Эконометрика — МИЭФ, 2025 midterm 1

МИЭФ Эконометрика 2025 midterm 1

QUESTION 1

Multiple-choice test

Which of the following is a difference between least absolute deviations, LAD, and ordinary least squares, OLS?

1. OLS and LAD are equally sensitive to outlying observations.
2. OLS is more computationally intensive than LAD.
3. OLS is more sensitive to outlying observations than LAD.
4. OLS is justified for very large samples, while LAD is justified for smaller samples.
5. OLS estimates the conditional median of the dependent variable, while LAD estimates the conditional mean.

QUESTION 2

Multiple-choice test

Which of the following regression models is nonlinear in parameters?

1. $y = \frac{1}{\beta_0 + \beta_1 x} + u.$
2. $y = \beta_0 + \beta_1 x^{1/2} + u.$
3. $\log y = \beta_0 + \beta_1 \log x + u.$
4. $\log y = \beta_0 + \beta_1 x + u.$
5. All regressions 1)-4) are nonlinear in parameters.

QUESTION 3

Multiple-choice test

A linear transformation of the explanatory variable X in the model

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

does not generally change:

1. The estimate of the intercept.
2. The estimate of the slope coefficient.
3. The standard error of the slope coefficient estimate.
4. The determination coefficient of the regression.
5. The standard error of the intercept estimate.

QUESTION 4

Multiple-choice test

A student investigates factors affecting GDP growth for 200 countries in 2023. GDP is first measured in constant 2021 US dollars at purchasing-power parity. He then measures GDP growth using national currencies at constant prices. All regressors remain the same, and the GDP growth rate remains the dependent variable. What can be said about the two regressions?

1. All coefficients are the same.
2. Only the intercept changes; all slope coefficients remain the same.
3. All coefficients are generally different.
4. Coefficients may differ, but R^2 is the same.
5. Coefficients may differ, but SSR is the same.

QUESTION 5

Multiple-choice test

A student estimates 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. She then estimates

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

Which statement is correct?

1. Model (1) is a restricted version of (2).
2. Model (2) is a restricted version of (1).
3. Both statements 1) and 2) are incorrect.
4. Models (1) and (2) are equivalent.
5. There is perfect multicollinearity in model (2).

QUESTION 6

Multiple-choice test

Suppose the following model is estimated by OLS and the Gauss-Markov conditions are satisfied:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (1 + \beta_3 - \beta_2) x_3 + u.$$

Then:

1. You can obtain an unbiased estimate of β_3.

2. You cannot obtain an unbiased estimate of β_3, but can obtain a consistent estimate.

3. You cannot obtain either an unbiased or a biased-but-consistent estimate of β_3.

4. You cannot obtain any estimate of β_3 because of perfect multicollinearity.

5. All the above statements are incorrect.

QUESTION 7

Multiple-choice test

In the regression model

$$y = \alpha + \beta x + u,$$

where u satisfies the Gauss-Markov conditions and is normally distributed, x contains random measurement errors that are independent, normally distributed, homoscedastic, not autocorrelated, and have zero expected values. Suppose $\beta > 0$ and the mean of x is negative. For large samples:

1. The estimate of α is biased upwards.
2. The estimate of α is biased downwards.
3. The estimate of α is unbiased.
4. The estimate of α may be biased upwards or downwards.
5. The estimate of β is biased upwards.

QUESTION 8

Multiple-choice test

If OLS is used in a simple regression model with heteroscedasticity, the population variance of the slope estimator is

$$\operatorname{var}(b_2) = \frac{\sum_{i=1}^{n} x_i^2 \sigma_i^2}{\left(\sum_{i=1}^{n} x_i^2\right)^2}. \tag{1}$$

Under homoscedasticity,

$$\operatorname{var}(b_2) = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}. \tag{2}$$

Let $\sigma_i^2 = \sigma^2 k_i$, where k_i are unknown nonnegative weights and $\sum k_i = n$. Then:

1. Expression (1) is always greater than (2).
2. Expression (1) is always less than (2).
3. Expression (1) is greater than or equal to (2).
4. Expression (1) is less than or equal to (2).
5. Expression (1) can be greater than, less than, or equal to (2), depending on the relationship between σ_i and x_i.

QUESTION 9

Multiple-choice test

Which of the following is a difference between the White test and the Breusch-Pagan test?

1. The White test detects heteroscedasticity of unknown form, while the Breusch-Pagan test detects heteroscedasticity of a specified form.
2. The Breusch-Pagan test detects heteroscedasticity of unknown form, while the White test detects heteroscedasticity of a specified form.
3. The number of regressors used in the White test is larger than the number used in the Breusch-Pagan test.
4. The number of regressors used in the Breusch-Pagan test is larger than the number used in the White test.
5. None of the above.

