
Econometrics I, Regression analysis, ST223G
Autumn semester 2019

Addendum to Written Examination in Econometrics I

On problem 1c) the information that $R_{ur}^2 = 0.4451$ was forgotten.



Stockholms
universitet

Department of Statistics

Correction sheet

Date: 191203

Room: Brunnsvikssalen

Exam: Econometrics I

Course: Econometrics

Anonymous code:

0042-JMZ

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						7
Teacher's notes	30	30	20	20						

Points	Grade	Teacher's sign.
100	A	JSS

1.

$$a) \text{educ} = \beta_0 + \beta_1 \text{motheduc} + \beta_2 \text{fatheduc} + \beta_3 \text{abil} + \beta_4 \text{abil}^2 + u$$

$$t \text{ value } (I(\text{abil}^2)) = \frac{0.050519}{0.008304} \approx 6.093 \quad P \quad (I(\text{abil}^2)) \text{ är signifikant}$$

df: 1725 $H_0: \beta_4 = 0$ $H_1: \beta_4 \neq 0$ \rightarrow skiljt ifrån 0 på alla rimliga nivåer för α .

Man att $I(\text{abil}^2)$ har en signifikant påverkan på "educ" och har samma tecken som abil (sida positiv) så kommer förhållandet mellan "ability" och "education" vara kvadrerat.

och med att båda koefficienterna är positiva innebär det även att förhållandet inte har en maximipunkt givet att man inte kan erhålla "negativa" utbildningar, vilket känns som ett rimligt antagande.

$$1. b) \quad \text{educ} = \beta_0 + \beta_1 \text{motheduc} + \beta_2 \text{fatheduc} + \beta_3 \text{abil} + \beta_4 \text{abil}^2 + u$$

$$H_0: \hat{\beta}_1 - \hat{\beta}_2 = 0 \quad H_A: \hat{\beta}_1 - \hat{\beta}_2 \neq 0 \quad \alpha = 0,05$$

$$t = \frac{\hat{\beta}_1 - \hat{\beta}_2}{\text{se}(\hat{\beta}_1 - \hat{\beta}_2)} \quad \text{se}(\hat{\beta}_1 - \hat{\beta}_2) = \sqrt{\text{Var}(\hat{\beta}_1 - \hat{\beta}_2)}$$

$$\begin{aligned} \text{Var}(\hat{\beta}_1 - \hat{\beta}_2) &= \text{Var}(\hat{\beta}_1) + \text{Var}(\hat{\beta}_2) - 2\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) \\ &= \text{se}(\hat{\beta}_1)^2 + \text{se}(\hat{\beta}_2)^2 - 2\text{Cor}(\hat{\beta}_1, \hat{\beta}_2) \end{aligned}$$

$$= 0,028096^2 + 0,019601^2 - 2(-0,000292821)$$

$$\approx 0,0017592$$

$$\Rightarrow \text{se}(\hat{\beta}_1 - \hat{\beta}_2) = \sqrt{0,0017592} \approx 0,04194$$

$$\Rightarrow t = \frac{0,190126 - 0,108439}{0,04194} \approx \underline{1,936}$$

kritiskt värde ($\alpha=0,05$) = 1,96 t är inte signifikant skillnad från 0.

Ho förkastas inte på $\alpha=0,05$

SU, DEPARTMENT OF STATISTICS

Room: B-vik Anonymous code: 0047-212 Sheet number: 3

1. c)

F-test

$$R_{\text{old}}^2 = 0.4451$$

$$\alpha = 0.05$$

$$R_{\text{R}}^2 = 0.4444$$

$$F = \frac{(R_{\text{old}}^2 - R_{\text{R}}^2)}{(1 - R_{\text{old}}^2)} \cdot \frac{df_{\text{un}}}{q}$$

$$df_{\text{un}}: n - k - 1 = 1230 - 6 - 1 = 1223$$

$$q = 2$$

$\left. \begin{array}{l} \text{kostnaden för collage 2017} : \beta_5 \\ \text{--- " --- " 2018} : \beta_6 \end{array} \right\} k = 6$

$H_0: \beta_5 = \beta_6 = 0$ H_A : inte H_0

$$F = \frac{(0.4451 - 0.4444)}{(1 - 0.4451)} \cdot \frac{1223}{2}$$

$$\approx 0.7714$$

$$\text{kritiskt värde: } 3.00$$

$$(1223, 2, 0.05)$$

H_0 förkastas ej vid $\alpha = 0.05$! Ej signifikant skillnad i β .

Den utökade modellen är inte en signifikant ökad förklaring på "edre" än den begränsade modellen, vilket de borde förstås.

2. a) Variabeln eco (bs) ser ut att vara lognormal och endast anta positiva värden (då det handlar om försäljningsstorlekar är det ganska logiskt).

Tobitmodellen kommer väl till pass i situationer som denna där variabeln endast antar positiva värden samt har en stor andel värden likt 0. I övrigt bör värdena ha en stor spridning.

Huruvida variabeln har en stor icke-negativ spridning (med en stor andel värden likt 0) är svårt att ange i sambandet.

Hur mycket är 40 lbs ekologiska äpplen för ett husstall?

Det är svårt att säga på de del som nämns. Flera andra faktorer kan ha påverkan på mängden äpplen som efterfrågas, ex svart, ekologiska, säsong etc.

Där för har jag svårt att ange en spridning av svar.

Att ca 50% av andelen (som utlöses av 248 st) efterfrågar 0 äpplen är också ett argument för att Tobit-modellen passar.

