Эконометрика — МИЭФ, 2023 midterm 1

МИЭФ | Эконометрика | 2023 | midterm 1

Рисунки пока рендерятся в тестовом режиме и могут отличаться от исходных материалов.

QUESTION 1

Multiple-choice test

Which of the following statements is true?

1. If the calculated value of the F statistic is higher than the critical value, we reject the alternative hypothesis in favor of the null hypothesis.

2. The F statistic is always nonnegative as SSR_r is never smaller than SSR_{ur}.

3. Degrees of freedom of a restricted model is always less than the degrees of freedom of an unrestricted model.

4. The F statistic is more flexible than the t statistic to test a hypothesis with a single restriction.

5. None of the above.

QUESTION 2

Multiple-choice test

In a regression model, if variance of the dependent variable Y, conditional on an explanatory variable X, is not constant, then:

1. The t statistics are invalid and confidence intervals are valid for small sample sizes.

2. The t statistics are valid and confidence intervals are invalid for small sample sizes.

3. The t statistics and confidence intervals are valid no matter how large the sample size is.

4. The t statistics and confidence intervals are both invalid no matter how large the sample size is.

5. The OLS estimators are biased, and hence no need to discuss t statistics and confidence intervals.

QUESTION 3

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. The disturbance term is correlated with the dependent variable.

3. One or more of the explanatory variables is jointly determined with the dependent variable.

4. Heteroscedasticity is present in the model.

5. There is correlation between some explanatory variables.

QUESTION 4

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 estimator

$$b = \frac{\sum_{i=2}^{n}(Y_i - Y_{i-1})}{\sum_{i=2}^{n}(X_i - X_{i-1})}$$

is, generally speaking:

1. An unbiased and efficient estimator of β_2.
2. An unbiased but inefficient estimator of β_2.
3. A biased estimator of β_2.
4. A non-linear estimator of β_2.
5. Non-stochastic.

QUESTION 5

Multiple-choice test

For the sample of 55 observations, functions (1) and (2) were estimated:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u \tag{1}$$

$$Y = \beta_0 + \beta_1 (X_1 - X_2) + u. \tag{2}$$

The determination coefficients for these models are $R_1^2 = 0.9$ and $R_2^2 = 0.7$, respectively. The F statistic for testing the hypothesis $\beta_1 = \beta_2$ in (1) equals:

1. 6.7.
2. 8.2.
3. 30.
4. 60.
5. You cannot test this hypothesis using (1) and (2).

QUESTION 6

Multiple-choice test

The function of expenditures for cosmetics depending on disposable personal income has been estimated using OLS for a representative sample of people:

$$Y = \beta_0 + \beta_1 D_1 + \beta_2 X + \beta_3 X(1 - D_2) + u,$$

where Y is expenditure for cosmetics, X is disposable personal income, $D_1 = 1$ for females and 0 for males, and $D_2 = 1$ for males and 0 for females.

For this regression the following is correct:

1. The estimates of intercept are the same for male and female subsamples, while the estimates of slope coefficient, generally speaking, differ for them.

2. The estimates of slope coefficient are the same for male and female subsamples, while the estimates of intercept, generally speaking, differ for them.

3. Both intercepts and slope coefficients estimated, generally speaking, differ for male and female subsamples.

4. Both intercepts and slope coefficients estimated are the same for male and female subsamples.

5. The combination of intercept and slope dummies is incorrect, and the model cannot be estimated.

QUESTION 7

Multiple-choice test

If you have estimated the parameters of the following model using OLS directly, with the Gauss-Markov conditions satisfied,

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \big(\beta_2(1+\beta_3)\big)x_3 + u,$$

then:

1. You can get an unbiased estimate of β_3.
2. You cannot get an unbiased estimate of β_3, but can easily get a consistent estimate of it.
3. You cannot get an unbiased, or biased but consistent, estimate of β_3.
4. You cannot get any estimate of β_3.
5. All the above statements are incorrect.

QUESTION 8

Multiple-choice test

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

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

The formula for the homoscedasticity case is

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

Let $\sigma_i^2 = \sigma^2 k_i$, where k_i are unknown weights and $\sum k_i = 1$. 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 nature of the relationship between σ_i and x_i.

QUESTION 9

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 $\beta < 0$ and the mean value of x is negative. When estimating the model using OLS, for large samples:

1. The estimator of α will be biased upwards.
2. The estimator of α will be biased downwards.
3. The estimator of α will be unbiased.
4. The estimator of α may be biased upwards or downwards.
5. The OLS estimator of α does not exist.

QUESTION 10

Multiple-choice test

For a simultaneous equations model with 7 equations, 7 endogenous variables and 7 exogenous variables, the following statement is true:

1. With that number of potential instruments, any equation is identified in the model.
2. An equation in the model is identified if and only if only exogenous variables are available on its right-hand side.
3. The number of potential instruments is insufficient to make all the equations identified.
4. No equation can be overidentified in the model.
5. None of the above.