QUESTION 10

Multiple-choice test

The Durbin-Wu-Hausman test can be used to detect:

I. Measurement-error bias.

II. Simultaneous-equations bias.

III. Endogeneity of explanatory variables.

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

QUESTION 11

Multiple-choice test

The economic model is

$$\begin{aligned} y_1&=\alpha+\tau y_2+\pi x_1+\phi x_2+\varepsilon x_3+u_1, \tag{1}\\ y_2&=\beta+\mu y_1+\gamma x_1+u_2. \tag{2} \end{aligned}$$

Here y_1 and y_2 are endogenous variables; x_1, x_2, and x_3 are exogenous variables; and u_1, u_2 satisfy the Gauss-Markov conditions. Indicate the correct statement:

1. You may apply TSLS to (1), but not to (2).
2. You may apply TSLS to (2), but not to (1).
3. You may apply TSLS to both (1) and (2).
4. You may not apply TSLS to either (1) or (2).
5. TSLS is not needed because OLS provides consistent estimates in both equations.

QUESTION 12

Multiple-choice test

For a simultaneous equations model with 9 equations, 9 endogenous variables, and 8 exogenous variables, which statement is true for any equation?

1. An equation is likely to be underidentified if 10 variables are missing from it.
2. An equation is likely to be exactly identified if 9 variables are missing from it.
3. An equation is likely to be overidentified if 9 variables are missing from it.
4. An equation is likely to be exactly identified if 10 variables are missing from it.
5. An equation is likely to be underidentified if 11 variables are missing from it.

Part 2. Free Response Questions — 2 hours.

Section A. Answer all questions from this section (original Questions 1-2).

QUESTION 13

Written Question 1 — 25 marks

An ICEF student preparing her diploma examines the dependence of painting prices P_i, in thousands of dollars, on the painting's age AGE_i, measured in decades from the current year to the year in which the work was created, and canvas size S_i, measured in square feet. Data on 19 paintings sold at one auction are used to estimate:

$$\widehat{P_i} = 2.35 + 0.028AGE_i + 0.037S_i, \qquad R^2 = 0.325, \qquad SSR = 0.417,$$

with standard errors

$$(0.57) \qquad (0.011) \qquad (0.020). \qquad (1)$$

(a) (13 marks)

- Interpret the coefficients of equation (1).
- Evaluate the significance of the individual coefficients and of equation (1) as a whole.
- S_i is insignificant, and the student considers excluding it. What are the likely consequences, given that the correlation between AGE and S is -0.93?

The supervisor recommends adding provenance information. Let $PV_i = 1$ if the painting has documented provenance and $PV_i = 0$ otherwise. The student estimates

$$\widehat{P_i} = 0.87 + 0.062AGE_i + 0.082S_i + 0.36PV_i, \qquad R^2 = 0.554, \qquad SSR = 0.276,$$

with standard errors

$$(0.71) \qquad (0.016) \qquad (0.024) \qquad (0.13). \qquad (2)$$

- Why are the intercepts different in equations (1) and (2)?
- What does the coefficient on PV_i mean? Is it significant?
- What assumptions about the coefficients on AGE_i and S_i are implicit in equation (2)?

(b) (12 marks) The supervisor also recommends adding PV_iAGE_i and PV_iS_i:

$$\widehat{P_i} = -0.097 + 0.085AGE_i + 0.11S_i + 2.69PV_i - 0.07PV_iAGE_i - 0.08PV_iS_i,$$

$$R^2 = 0.691, \qquad SSR = 0.191,$$

with standard errors

$$(0.75) \qquad (0.017) \qquad (0.025) \qquad (3.55) \qquad (0.05) \qquad (0.14). \tag{3}$$

- What is the meaning of this recommendation? How do the marginal effects differ for paintings with and without provenance?
- Is provenance statistically significant in equation (3)?
- The alternative approach is a Chow test. Suppose equation (1), estimated separately for paintings without provenance, gives $SSR = 0.1908$, while for paintings with provenance it gives $SSR = 0.000122$. Explain the Chow test and compare its result with the previous test.

QUESTION 14

Written Question 2 — 25 marks

A student studies selling prices of 400 houses in the Moscow region, measured in thousands of roubles. The data include property size in square metres, number of bedrooms, and a dummy for air conditioning.

The estimated equation is

$$\log(price_i) = 9.894 + 0.300\log(size_i) + 0.078 bedrooms_i + 0.212 airco_i, \qquad n = 400. \quad (1)$$

Conventional standard errors are

$$(0.232) \qquad (0.028) \qquad (0.015) \qquad (0.024),$$

and heteroscedasticity-robust standard errors are

$$[0.233] \qquad [0.028] \qquad [0.018] \qquad [0.023].$$

(a) (13 marks)

- What is the economic meaning of the coefficient on $\log(size)$?
- Interpret the coefficient on $bedrooms$.
- Interpret the coefficient on $airco$.
- What is heteroscedasticity? Why is it likely in this sample? What are its consequences for econometric estimation?
- What are heteroscedasticity-consistent standard errors and what are they used for? Comment on the differences between the two sets of standard errors. Are the coefficient estimates significant?

