Faculty of Economics, University of Belgrade
IMQF studies
Year: 2021/2022
Professor: Zorica Mladenović and Aleksandra Nojković

Time allowed: 150 minutes
Total marks: 100

Exam in Intermediate Econometrics (Test 1, 50 marks)
Date: February/2022

1. Consider the simple linear regression model:

$$y_i=\beta_1 + \beta_2 x_{i2}+\varepsilon_i \quad , i = 1,2, \dots, n.$$

under classical linear regression model assumptions (Gauss-Markov assumptions).

a) Discuss whether the OLS estimator for the slope β_2 is unbiased (you are not asked to derive OLS estimator).

- Presentation 2/Slides 31 and 32

Under assumptions A(1) and A(2) OLS an estimator is unbiased: E(b)=β.

A(1): Error terms have mean zero $E(\varepsilon_i) = 0$.
A(2): All error terms are independent of all X variables:
$\{\varepsilon_1, \dots, \varepsilon_N\}$ is independent of $\{x_1, \dots, x_N\}$.

(2 marks)

b) Derive OLS estimator for a simple linear model by assuming that the intercept is zero.

- Similar (simpler model) to Presentation 2/Slide 14

$$SS = \sum_{i=1}^{N} (Y_i - b_2 X_i)^2$$

$$\frac{\partial SS}{\partial b_2} = \sum_{i=1}^{N} 2(Y_i - b_2 X_i)(-X_i) = 0 \Rightarrow b_2 = \frac{\sum_{i=1}^{N} X_i Y_i}{\sum_{i=1}^{N} X_i^2}.$$

(2 marks)

2. An economist is interested in estimating the production function defined by Cobb-Douglas specification:

$$Y = \beta_1 L^{\beta_L} K^{\beta_K} e^{\varepsilon}$$

where Y is production, L is labour, K is capital stock and ε is error term.

a) Provide the interpretation of the parameter β_L. Explain your answer.

- Similar to Presentation 2/Slides 72 and 73

$$\frac{\text{dlnY}}{\text{dlnX}} = \frac{\text{\%changeY}}{\text{\%changeX}} = \text{Elasticity of Y wrt change in X}$$

An elasticity measures the *relative* change in the dependent variable y_i due to a *relative* change in x_{ik}.

(2 marks)

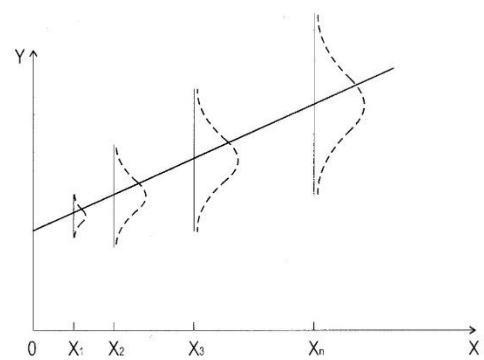
b) There is a suspect that the errors exhibit heteroscedasticity. Explain what should be considered by the concept of heteroscedasticity. Enhance your answer with the help of a graphical illustration.

- Presentation 3/Slides 7-9

Heteroskedasticity arises if different error terms do not have the same variance. When do we expect this?

- **Variances depend upon one or more explanatory variables (e.g., firm size);**
- Variances evolve over time (time-varying volatility);

Similar to the below illustrations:



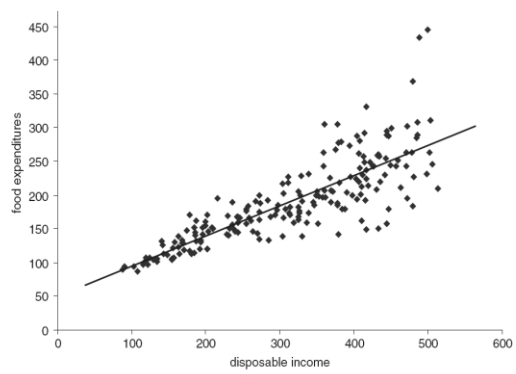


Figure 4.1 An Engel curve with heteroskedasticity

(2 marks)

3. The following output on log of wages (*lnwage*) was obtained using LFS data on 1462 women:

Dependent Variable: LNWAGE
Method: Least Squares
Date: 02/26/22 Time: 13:02
Sample (adjusted): 3 4375
Included observations: 1462 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.237523	0.098671	73.34979	0.0000
AGE	0.006016	0.002468	2.437506	0.0149
EXPERIENCE	0.004121	0.002443	1.686589	0.0919
EDUC	0.106554	0.005467	19.48904	0.0000
OWN3	0.223468	0.062584	3.570701	0.0004
SERVICES	0.210678	0.031693	6.647469	0.0000
AGRIC	-0.104414	0.061120	-1.708355	0.0878
OWN3*SERVICES	-0.148638	0.069677	-2.133234	0.0331

