Эконометрика — МИЭФ, 2024 midterm 1

МИЭФ | Эконометрика | 2024 | midterm 1

QUESTION 1

Multiple-choice test

A change in the unit of measurement of the dependent variable in a model does not lead to a change in:

1. The standard error of the regression.
2. The sum of squared residuals of the regression.
3. The determination coefficient of the regression.
4. The confidence intervals of the regression.
5. All items 1)-4) remain unchanged.

QUESTION 2

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:

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

Which statement is correct?

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

QUESTION 3

Multiple-choice test

In the following equation, GDP is gross domestic product and FDI is foreign direct investment:

$$\log(GDP) = 2.65 + 0.527\log(bankcredit) + 0.222FDI,$$

with standard errors

$$(0.13) \qquad (0.022) \qquad (0.017).$$

Which statement is true?

1. If GDP increases by 1%, bank credit increases by 0.527%, holding FDI constant.
2. If bank credit increases by 1%, GDP increases by 0.527%, holding FDI constant.
3. If GDP increases by 1%, bank credit increases by $\log(0.527)$%, holding FDI constant.
4. If bank credit increases by 1%, GDP increases by $\log(0.527)$%, holding FDI constant.
5. If bank credit increases by 1 unit, GDP increases by 0.527%, holding FDI constant.

QUESTION 4

Multiple-choice test

A student estimates 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. He then estimates

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

Which statement is correct?

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

QUESTION 5

Multiple-choice test

A student regresses $\log(EARN)$ on S (years of schooling), $ASVABC$ (ability indicator), $FEMALE$, $ETHWHITE$, and the interaction $FEMALE \times ETHWHITE$. He then replaces the interaction by $MALE \times ETHNW$, where

$$MALE = 1 - FEMALE, \qquad ETHNW = 1 - ETHWHITE.$$

Which statement is correct?

1. The estimated slope coefficients of S and $ASVABC$ will generally change.

2. The coefficient of $MALE \times ETHNW$ will be the same as that of $FEMALE \times ETHWHITE$ in the initial regression.

3. The coefficient of $MALE \times ETHNW$ will have the same absolute value as that of $FEMALE \times ETHWHITE$, but the opposite sign.

4. The coefficients of $FEMALE$ and $ETHWHITE$ will stay the same.

5. The intercept will stay the same.

QUESTION 6

Multiple-choice test

In the regression model

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

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

1. The estimate of α will be biased upwards.
2. The estimate of α will be biased downwards.
3. The estimate of α will be unbiased.
4. The estimate of α may be biased upwards or downwards, depending on the sign of the mean value of x.
5. The estimate of α may be biased upwards or downwards, depending on the sign of α.

QUESTION 7

Multiple-choice test

Which term or terms in the general form of the t statistic are computed differently between the usual OLS t statistic and the heteroscedasticity-consistent t statistic?

1. Estimate, standard error, and hypothesized value.
2. Estimate only.
3. Standard error only.
4. Estimate and standard error.
5. Estimate and hypothesized value.

QUESTION 8

Multiple-choice test

The correct model specification is

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

but the fitted specification is

$$\widehat{Y} = \widehat{\beta}_1 + \widehat{\beta}_2 X_2.$$

The bias of the intercept estimate, $\text{bias}(\widehat{\beta}_1)$, is:

1. Equal to zero.
2. Strictly proportional to β_1.
3. Strictly proportional to β_2.
4. Strictly proportional to β_3.
5. None of the above.

QUESTION 9

Multiple-choice test

Which statement about measurement error is true?

1. If measurement error in an independent variable has zero mean, OLS estimators are unbiased and consistent.

2. If measurement error in an independent variable has zero mean, OLS estimators are biased but consistent.

3. If measurement error in an independent variable is uncorrelated with the variable, OLS estimators are unbiased.

4. If measurement error in a dependent variable is correlated with the independent variables, OLS estimators are unbiased.

5. None of the above.

QUESTION 10

Multiple-choice test

A student estimated a regression of real manufacturing on real GDP for 130 countries in 2022, without data for the USA and China. After receiving data for those two countries, he wants to check the model's quality for predicting GDP for them. He should:

1. Run the same regression for 132 observations and use the standard error of regression as the standard error of the prediction error for either country.

2. Add one common intercept dummy for the two countries, run the regression with 132 observations, and use the standard error of the dummy coefficient as the standard error of prediction error for both countries.

3. Add separate intercept dummies for the USA and China, run the regression with 132 observations, and use the standard errors of the two dummy coefficients as the standard errors of prediction errors.

4. Add one common intercept and one common slope dummy for the two countries, run the regression with 132 observations, and use the standard error of the new intercept dummy coefficient as the standard error of prediction error for both countries.

