Эконометрика — МИЭФ, 2025 final

МИЭФ | Эконометрика | 2025 | final

QUESTION 1

Multiple-choice test

In the regression model

$$Y_i = \beta_0 + \beta_1 \log(X_i) + u_i,$$

the interpretation of β_1 is:

1. A 1-unit increase in X leads to a β_1-unit change in Y.
2. A 1% increase in X leads to a $\beta_1/100$-unit change in Y.
3. A 1-unit increase in $\log(X)$ leads to a $\beta_1\%$ change in Y.
4. The elasticity of Y with respect to X.
5. The marginal effect of X on Y.

QUESTION 2

Multiple-choice test

To test $H_0 : 5\beta_1 = \beta_2$ in a regression, the appropriate test statistic is:

1. F-test comparing restricted and unrestricted models.
2. t-test:

$$t = \frac{5\hat{\beta}_1 - \hat{\beta}_2}{\text{se}(5\hat{\beta}_1 - \hat{\beta}_2)}.$$

3. χ^2-test using the likelihood ratio.
4. Breusch-Pagan test for heteroskedasticity.
5. Durbin-Watson test for autocorrelation.

QUESTION 3

Multiple-choice test

If heteroskedasticity is present in a regression, the OLS estimators remain:

1. Biased and inconsistent.
2. Unbiased and efficient.
3. Unbiased but standard errors are invalid.
4. Consistent only if robust standard errors are used.
5. Efficient under the Gauss-Markov assumptions.

QUESTION 4

Multiple-choice test

Cointegration between two $I(1)$ variables implies:

1. Both are stationary.
2. Some linear combination is $I(0)$.
3. Any linear combination is $I(0)$.
4. Any linear combination is $I(1)$.
5. Both have unit roots.

QUESTION 5

Multiple-choice test

In the demand-supply system:

$$\begin{aligned} beer &= \alpha_1 price + \alpha_2 GDP + u_1 \quad && \text{(Demand)}, \\ beer &= \beta_1 price + \beta_2 sunshine + u_2 \quad && \text{(Supply)}, \end{aligned}$$

$price$ is endogenous because:

1. It is correlated with u_1.
2. It is measured with error.
3. Sunshine is omitted.
4. GDP is exogenous.
5. The model is overidentified.

QUESTION 6

Multiple-choice test

A valid instrumental variable Z must satisfy:

1. $\text{Cov}(Z, u) = 0$ and $\text{Cov}(Z, X) \neq 0$.
2. $\text{Cov}(Z, u) \neq 0$ and $\text{Cov}(Z, X) = 0$.
3. $\text{Cov}(Z, X) = 0$ and $\text{Cov}(Z, Y) \neq 0$.
4. $\text{Cov}(Z, u) \neq 0$ and $\text{Cov}(Z, X) \neq 0$.
5. Z must be binary.

QUESTION 7

Multiple-choice test

The marginal effect of a regressor X in a Probit model is:

1. The coefficient β.
2. $\beta \times \phi(\beta X)$, where ϕ is the standard normal PDF.
3. The value of e^{β}.
4. The change in Y for a 1-unit increase in X.
5. The R^2 of the regression.

QUESTION 8

Multiple-choice test

Indicate the incorrect statement among the following ones:

1. If X_t is a random walk with drift, the series of first differences $\Delta X_t = (X_t - X_{t-1}) = \beta_1 + \varepsilon_t$, where ε_t is white noise, is stationary.

2. The time trend $X_t = \beta_1 + \beta_2 t + \varepsilon_t$ is a non-stationary series.

3. The MA(1) process $X_t = \varepsilon_t + \alpha_2 \varepsilon_{t-1}$ is stationary.

4. The AR(1) process $X_t = \beta_2 X_{t-1} + \varepsilon_t$, with $-1 < \beta_2 < 1$, is asymptotically stationary.

5. The stationarity of an ARMA process is determined by its MA part.

QUESTION 9

Multiple-choice test

For the model

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

where X_i are non-stochastic and the Model A assumptions are satisfied, the following three estimators of β_2 are proposed:

$$b_1 = \frac{\bar{Y}}{\bar{X}}, \qquad b_2 = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_i (X_i - \bar{X})^2}, \qquad b_3 = \frac{\sum_i X_i Y_i}{\sum_i X_i^2}.$$

The following is correct for these estimators:

1. All the estimators b_1, b_2 and b_3 are unbiased.
2. All the estimators b_1, b_2 and b_3 are biased.
3. The estimator b_2 is unbiased, while b_1 and b_3 are biased.
4. The estimators b_1 and b_2 are unbiased, while b_3 is biased.
5. The estimators b_2 and b_3 are unbiased, while b_1 is biased.