R-squared	0.284168	Mean dependent var	9.029893
Adjusted R-squared	0.280722	S.D. dependent var	0.531330
S.E. of regression	0.450623	Akaike info criterion	1.249084
Sum squared resid	295.2503	Schwarz criterion	1.278017
Log likelihood	-905.0803	Hannan-Quinn criter.	1.259876
F-statistic	82.45772	Durbin-Watson stat	1.355271
Prob(F-statistic)	0.000000		

where regressors in baseline model are:

experience – years of work experience;
age – age in years;
educ – number of years of full-time education;
own3 – 1 if works in private sector, 0 otherwise;
agric – 1 if works in agriculture, 0 otherwise;
services – 1 if works in services, 0 otherwise;
*own3*services* – interaction variable.

Note on all above requests: *Clearly explain the null and alternative hypothesis, the test statistics, and rejection rule.*

a) Test the null hypothesis that the coefficient of *educ* is zero (at 5% significance level).

- Presentation 2 / Slides 61-63

$H_0: b_{4(educ)}=0$
$H_1: b_{4(educ)}\neq 0$

Corresponding prob. of t-stat. (0.000) → We reject H_0.

(2 marks)

b) At the 5% significance level, test the joint significance of the regressors.

- Presentation 2 / Slides 48-47 and 64-66

$H_0: R^2=0$
$H_1: R^2 \neq 0$

Corresponding prob. of F-stat. (0.000) → We reject H_0.

(2 marks)

c) Is the return to age equal as to the return-to-work experience in our model (can we state that the coefficients of age and experience are equal)?

Wald Test:
Equation: EQ01

Test Statistic	Value	df	Probability
t-statistic	0.400139	1454	0.6891
F-statistic	0.160111	(1, 1454)	0.6891
Chi-square	0.160111	1	0.6891

Null Hypothesis: C(2)=C(3)
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(2) - C(3)	0.001895	0.004736

Restrictions are linear in coefficients.

- Presentation 2 / Slides 61-63

$H_0: b_{2(age)} = b_{3(experiance)}$
$H_1: b_{2(age)} \neq b_{3(experiance)}$

Corresponding prob. of t-stat. (0.6891) → We do not reject H_0.

(3 marks)

d) Use the result below to test the joint significance of the dummy variables excluded from baseline model:

Dependent Variable: LNWAGE
Method: Least Squares
Date: 02/26/22 Time: 13:29
Sample (adjusted): 3 4375
Included observations: 1462 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.352860	0.098281	74.81449	0.0000
AGE	0.005707	0.002521	2.264253	0.0237
EXPERIENCE	0.002597	0.002498	1.039854	0.2986
EDUC	0.114371	0.005480	20.87117	0.0000

R-squared	0.243266	Mean dependent var	9.029893
Adjusted R-squared	0.241709	S.D. dependent var	0.531330
S.E. of regression	0.462682	Akaike info criterion	1.299179
Sum squared resid	312.1208	Schwarz criterion	1.313645
Log likelihood	-945.6995	Hannan-Quinn criter.	1.304575
F-statistic	156.2337	Durbin-Watson stat	1.263129
Prob(F-statistic)	0.000000		

- Presentation 2 / Slide 81

$H_0: b_5 = b_6 = b_7 = b_8 = 0$
$H_1: H_0 \textit{ is not true.}$

$$F = \frac{(R_1^2 - R_0^2)/J}{(1 - R_1^2)/(N - K)}$$

$F = \frac{(0.284168-0.243266)/4}{(1-0.284168)/(1462-8)}$ =20.56626 > F*→ We reject H_0.

The set of dummy variables should not be excluded from baseline model.

(4 marks)

e) In order to test for general model misspecifications, we performed Ramsey's RESET test (we used two fitted terms). Is there a specification problem in our model?

Ramsey RESET Test
Equation: EQ01
Omitted Variables: Powers of fitted values from 2 to 3
Specification: LNWAGE C AGE EXPERIENCE EDUC OWN3
SERVICES AGRIC OWN3*SERVICES

	Value	df	Probability
F-statistic	9.843611	(2, 1452)	0.0001
Likelihood ratio	19.68963	2	0.0001

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	3.949656	2	1.974828
Restricted SSR	295.2503	1454	0.203061
Unrestricted SSR	291.3006	1452	0.200620

LR test summary:

	Value
Restricted LogL	-905.0803
Unrestricted LogL	-895.2354

- Presentation 2 / Slides 75-79

H_0: Regression specification is correct.
$H_1: H_0$ *is not true.*

Corresponding prob. of F-stat. (0.0001) → We reject H_0.

(3 marks)

f) Check for heteroskedasticity using result of the White test (no cross products/terms).