Ja. R

2. b)

$$y = 1.002687 + (-5.821484)ecoprc + 5.655171reyprc + 0.006638famine + 0.130315hsize + u$$

$u | X \sim N(0, \sigma^2)$ R

$y = \max(0, y^*)$ R

(0.67)
(0.84)
(1.06)
(0.004)
(0.045)
R

"hsize" var ett p-värde på 0.17 vilket endast gör variabeln signifikant på " $\alpha=0.20$ -nivå", och därför ganska ointressant i sambandet. R

Att "ecoprc" har en negativ påverkan på efterfrågan lärars (logiskt), den är också statistiskt signifikant.

"reyprc" verkar också ha en påverkan på efterfrågan, även om den har något större "sci". Variabeln illustrerar antagligen hur konsumenten blir mer beredda att sika sig till andra alternativ när priset på vanliga äpplen ökar, vilket då ökar efterfrågan på ekologiska äpplen.

"Famine" ser ut att ha en väldigt lågt värde på efterfrågan för ekologiska äpplen. Huruvida det beror på egenrådgivning "ekologisk" eller generell korrelation i efterfrågan på brukttypen "äpple" är svårt att säga på.

Vårt att förtydliga är "ecoprc" och "reyprc" p-värden i kontrast till resterande variabler. De är betydligt mer signifikanta i sambandet. R

c) Att priset för ekologiska produkter har en minskande effekt på efterfrågan lärars (logiskt) men kanske på att dessa produkter tenderar att kosta mer än icke-ekologiska alternativ och därmed redan görde ha en efterfrågan som är mer känslig för pris förändringar. R

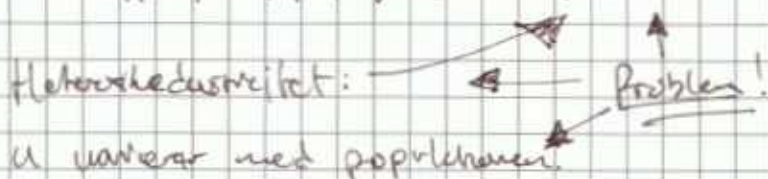
Att priset för "vanliga" äpplen verkar ha en motsatt effekt lärars också (logiskt) då det illustrerar hur marginalen i pris mellan "ekos" och vanliga varor påverkas av både produktens pris.

Variablerna "ecoprc" och "reyprc" relativt låga koefficienter skulle kunna tolkas som att konsumenten ser varorna som goda substitut. R

$$3. \quad \text{Marketing} = \beta_0 + \beta_1 \text{Population} + \beta_2 \text{Tax} + \beta_3 \text{Prop. young} + u$$

$$E(u | \text{pop.}, \text{tax}, \text{prop. young}) = 0$$

$$\text{Var}(u | \text{pop.}, \text{tax}, \text{prop. young}) = \sigma^2 \text{population}^2$$

heteroskedastiskitet:  Problem!
u varierar med populationen.

Vi behöver en modell med homoskedastiskitet för att uppfylla
MLR.5: att OLS-estimatorer har minimal varians för alla estimatorer.

Detta uppnår vi genom att skriva om modellen så
att $\text{Var}(u) = \sigma^2$

$$\text{eller} = \frac{\sigma^2 \cdot \text{POP}^2}{\text{POP}^2} \quad \text{i vårt fall.} \quad \text{Vi kan visa att:}$$

$$\begin{aligned} \text{Var}\left(\frac{u}{\text{POP}}\right) &= \text{Var}\left(\frac{1}{\text{POP}} \cdot u\right) = \frac{1}{\text{POP}^2} \text{Var}(u) = \frac{1}{\text{POP}^2} \cdot \sigma^2 \text{POP}^2 = \\ &= \frac{\sigma^2 \text{POP}^2}{\text{POP}^2} = \sigma^2 \quad \underline{\underline{R}} \end{aligned}$$

Var modellen bör alltid skrivas om till:

$$\frac{\text{Marketing}}{\text{POP}} = \frac{\beta_0}{\text{POP}} + \beta_1 \frac{\text{POP}}{\text{POP}} + \beta_2 \frac{\text{tax}}{\text{POP}} + \beta_3 \frac{\text{prop. young}}{\text{POP}} + \frac{u}{\text{POP}}$$

för att MLR.5 ska gälla och säkerställa att vår modell
innehåller den minsta variansen för samtliga linjära, " unbiased " estimatorer. R

$$4. \quad \text{corn} = \alpha_1 \text{ price} + \beta_1 \text{ income} + u_1$$

$$\text{corn} = \alpha_2 \text{ price} + \beta_2 \text{ rainfall} + \gamma_2 \text{ rainfall}^2 + u_2$$

Exogenous variables: income, rainfall, rainfall²

För att en ekvation ska kunna identifieras/estimeras måste dess endogena förklarande variabel, i det här fallet "price", kunna bytas ut mot ett instrument "besittande av en exogen variabel" i sin systemekvation som inte knörs med i den första ekvationen.

Den här exogena variabel måste vara korrelerad med "corn" och den endogena variabel men okorrelerad med feltermerna " u_i ".

För att identifiera den övre ekvationen behöver vi alltså en exogen variabel från den andra modellen, ex "rainfall", som instrument för "price" i den övre modellen.

Vi kan också använda "income" som instrument för "price" i den nedre modellen.

Vi använder oss av "Order condition" och "Rank condition":

Order condition: För att identifiera den första ekvationen behöver minst 1 exogen variabel exkluderas från den första ekvationen. *Om Villen då?*

Rank condition: Rank villkoret kräver att den exkluderade variabeln (här den första ekvationens återgång i den andra ekvationen) med en icke-negativ koefficient.

Jä, modellen är identifierad!

R