QUESTION 10

Multiple-choice test

A student estimated by OLS the production function

$$y = \gamma_1 + \alpha k + \beta l + u, \tag{1}$$

where y is the output growth rate, k is the capital growth rate, and l is the labour growth rate. Then he decided to estimate by OLS the function

$$y - k - l = \gamma_2 + \mu k + \rho l + u. \tag{2}$$

Which statement of the following ones is correct?

1. $\hat{\mu} = \hat{\alpha}$.
2. $\hat{\rho} = \hat{\beta}$.
3. $R_1^2 = R_2^2$.
4. $SSR_1 = SSR_2$.
5. $SST_1 = SST_2$.

QUESTION 11

Multiple-choice test

What is an assumption underlying the random effects model that must be satisfied for its estimates to be consistent?

1. The error term must be homoscedastic across entities.
2. The unobserved individual effect must be uncorrelated with the independent variables.
3. All variables must be normally distributed.
4. There must be no time trends affecting the dependent variable.
5. The number of entities must equal the number of time periods.

QUESTION 12

Multiple-choice test

In a panel data context, which statement about time dummies is accurate?

1. They are unnecessary if using a fixed effects model.
2. They help control for time-specific shocks that affect all entities equally.
3. They can only be included in random effects models.
4. They complicate the estimation process without providing additional information.
5. They should always be included regardless of the chosen model.

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.

When an ICEF student went to China on an exchange programme, he discovered that Chinese people drink coffee as well as the traditional tea. The famous Starbucks (SB) is catching up with the young company Luckin Coffee (LC) with its own coffee service policy. The student obtained data from his young Chinese friends on their expenses for coffee purchased in SB coffee shops, their income, and also service prices at SB and prices of similar services at LC. To study the competition between these companies, he estimated a model in which the logarithm of SB coffee expenses $cofSB_i$ is regressed on the log of the price of coffee purchased at Starbucks pSB_i and the log of the price of coffee purchased at Luckin Coffee pLC_i, with the following results:

$$\widehat{cofSB_i} = -1.75 - 1.34pSB_i - 0.27pLC_i, \qquad R^2 = 0.78. \tag{1}$$

Standard errors:

$$(3.27) \qquad (0.32) \qquad (0.18)$$

Standard errors are in brackets, the sample size is 32, and all variables are in logarithms.

(a) (13 marks)

- Based on the interpretation of the coefficients of model (1), explain whether they are consistent with expectations according to economic theory. Are these coefficients significant? Should the insignificant variable be excluded from the model? Why or why not?

Following his professor's advice, the student added to the model log of income inc_i:

$$\widehat{cofSB_i} = -6.27 + 1.05inc_i - 0.924pSB_i + 0.84pLC_i. \tag{2}$$

Standard errors:

$$(1.48) \qquad (0.15) \qquad (0.19) \qquad (0.39)$$

- Unfortunately, the student forgot to include the R-squared value in model (2). Help him to find it without recalculating the regression, relying only on the relationship

between the corresponding t and F statistics.

- What do you think are the advantages of model (2) over model (1)?

(b) (12 marks)

- Let β_{pSB} and β_{pLC} be the coefficients of the variables pSB and pLC in (2). What would a student need to do to obtain the information necessary to test the hypothesis

$$\beta_{pLC} = -\beta_{pSB} \quad \text{vs.} \quad \beta_{pLC} \neq -\beta_{pSB}?$$

Explain in detail.

- What is the economic meaning of the hypotheses to be tested? How to test the hypothesis

$$\beta_{pLC} = -\beta_{pSB} \quad \text{against} \quad \beta_{pLC} < -\beta_{pSB}?$$

What information is required and how to get it?

QUESTION 14

Written part, Section A — original Question 2 — 25 marks

Consider two types of time series: random walk

$$X_t = X_{t-1} + \varepsilon_t$$

and trend

$$Y_t = t + \varepsilon_t.$$

We assume that a random term ε_t is white noise, meaning that

$$E(\varepsilon_t) = 0, \qquad E(\varepsilon_t^2) = \sigma_\varepsilon^2, \qquad E(\varepsilon_t \varepsilon_s) = 0 \text{ if } s \neq t,$$

and is independent of the lagged values of the time series under consideration. It is also assumed that the initial state of any series under consideration, if it exists, is fixed.

(a) (13 marks)

- What is a stationary time series? What is the order of integration of a time series? What series are called integrated $I(0)$? What series are called integrated $I(1)$, and so on? What is the integration order of the white noise ε_t above?

- What is the order of integration of the random walk $X_t = X_{t-1} + \varepsilon_t$? Explain.

- What is the order of integration of the trend series $Y_t = t + \varepsilon_t$? Explain.