Heteroskedasticity Test: White
Null hypothesis: Homoskedasticity

F-statistic	5.098529	Prob. F(7,1454)	0.0000
Obs*R-squared	35.02632	Prob. Chi-Square(7)	0.0000
Scaled explained SS	94.99779	Prob. Chi-Square(7)	0.0000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 02/26/22 Time: 13:09
Sample: 3 4375
Included observations: 1462

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.360207	0.055371	6.505354	0.0000
AGE^2	-1.87E-05	2.87E-05	-0.649112	0.5164

EXPERIENCE^2	6.42E-05	6.71E-05	0.955705	0.3394
EDUC^2	-0.000163	0.000238	-0.685428	0.4932
OWN3^2	-0.152246	0.064998	-2.342301	0.0193
SERVICES^2	-0.174476	0.032976	-5.290958	0.0000
AGRIC^2	-0.016311	0.063524	-0.256773	0.7974
OWN3*SERVICES^2	0.201814	0.072499	2.783665	0.0054

R-squared	0.023958	Mean dependent var	0.201950
Adjusted R-squared	0.019259	S.D. dependent var	0.473096
S.E. of regression	0.468518	Akaike info criterion	1.326972
Sum squared resid	319.1662	Schwarz criterion	1.355906
Log likelihood	-962.0167	Hannan-Quinn criter.	1.337765
F-statistic	5.098529	Durbin-Watson stat	1.866338
Prob(F-statistic)	0.000010		

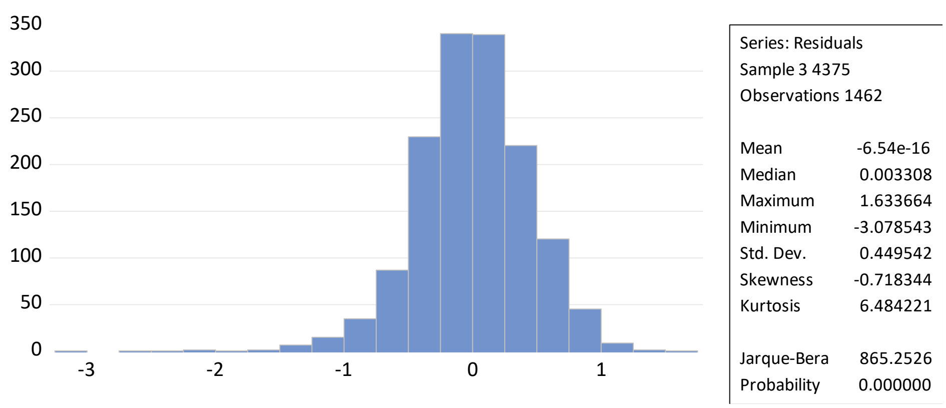
- Presentation 3 / Slides 27-29

H_0: Error terms are homoscedastic
H_1: Error terms are heteroscedastic

WH – statistics = 1462*0.023958=35.02632 > F* (corresponding prob. 0.000) → We reject H_0.

(3 marks)

g) Check for normality of residuals in the regression model.



- Presentation 1 / Slides 34 and 96

A convenient fifth assumption is that all error terms have a normal distribution. We specify:

(A5): $\varepsilon_i \sim NID(0, \sigma^2)$

Ho: Residuals have N distribution.
$H_1: H_0$ *is not true.*

JB = 865.25 (prob. 0.000) → We reject H_0.

(4 marks)

4. Derive the approximative relation between Durbin-Watson (DW) test-statistic and autocorrelation coefficient of order one (ρ).

- Presentation 3 / Slides 48

$$dw \approx 2 - 2\hat{\rho}$$

$$dw = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

$$dw = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} = \frac{\sum_{t=2}^{n} e_t^2 - 2\sum_{t=2}^{n} e_t e_{t-1} + \sum_{t=2}^{n} e_{t-1}^2}{\sum_{t=1}^{n} e_t^2},$$

$$\sum_{t=2}^{n} e_t^2 \approx \sum_{t=2}^{n} e_{t-1}^2 \rightarrow dw \approx 2\left(1 - \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}\right) \rightarrow dw \approx 2(1 - \hat{\rho})$$

$\hat{\rho}$ – *estimated autocorrelation coefficient of order one.*

(3 marks)

5. Consumers expenditure on food (*lcons*) are estimated based on quarterly data as follows:

Dependent Variable: LCONS
Method: Least Squares
Date: 02/26/22 Time: 13:35
Sample (adjusted): 1985Q2 1994Q2
Included observations: 37 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.488356	0.575327	-0.848831	0.4021
LCONS(-1)	0.818289	0.103707	7.890392	0.0000
LPRICE	-0.120416	0.086416	-1.393442	0.1728
LDISP	0.411340	0.169728	2.423524	0.0210

R-squared	0.758453	Mean dependent var	4.608665
Adjusted R-squared	0.736494	S.D. dependent var	0.051985
S.E. of regression	0.026685	Akaike info criterion	-4.307599
Sum squared resid	0.023500	Schwarz criterion	-4.133446
Log likelihood	83.69058	Hannan-Quinn criter.	-4.246202
F-statistic	34.53976	Durbin-Watson stat	1.727455
Prob(F-statistic)	0.000000		

where: *ldisp* = disposable income and *lprice* =the relative price index of food.