(b) (12 marks) The student observes that the variance of $\log(price)$ increases with $\log(size)$ and assumes

$$\sigma_{u_i} = k\log(size_i),$$

in addition to the Gauss-Markov assumptions and normality.

- To test for heteroscedasticity, she orders the 400 observations by size. Estimating equation (1) on the 150 observations with the smallest $\log(size)$ gives $SSR_1 = 6.214$.

Estimating it on the 100 observations with the largest $\log(size)$ gives $SSR_2 = 7.781$. The unequal sample sizes are used because there are fewer high-priced properties. Carry out the appropriate test, state the hypotheses, calculate the statistic, and draw a conclusion. What are the consequences of violating the equal-subsample-size rule?

- A friend recommends White's test. For the auxiliary equation with cross terms, the student obtains $R^2 = 0.0433$. Complete the test, including its distribution, degrees of freedom, critical value, and conclusion.

Section B. Answer one question from this section (original Question 3 or Question 4).

QUESTION 15

Written Question 3 — 25 marks

Consider two regression models without an intercept:

$$Y_i = \beta_1 Z_i + \beta_2 X_i + u_i, \tag{1}$$

$$Y_i = \beta_2 X_i + u_i, \tag{2}$$

where

$$E(u_i) = 0, \qquad E(u_i^2) = \sigma^2, \qquad E(u_i u_j) = 0 \text{ for } i \neq j,$$

and Z and X are non-stochastic.

(a) (10 marks) Let (1) be the true model, but estimate (2) by OLS:

$$\widehat{Y} = \widehat{\beta}_2 X.$$

- Show that

$$\widehat{\beta}_2^* = \frac{\sum_i X_i Y_i}{\sum_i X_i^2}$$

is biased.

- Which factors determine the direction of the bias?
- Under what conditions is there no bias?

(b) (8 marks) Consider

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{3}$$

and

$$Y_i = \beta_2 X_i + u_i, \tag{4}$$

with the same disturbance assumptions, non-stochastic X, and $i = 1, \ldots, n$.

- Let (3) be true and estimate (4) by OLS. Rewrite (3) as $Y_i = \beta_1 Z_i + \beta_2 X_i + u_i$, where $Z_i = 1$. Using part (a), find the omitted-variable bias. How does $\overline{X}$ affect its direction, and together with which other factors?

- Now let (4) be true and estimate (3) by OLS. Briefly describe the consequences for the properties of $\widehat{\beta}_2$.

(c) (7 marks) Obtain the expression for

$$E\left(\widehat{\beta}_2^* - \widehat{\beta}_2\right)$$

by directly comparing

$$\widehat{\beta}_2^* = \frac{\sum_i X_i Y_i}{\sum_i X_i^2}$$

from regression (4) with

$$\widehat{\beta}_2 = \frac{\sum_i (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_i (X_i - \overline{X})^2}$$

from regression (3).

QUESTION 16

Written Question 4 — 25 marks

Consider the following simultaneous-equations model of equilibrium wheat consumption:

$$wheat^D = \alpha_1 price + \alpha_2 GDP + u_D, \tag{1}$$

$$wheat^S = \beta_1 price + \beta_2 sunshine + \beta_3 flood + u_S. \tag{2}$$

The errors u_D and u_S are i.i.d. with zero means and constant variances. Here:

- $wheat$ is per-capita wheat consumption, in kilograms;
- $price$ is the domestic price per kilogram;
- GDP is per-capita income;
- $sunshine$ is sunshine hours or solar radiation;
- $flood$ is the number of severe flooding events during the last wheat-growing season.

The variables GDP, $sunshine$, and $flood$ are exogenous; $wheat$ and $price$ are endogenous. The model is estimated using cross-sectional data.

(a) (10 marks)

- Explain endogenous and exogenous variables.
- Using the market-clearing condition $wheat^S = wheat^D$, derive reduced-form equations for $price$ and $wheat$.
- Briefly show why estimating equations (1) and (2) by OLS is problematic. A derivation of the large-sample bias is not required.

(b) (8 marks)

- What does it mean for an econometric equation to be identified or not identified? The order condition is not required for this definition.
- Determine whether equations (1) and (2) are exactly identified, underidentified, or overidentified using the order condition.

- A seminar participant proposes adding

$$sunshine^2 = (sunshine)^2,$$

so the supply equation becomes

$$wheat^S = \beta_1 price + \beta_2 sunshine + \beta_3 sunshine^2 + \beta_4 flood + u_S. \tag{2*}$$

What is the rationale for this proposal? What signs should be expected for the coefficients on $sunshine$ and $sunshine^2$?

- How does adding $sunshine^2$ change identification of equations (1) and (2*)?

(c) (7 marks) Which method can be used to obtain consistent estimates of equation (1)? Briefly describe how to apply it.