5. Add separate intercept and slope dummies for the USA and China, run the regression with 132 observations, and use the standard errors of the new intercept dummy coefficients as the standard errors of prediction errors.

QUESTION 11

Multiple-choice test

The following equations form a simultaneous equations model:

$$\begin{aligned} K_1&=\gamma_1+\alpha_1K_2+\beta_1Z_1+u_1, \tag{1}\\ K_2&=\gamma_2+\alpha_2K_1+\beta_2Z_2+u_2. \tag{2} \end{aligned}$$

The error term in the reduced-form equation for K_2 will be:

1. A quadratic function of u_1 and u_2, correlated with Z_1 and Z_2.
2. A quadratic function of u_1 and u_2, uncorrelated with Z_1 and Z_2.
3. A linear function of u_1 and u_2, correlated with Z_1 and Z_2.
4. A linear function of u_1 and u_2, uncorrelated with Z_1 and Z_2.
5. A linear function of u_1, u_2, γ_1, and γ_2, uncorrelated with Z_1 and Z_2.

QUESTION 12

Multiple-choice test

A student estimates the relationship between Gross Regional Product, GRP, and the Gini coefficient, $GINI$, for 22 provinces of China in 2022:

$$GRP_{i,2022} = \alpha_1 + \alpha_2 GINI_{i,2022} + \alpha_3 GRP_{i,2021}, \quad (1)$$

$$GINI_{i,2022} = \beta_1 + \beta_2 GINI_{i,2021} + \beta_3 GRP_{i,2022} + \beta_4 GRP_{i,2021}. \quad (2)$$

Then:

1. Equation (1) is underidentified.
2. Equation (1) is overidentified.
3. Equation (2) is overidentified.
4. Equation (2) is exactly identified.
5. None of the above.

Part 2. Free Response Questions — 1 hour 30 minutes.

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

QUESTION 13

Written Question 1 — 25 marks

A student investigates the market for private mathematics teachers in Moscow, with particular interest in those who can teach in English. He takes a random sample of 30 teacher profiles from 300 profiles registered on an internet site.

Variables are:

- $PRICE_i$: price of a standard two-hour lesson, in thousands of roubles;
- $DIST_i$: number of metro stations from the centre of Moscow to the teacher's place;
- $HOME_i$: dummy equal to 1 if the tutor visits the client;
- ENG_i: dummy equal to 1 if the tutor can teach mathematics in English.

The estimated regressions are:

$$\widehat{PRICE}_i = 6.59 - 0.16 DIST_i, \qquad R^2 = 0.185, \tag{1}$$

$$\widehat{PRICE}_i = 4.51 + 2.54 HOME_i, \qquad R^2 = 0.40, \tag{2}$$

$$\widehat{PRICE}_i = 5.13 - 0.08 DIST_i + 1.95 HOME_i + 0.27 DIST_i HOME_i, \qquad R^2 = 0.437, \tag{3}$$

$$\widehat{PRICE}_i = 4.52 - 0.08 DIST_i + 2.18 HOME_i + 1.58 ENG_i - 0.39 HOME_i ENG_i, \qquad R^2 \tag{4}$$

(a) (12 marks)

- Interpret equation (1). What disadvantages does it have in the context of the problem?
- Interpret equation (2). What disadvantages does it have?
- How does the meaning and the set of assumptions in equation (3) differ from equation (1)?
- Interpret all coefficients in equation (4).

(b) (13 marks)

- Is the factor "distance", represented jointly by $DIST_i$ and $DIST_i HOME_i$, significant in equation (3)?
- Is the factor "teaching at the student's place", represented jointly by $HOME_i$ and $DIST_i HOME_i$, significant in equation (3)?
- Are all dummy variables included in equation (4) jointly significant?
- Which variables are missing from models (3) and (4) if the student wants to evaluate the full significance of both dummy variables? How can this be done?

QUESTION 14

Written Question 2 — 25 marks

A researcher studies factors affecting the volume of paid services per capita V_i in 82 Russian regions, measured in roubles. He suggests that it depends primarily on average monthly income per capita I_i, which ranges from 14,000 to 70,000 roubles.

(a) (13 marks) The researcher estimates

$$\widehat{V_i} = -370.7 + 0.17 I_i, \qquad R^2 = 0.78. \tag{1}$$

Conventional standard errors are

$$(295.6) \qquad (0.010),$$

and White heteroscedasticity-consistent standard errors are

$$[614.3] \qquad [0.024].$$

The Breusch-Pagan statistic $Obs \cdot R^2$ equals 26.79.

- What is heteroscedasticity? Explain how it could arise in this setting.
- Interpret equation (1). Which characteristics of the estimates may indicate heteroscedasticity?
- Explain the Breusch-Pagan test and how to use the statistic $Obs \cdot R^2 = 26.79$. Are there signs of heteroscedasticity? What should the researcher do next?

