

Эконометрика — МИЭФ, 2024 final

МИЭФ

Эконометрика

2024

final

QUESTION 1

Multiple-choice test

In the model

$$Y_i = \beta_1 + u_i,$$

the OLS estimator of β_1 equals:

1. $b_1 = \bar{Y}$.
2. $b_1 = 0$.
3. $b_1 = \bar{u}$.
4. $b_1 = \frac{\sum_i Y_i^2}{\sum_i Y_i}$.
5. $b_1 = \frac{\sum_i Y_i}{n - 1}$.

QUESTION 2

Multiple-choice test

The regression model

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

had been transformed into one with deviations of the explanatory variables from their means:

$$X_2^* = X_2 - \bar{X}_2, \quad X_3^* = X_3 - \bar{X}_3.$$

In the transformed model, fitted for the same sample, generally speaking:

1. The estimates of the slope coefficients will change, while the estimate of the intercept will stay the same.
2. The estimates of both the slope coefficients and intercept will stay the same.
3. The estimates of the slope coefficients will stay the same, while the estimate of the intercept will change.
4. The estimates of both the slope coefficients and intercept will change.
5. The transformed model cannot be estimated due to multicollinearity.

QUESTION 3**Multiple-choice test**

Which method can be used for detecting heteroskedasticity of the type

$$\sigma_i^2 = \gamma X_i?$$

1. Goldfeld-Quandt test.
2. Durbin-Watson test.
3. Dickey-Fuller test.
4. Student's test.
5. None of the above.

QUESTION 4**Multiple-choice test**

The researcher wants to examine the life expectancy level in the world and estimates the following model:

$$\widehat{LE}_i = \hat{\beta}_1 + \hat{\beta}_2 EUR_i, \quad SSR = 316.$$

Here LE_i is life expectancy in the country in years, and EUR_i is a dummy variable equal to 1 if the country is European. It is known that there are 160 countries in the sample and 40 of them are European. What is the estimated covariance between $\hat{\beta}_1$ and $\hat{\beta}_2$?

1. -1.03%.
2. -1.67%.
3. -2.39%.
4. -2.89%.
5. -3.33%.

QUESTION 5

Multiple-choice test

The researcher estimated the model of the wage rate in thousand rubles gained by a specialist in the banking sector:

$$\ln w_i = 11.3 + 0.75EDUC_i + 1.2MALE_i + 0.22 \ln EXPER_i + u_i.$$

Here w is the wage rate received by an individual, $EDUC$ is the number of years of graduate study, $MALE$ is a binary variable equal to 1 if the person is male, and $EXPER$ is the number of years of working experience. Indicate the correct statement:

1. One more year of working experience increases the wage rate by 0.22 thousand rubles, on average, others equal.
2. One more year of graduate study increases the wage rate by 75%, on average, others equal.
3. One more year of working experience increases the wage rate by 22%, on average, others equal.
4. Males get 20% higher wages than females, on average, others equal.
5. Males get higher wages than females by 1.2 thousand rubles, on average, others equal.

QUESTION 6

Multiple-choice test

The Durbin-Wu-Hausman test can be used for detection of the following:

- I. Heteroscedasticity.
- II. Measurement errors.
- III. Simultaneous equations bias.
- IV. Endogeneity of explanatory variables.

1. I, II and III only.
2. III and IV only.
3. II and IV only.
4. II, III and IV only.
5. All I-IV.

QUESTION 7

Multiple-choice test

In econometrics, simultaneity bias arises when:

1. Strictly exogenous explanatory variables determine the dependent variable through a step-by-step process.
2. One or more of the explanatory variables is jointly determined with the dependent variable.
3. The disturbance term is correlated with the dependent variable.
4. Heteroscedasticity is present in the model.
5. There is correlation between some explanatory variables.

QUESTION 8

Multiple-choice test

To obtain consistent estimates of the parameters of the standard partial adjustment model

$$Y_t^* = \beta_1 + \beta_2 X_t + u_t, \quad Y_t - Y_{t-1} = \lambda(Y_t^* - Y_{t-1}),$$

one should use:

1. Koyck distribution.
2. Koyck transformation.
3. Error correction model.
4. ADL(1,0) model.
5. ADL(1,1) model.

QUESTION 9

Multiple-choice test

Which of the following statements is true?

1. The expectation of a random walk with drift process is a linear function of time.
2. The expectation of a random walk with drift process does not exist.
3. The expectation of a random walk with drift process is zero.
4. The expectation of a random walk with drift process is constant in time.
5. The expectation of a random walk with drift process is a quadratic function of time.

QUESTION 10**Multiple-choice test**

Given the estimated regression

$$z = 1.55 + 0.5X - 0.3X^2,$$

find the marginal partial effect at $X = 2$, applying the Probit model and its marginal probability function

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

You may use $f(z = 1.35) = 0.1604$.

1. 0.1502.
2. 0.1604.
3. -0.1123.
4. -0.2016.
5. None of the above.