Then we estimated the auxiliary regression using residuals from the above regression as explanatory variable:

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 02/26/22 Time: 13:47
Sample: 1985Q2 1994Q2
Included observations: 37
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.153347	0.607265	0.252521	0.8023
LCONS(-1)	-0.054709	0.123515	-0.442932	0.6608
LPRICE	0.003521	0.086942	0.040502	0.9679
LDISP	0.018085	0.171957	0.105171	0.9169
RESID(-1)	0.174392	0.211345	0.825154	0.4154

R-squared	0.020834	Mean dependent var	8.51E-16
Adjusted R-squared	-0.101562	S.D. dependent var	0.025549
S.E. of regression	0.026815	Akaike info criterion	-4.274599
Sum squared resid	0.023010	Schwarz criterion	-4.056908
Log likelihood	84.08009	Hannan-Quinn criter.	-4.197853
F-statistic	0.170220	Durbin-Watson stat	1.855257
Prob(F-statistic)	0.952013		

Using both of the above results, test for evidence of first-order autocorrelation. Clearly explain the null and alternative hypothesis, the test statistics, rejection rule and assumptions of underlying test. What is the name of the test that that you find adequate in this case?

- Presentation 3 / Slides 46-47

H_o: No autocorrelation of order one.
$H_1: H_0$ *is not true*.

Breusch-Godfrey Serial Correlation LM Test (Durbin-Watson is not appropriate when lagged dependent variable is included in a model) :

(T-1)*R^2=37*0.0208=0.770865 < F*→ We do not reject H_0.

(4 marks)

6. A model of wages is specified for males as:

$$\ln(w_i) = \beta_1 + \beta_2 exp_i + \beta_3 expsq_i + \beta_4 educ_i + \beta_5 M_i + \beta_6 MR_i + \varepsilon_i + \varepsilon_i \quad (1)$$

where *w* = gross hourly wages, *exp* (*expsq*) = years of work experience (squared) and *educ* = numbers of year of full education; *M* =1 if married, 0 otherwise, and *R* = 1 if lives in a rural area, 0 otherwise.

a) You are worried about the effect of an omitted variable *ability* (unobserved heterogeneity). What effect might the omitted relevant variable *ability* have on the OLS estimate of the coefficient on *educ* in (1)?

- Presentation 4 / Slides 5-9 and 22-29

OLS estimator is biased (upward) and inconsistent.

(2 marks)

b) Why is there no dummy variable for male respondents that live in city areas?

- Presentation 2 / Slide 38

In order to avoid the "dummy variable trap"

(2 marks)

c) The data set includes a variable for the *IQ* (Intelligent Quotient) score, which can serve as a proxy for *ability*. How will including the *IQ* variable in (1), which is then estimated by OLS, change your coefficient estimates in (1)?

Presentation 4 / Slides 5-9 and 22-29

The OLS estimator on education will decrease.

(2 marks)

d) Instead of estimating the model outlined in (c) by OLS you have decided to estimate equation (1) using IV estimation and you believe you have two potential instruments for educ: (i) *IQ score*, and (ii) *education level of siblings*. Evaluate each of these instruments on the ground of instrument relevance and instruments exogeneity.

- Presentation 4 / Slides 16-21

The instrumental variables estimator is a consistent estimator for β_2 provided **the instruments are valid**

- This requires that they are both
  1) *Exogenous*, i.e., E{ $\varepsilon_t z_t$ } = 0

  and

  2) *Relevant*, i.e., cov{ x_t , z_t } $\neq$ 0.

(2 marks)

3) What, if anything, are key differences between OLS estimation as outlined in (c) and IV estimation as outlined in (d)?

- Presentation 4 / Slides 5-9 and 16-29

OLS estimator is inconsistent while IV is consistent.

(2 marks)

4) Using two instruments suggested in (d), carefully explain how you might go about undertaking a test of (i) instrument relevance, and (ii) instrument exogeneity.

(i)

- Presentation 4 / Slide 48

Briefly explain the idea of Stock-Watson (optionally Stock-Yogo test, too).

(ii)

- Presentation 4 / Slides 49-53

Briefly explain the idea of J-test.

(4 marks)