(b) (12 marks) To eliminate heteroscedasticity, the researcher estimates

$$\frac{\widehat{V_i}}{I_i} = 0.19 - 809.2 \left(\frac{1}{I_i}\right), \qquad R^2 = 0.095,$$

with standard errors

$$(0.01) \qquad (279.6), \tag{2}$$

and a Breusch-Pagan statistic $Obs \cdot R^2 = 1.14$.

He also estimates

$$\log \widehat{V_i} = -3.53 + 1.16 \log I_i, \qquad R^2 = 0.81,$$

with standard errors

$$(0.65) \qquad (0.064), \tag{3}$$

and a Breusch-Pagan statistic $Obs \cdot R^2 = 1.73$.

- Explain why specifications (2) and (3) may eliminate heteroscedasticity.
- Was each method successful? Explain.
- In equation (2), R^2 is much smaller than in equation (1), and the coefficient on the explanatory variable appears negative. Does this indicate poor statistical quality? Restore the dependence of V_i on I_i estimated by WLS.
- Why is the slope coefficient in equation (3) so different from that in equation (1)?

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

QUESTION 15

Written Question 3 — 25 marks

Consider the model without an intercept:

$$y_t = \beta x_t + u_t, \qquad t = 1, 2, \ldots, T.$$

The regressor x_t is measured with error. Only

$$x_t^* = x_t + v_t$$

is observed, with

$$E(u_t) = E(v_t) = 0,$$

$$E(u_t v_t) = E(x_t u_t) = E(x_t v_t) = 0.$$

(a) (12 marks)

- Let $\widehat{\beta}$ be the OLS estimator from regressing y_t on x_t^*. Show that $\widehat{\beta}$ is inconsistent.
- Explain endogeneity. Explain why measurement error in this setting produces endogeneity.

(b) (7 marks) Now suppose x_t is measured without error, but y_t is measured with error:

$$y_t^* = y_t + w_t,$$

where

$$E(w_t) = 0, \qquad E(u_t w_t) = E(x_t w_t) = E(v_t w_t) = 0.$$

- Let $\widehat{\beta}$ be the OLS estimator from regressing y_t^* on x_t. Is it consistent? Explain in detail.
- Briefly describe the implications for OLS when measurement errors are simultaneously present in the independent variable x_t and the dependent variable y_t.

(c) (6 marks) Suppose there is also an instrument z_t for x_t in the setting from part (a).

- Explain in detail how to perform a Hausman test in Davidson-MacKinnon form to determine how serious the endogeneity problem caused by measurement error in x_t

is.

- How can the test results be used? What options are available, and what are their comparative advantages and disadvantages?

QUESTION 16

Written Question 4 — 25 marks

An economist investigates the relationship between wages and prices:

$$w_t = \alpha_0 + \alpha_1 p_t + \alpha_2 y_t + \alpha_3 z_t + u_t, \tag{1}$$

$$p_t = \beta_0 + \beta_1 w_t + \beta_2 y_t + \beta_3 z_t + v_t, \tag{2}$$

where w is the rate of growth of money wages, p is the rate of growth of prices, y is the rate of growth of productivity, and z is the unemployment rate. Both y and z are assumed exogenous.

(a) (9 marks) The economist first considers the simplified version obtained by setting

$$\alpha_2 = \beta_2 = \beta_3 = 0 :$$

$$w_t = \alpha_0 + \alpha_1 p_t + \alpha_3 z_t + u_t, \tag{3}$$

$$p_t = \beta_0 + \beta_1 w_t + v_t. \tag{4}$$

- Derive the reduced-form system for (3)-(4).
- What does the reduced form imply about the properties of the OLS estimators for both structural equations?
- What is identification? What can be said about identification of equations (3) and (4)?
- Give an example of restrictions on $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3$ under which both equations of the general model become exactly identified.

(b) (8 marks) The economist next sets

$$\alpha_2 = \alpha_3 = 0 :$$

$$w_t = \alpha_0 + \alpha_1 p_t + u_t, \tag{5}$$

$$p_t = \beta_0 + \beta_1 w_t + \beta_2 y_t + \beta_3 z_t + v_t. \tag{6}$$

- What can be said about identification of equations (5)-(6)? Explain using the order condition.

- Which methods can consistently estimate equation (5)? No details are required.
- How would your conclusions change if w_{t-1} were used on the right-hand side of equation (6) instead of w_t? How would this affect the choice of a consistent estimation method for both equations?

(c) (8 marks)

- Briefly explain TSLS. In which cases do TSLS estimates:
    1. outperform other methods;
    2. provide no benefits;
    3. become inapplicable?
- For equations (5)-(6), explain how to use TSLS to estimate α_1 using the available instruments. State what should be done in the first and second stages and why the resulting estimator is consistent.