QUESTION 11**Multiple-choice test**

In a Random Effects (RE) model, the error term u_{it} can be decomposed into:

1. A time-specific component and an individual-specific component.
2. An observation-specific component and a time-specific component.
3. A time-specific component and a group-specific component.
4. An individual-specific component and an observation-specific component.
5. An individual-specific component and a group-specific component.

QUESTION 12

Multiple-choice test

In the context of panel data, the term "unbalanced panel" refers to:

1. A panel with more cross-sectional units than time periods.
2. A panel with heteroskedastic errors.
3. A panel where all cross-sectional units are observed for the same time periods.
4. A panel with correlated random effects.
5. A panel where the number of time periods differs across cross-sectional units.

QUESTION 13**Written part, Section A — original Question 1 — 25 marks**

Part 2. Written examination. One session, 2 hours without break.

SECTION A. Answer all questions 1-2 from this section.

A student has aggregate annual data on consumer expenditure on cloth $CLOT$, disposable personal income DPI , both measured at current prices in U.S. dollars billions, a price index for clothes $PCLOT$, and a price index for total personal expenditure $PTPE$, for the United States for the period 1998-2022, 25 observations. The student runs the following regressions. Standard errors are in parentheses; SSR is the sum of squared residuals. When considering equations, issues related to nonstationary time series should be ignored.

$$\widehat{\log CLOT} = -1.32 + 0.80 \log DPI - 0.61 \log PCLOT + 0.60 \log PTPE, \quad R^2 = 0.998. \quad (1)$$

Standard errors and additional statistics:

$$(0.12) \quad (0.06) \quad (0.11) \quad (0.05), \quad SSR = 0.0045, \quad DW = 1.66.$$

$$\widehat{\log \left(\frac{CLOT}{PTPE} \right)} = -1.80 + 0.58 \log \left(\frac{DPI}{PTPE} \right) + 0.13 \log \left(\frac{PCLOT}{PTPE} \right), \quad R^2 = 0.937. \quad (2)$$

Standard errors and additional statistics:

$$(0.19) \quad (0.09) \quad (0.10), \quad SSR = 0.0172, \quad DW = 0.66.$$

(a) (12 marks)

- Give an interpretation to regression coefficients of equations (1) and (2), commenting on the meaning of the variables used in these models and values of coefficients.
- Are coefficients of equations (1) and (2) significant? What conclusions can be drawn from this?
- Are coefficients of equations (1) and (2) significant? What conclusions can be drawn from this? Are equations (1) and (2) significant in general?

(b) (13 marks)

- Explain why the second regression may be considered as a restricted version of the first. What is the restriction? Test the restriction. What information — on R^2 or SSR or both — can be used for this test?
- What are the benefits and risks of imposing a restriction in a regression model? Consider the restricted and unrestricted equation in turn as correct specification. Which equation should be

chosen for further study of the problem under consideration based on the analysis performed, and why?

- The teacher advised the student to also pay attention to the Durbin-Watson statistics in equations (1) and (2). Perform the required test and draw conclusions.

QUESTION 14**Written part, Section A — original Question 2 — 25 marks**

Working on her coursework, a student of ICEF collected data on a sample of 36 ICEF graduates from different years of graduation working in Russia. She is interested in studying their current salary, SAL , in thousands of roubles per month. Explanatory variables are AGE , age of respondent in years, age squared AGE^2 , and also some dummy variables: $MSCA$, Master of Science Degree Abroad, equal to 1 for those graduates who have received a master's degree abroad and 0 otherwise; NFE , no further education, equal to 1 for those graduates who received a master's degree neither abroad nor in the country; and $MALE$, equal to 1 for male and 0 for female. Here are the results of estimation of two regressions using different sets of variables. Standard errors are in brackets.

$$\widehat{SAL}_i = 15.18 + 0.254AGE_i - 0.003AGE_i^2, \quad R^2 = 0.48. \quad (1)$$

Standard errors:

$$(0.18) \quad (0.021) \quad (0.0015)$$

$$\widehat{SAL}_i = 10.068 + 0.18AGE_i - 0.002AGE_i^2 + 4.021MSCA_i - 5.123NFE_i + 0.025MALE_i, \quad R^2 = 0.63$$

Standard errors:

$$(0.049) \quad (0.03) \quad (0.001) \quad (0.456) \quad (2.237) \quad (0.016)$$

(a) (12 marks)

- Give interpretation to the coefficients of equation (1). Evaluate the marginal effect of AGE for $AGE = 25$, $AGE = 42$, and $AGE = 60$.
- Is this marginal effect of AGE significant?
- What is the role of the AGE^2 variable in equations (1)-(2)? Noticing that this variable was insignificant in both equations, the student decided to exclude it from the equations. Try to predict how this might affect the coefficients of model (1). Explain your reasoning.