(b) (12 marks)

- What does it mean to say that two time series are cointegrated? How can cointegration of time series be used to improve the quality of econometric models?

- How do we check in practice that two time series are cointegrated?

- Consider two time series from (a): a random walk $X_t = X_{t-1} + \varepsilon_t$ and a trend $Y_t = t + \xi_t$, where ε_t and ξ_t are independent white noises. Are these time series cointegrated?

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 student studying the impact of a Master's degree on earnings collected data on the monthly income, E, in thousands of rubles and work experience W in years, for 168 ICEF undergraduates without further education and 42 ICEF Masters graduates. Defining a dummy variable M to be equal to 1 in the case of a Master degree and 0 in the case of a bachelor degree only, and a slope dummy MW as the product of M and W, the researcher runs the following regressions. Standard errors are in parentheses; SSR is the sum of squares of residuals.

Undergraduates:

$$\widehat{E}_i = 44.3 + 1.23W_i, \qquad SSR = 14710.3. \tag{1}$$

Graduates:

$$\widehat{E}_i = 56.2 + 0.74W_i, \qquad SSR = 5028.2. \tag{2}$$

Combined:

$$\widehat{E}_i = 46.5 + 1.18W_i, \qquad SSR = 22774.8. \tag{3}$$

Combined:

$$\widehat{E}_i = 44.6 + 1.16W_i + 9.47M_i, \qquad SSR = 19765.1. \tag{4}$$

Combined:

$$\widehat{E}_i = 44.3 + 1.23W_i + 11.9M_i - 0.49MW_i, \qquad SSR = 19738.5. \tag{5}$$

(a) (10 marks)

- Explain the meaning of all the coefficients in equations (1), (2), and (3).
- The student noticed that some coefficients of equations (1)-(5) coincide and others are related by simple arithmetic relationships. Find and explain such facts.

- Explain why the intercepts and slope coefficients of variable W in equations (4) and (5) are different. What advantage does equation (5) have over equation (4) for analysing graduate earnings?

- What is the interpretation of the coefficient of variable M in equation (4)? Is this coefficient significant?

(b) (10 marks)

- Test whether the earnings function for Master's graduates differs from that for Bachelor's graduates using an appropriate test for dummy variables based on equation (5).

- Now answer the same question using a different approach based on the Chow test. Explain the logic of the test: the null hypothesis, test statistic, and decision rule. How does this test relate to the previous one?

(c) (5 marks)

- A student also wants to investigate whether gender, $G = 1$ for males and $G = 0$ otherwise, affects earnings. What regression equation would you advise him to estimate in order to make this estimate as accurate as possible? Give reasons for your suggestion.

QUESTION 16

Written part, Section B — original Question 4 — 25 marks

The student is interested in the influence of the price level and income on the consumption of mineral water W_{it} imported into Russia. Suppose that the time series B_t (Borjomi), E_t (Evian), N_t (Narzan), P_t (Perrier), S_t (San Pellegrino), and V_t (Vittel) for 12 years, 2013-2024, form the panel, being the observations of the expenditures on a particular type i of water as a representative of the variable water W_{it}. INC_t is income, both in thousand rubles, and PRW_{it} is the relative price index for the corresponding type of water.

Estimated equations are:

$$\widehat{\log W}_{it} = -5.77 + 0.76 \log DPI_t - 0.08 \log PRW_{it}, \qquad R^2 = 0.054 \qquad 1 \text{ - Pooled.}$$

$$\widehat{\log W}_{it} = -3.21 + 0.95 \log DPI_t - 0.78 \log PRW_{it}, \qquad R^2 = 0.921 \qquad 2 \text{ - Fixed (LSDV).}$$

$$\widehat{\log W}_{it} = -3.22 + 0.95 \log DPI_t - 0.77 \log PRW_{it}, \qquad R^2 = 0.916 \qquad 3 \text{ - Random.}$$

(a) (10 marks)

- What are the advantages of panel data analysis comparing to cross-section regression and time series analysis? Explain the main elements of a typical panel data model. What is unobserved heterogeneity and how is it treated in pooled, fixed, and random effects models?

- Explain the idea of the LSDV model. How does this model reflect unobserved heterogeneity? What are fixed effects?

(b) (8 marks)

- What test can justify the advantage of the LSDV model over pooled regression? Run this test and draw conclusions. What other arguments are there for LSDV regression?

(c) (7 marks)

- What are the assumptions on which the use of the random effects method is based? What are the consequences of violating these assumptions?

- What is the Durbin-Wu-Hausman (DWH) test to choose between random effects and fixed effects panel model? What is its main idea, what pair of hypotheses are being used, and what is the distribution of the test statistic? The statistic of the DWH test was 2.72; what conclusion can be drawn from this? What would be your final conclusion?