(b) (13 marks)

- How many categories of education level of ICEF graduates do dummy variables $MSCA$ and NFE describe? What is the reference category for equation (2)? What do the coefficients of dummy variables tell about salaries of ICEF graduates?
- Are the coefficients of the variables $MSCA$, NFE , and $MALE$ significant? Are they jointly significant?

- The student decided to replace the variable NFE , no further education, with $MSCR$, getting a master's degree in Russia, leaving all other variables in the equation. Indicate what changes will occur in equation (2), and how exactly they will change, and which coefficients will remain the same. Answer the same questions about the standard errors of the coefficients and the value of R-squared.

To answer these questions, simply write a new equation, writing out only those quantities whose values you can accurately predict and leaving blanks for those quantities whose value cannot be predicted.

No explanation is required.

$$\widehat{SAL}_i = \text{---} + \text{---}AGE_i + \text{---}AGE_i^2 + \text{---}MSCA_i + \text{---}MSCR_i + \text{---}MALE_i, \quad R^2 = \text{---} \quad (3)$$

Standard-error template:

(---) (---) (---) (---) (---) (---)

QUESTION 15

Written part, Section B — original Question 3 — 25 marks

SECTION B. Answer only ONE question from this section: Question 3 OR Question 4.

A variable Y is determined by a variable X , the relationship being

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

where u_i is a disturbance term that satisfies regression Model B assumptions. The values of X are drawn randomly from a population with variance σ_X^2 .

(a) (12 marks)

- Assuming X_i and u_i are correlated, briefly show that the OLS estimator of β_2 is inconsistent.
- Suppose that there exists a third variable Z that is correlated with X but independent of u . Demonstrate that if the researcher had regressed Y on X using Z as an instrument for X , the slope coefficient $\hat{\beta}_2^{IV}$ would have been a consistent estimator of β_2 .
- What is the difference between an instrumental variable, instrument, and a proxy variable?
- What are the rules for choosing valid instruments?

(b) (13 marks)

- During discussion of these results, some of the participants suggested using the TSLS approach instead, saying that it has many advantages compared to the instrumental variable approach. What is the TSLS estimator here?
- Show that in this case the TSLS estimator is the same as the instrumental variable estimator.

QUESTION 16**Written part, Section B — original Question 4 — 25 marks**

A student is interested in the influence of the price level and income on tourists' expenditures on various types of entertainment in a large seaside resort, E_{it} , in thousands of dollars. Suppose that the time series of expenditures on D_t , diving; S_t , surfing; P_t , paragliding; Y_t , yachting; F_t , fishing; and T_t , spa treatment, for ten years, 2008-2017, form the panel as observations for particular units of some general type of entertainment, E_{it} . DPI_t is disposable personal income, also in thousands of dollars, and PRE_{it} is the relative price index for the corresponding E_{it} . Let the model under investigation be

$$LGE_{it} = \beta_1 + \beta_2 LGDPI_t + \beta_3 LGPRE_{it} + u_{it},$$

where

$$LGE = \log(E), \quad LGDPI = \log(DPI), \quad LGPRE = \log(PRE).$$

She estimated three equations:

$$\widehat{LGE}_{ij} = -4.66 + 0.93LGDPI_j - 0.11LPRE_{ij}, \quad R^2 = 0.048, \quad RSS = 15.18 \quad 1 - \text{Pooled. (1)}$$

Standard errors:

$$(5.84) \quad (0.66) \quad (0.06)$$

$$\widehat{LGE}_{ij} = -2.13 + 0.98LGDPI_j - 0.68LPRE_{ij}, \quad R^2 = 0.994, \quad RSS = 0.0957 \quad 2 - \text{Fixed. (2)}$$

Standard errors:

$$(0.51) \quad (0.06) \quad (0.05)$$

$$\widehat{LGE}_{ij} = -2.15 + 0.98LGDPI_j - 0.67LPRE_{ij}, \quad R^2 = 0.899, \quad RSS = 0.106 \quad 3 - \text{Random. (3)}$$

Standard errors:

$$(0.57) \quad (0.06) \quad (0.05)$$

(a) (12 marks)

- What are the advantages of panel data analysis comparing to cross-section regression and time series analysis? What relative advantages of panel data models can be indicated based on the comparison of the estimated equations (1)-(3)?
- The student decided to use the approach based on the fixed effects panel data model. Help the student to understand what the LSDV method is and how to use it for estimation of the panel data model under consideration.

- The fixed effect for yachting is 0.53, and for fishing -0.85 ; explain what this means for analysing equation (2).

(b) (13 marks)

- How to test the presence of unobserved heterogeneity in the fixed effect model, or, to put it another way, that there are considerable differences between the pooled model and the model with fixed effects? Give some details: indicate the corresponding test and the data needed for it, null hypothesis, distribution of the test statistic, the number of degrees of freedom, and the decision rule.
- What test can you use to choose between random effects and fixed effects panel model? The statistic of this test turned out to be 1.76; what conclusion can be drawn from this? What risks are associated with choosing models with fixed and random effects? What meaningful considerations can be used here when choosing the type of model? What is your final choice?