QUESTION 11

Multiple-choice test

The economic model is described by the following simultaneous equations:
\begin{aligned} y_1&=\delta+\tau y_2+\pi x_2+u_2, \tag{1}\\ y_2&=\alpha+\pi y_1+\gamma x_1+\phi x_2+u_1. \tag{2} \end{aligned}

Here y_1 and y_2 are endogenous variables; x_1 and x_2 are stochastic exogenous variables; and u_1 and u_2 are disturbance terms satisfying the Gauss-Markov conditions. Indicate the correct statement:

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

QUESTION 12

Multiple-choice test

The model with the dependent variable P_i (monthly pension), as a function of work experience WE_i and average earnings $EARN_i$, is being considered:

$$P_i = \beta_1 + \beta_2 WE_i + \beta_3 EARN_i + u_i.$$

The value of pension is restricted by the values P_U and P_L from the top and from the bottom, but there are no actual observations in the sample with $P = P_U$ or $P = P_L$. The student decided to estimate a Tobit model for this sample. Indicate the correct statement:

1. The Tobit estimators of the model coefficients are biased and inconsistent.
2. The Tobit estimators of the model coefficients are biased but consistent.
3. The Tobit model estimates will be the same as the OLS estimates here.
4. The Tobit model may not be estimated for this sample.
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 is investigating factors that affect schoolchildren's consumption of unhealthy food at fast-food restaurants, such as McDonald's. Let Y_i be the average number of hamburgers consumed per month in 2021 and let X_i be age. The student wants to understand whether the dependence differs between boys and girls. She introduces a dummy variable D_i equal to 1 for boys and 0 for girls.

Using a sample of 17 boys and 13 girls, for a total of 30 observations, she first runs the simple regression

$$\widehat{Y_i} = -0.56 + 0.24X_i, \qquad R^2 = 0.17,$$

with standard errors

$$(0.53) \qquad (0.10). \tag{1}$$

Assuming that boys eat more frequently at a fast-food restaurant, she defines a slope dummy variable $(XD)_i = X_i D_i$ and fits the regression

$$\widehat{Y_i} = -1.43 + 0.19X_i + 0.52D_i + 0.78(XD)_i, \qquad R^2 = 0.36,$$

with standard errors

$$(0.36) \qquad (0.07) \qquad (0.33) \qquad (0.42). \tag{2}$$

(a)

- What is the meaning of the coefficients of regression (2)?
- Is there any difference in the influence of X_i on Y_i between boys and girls? How should the significance of this difference be tested?
- Can the answer to the previous question be obtained using the Chow test? What additional information is needed, and how can it be obtained?
- Can the answer to the previous question be obtained using the Chow test? What additional information is needed, and how can it be obtained?

(b) When the student showed her results to the supervisor, the supervisor advised her to evaluate a simplified regression of the form

$$Y_i = \beta_0 + \beta_1 D_i + u_i. \tag{3}$$

This regression does not take into account the effect of age X_i. The student did not have a computer with her to recalculate the coefficients. The supervisor noted that it is sufficient to know the average number of hamburgers consumed per month for girls, $\overline{Y}_0$, and boys, $\overline{Y}_1$, because it can be shown that

$$\beta_1 = \overline{Y}_1 - \overline{Y}_0 \qquad \text{and} \qquad \beta_0 = \overline{Y}_0.$$

- Show that these statements are true for regression (3).
- Provide the intuition behind these statements.

QUESTION 14

Written Question 2 — 25 marks

A student tries to determine how expenditure on education E_i, in billions of dollars, relates to GDP Y_i, in billions of dollars, and population P_i, in millions, using data on 34 developed and developing countries with high, medium, and low aggregate income for 2020. Here and below, e_i denotes the regression residual.

She estimates

$$\widehat{E}_i = -4.52 + 0.043Y_i + e_i, \qquad R^2 = 0.75, \qquad i = 1, \ldots, 34,$$

with standard errors

$$(3.40) \qquad (0.004). \tag{1}$$

(a)

- Why may the student fear the presence of heteroscedasticity? Explain using your understanding of heteroscedasticity.
- How could heteroscedasticity influence the regression results?
- How can heteroscedasticity be detected using graphs? Specify the relevant graphs.
- The student arranges countries by Y_i and runs two regressions in specification (1). For the 10 countries with the highest Y_i values, she obtains $SSR_1 = 5795.4$. For the 14 countries with the lowest Y_i values, she obtains $SSR_2 = 41.7$. Conduct an appropriate heteroscedasticity test using this information.
- The supervisor advises the student to use per-capita values E_i/P_i and Y_i/P_i instead of the absolute values E_i and Y_i. The student again arranges countries by Y_i/P_i. For the 10 countries with the highest Y_i/P_i, she obtains $SSR_1 = 0.19$; for the 14 countries with the lowest Y_i/P_i, she obtains $SSR_2 = 0.33$. Explain the idea behind this advice and assess its usefulness using an appropriate test.

(b) The student next estimates the multiple regression

$$\widehat{E}_i = -1.57 - 0.0056Y_i + 0.88P_i + e_i, \qquad R^2 = 0.98,$$

with standard errors

$$(0.94) \qquad (0.0027) \qquad (0.044). \tag{2}$$

- Compare the coefficients on Y_i in equations (1) and (2). How have the meaning, value, and significance of the coefficient changed? Which value seems more reasonable, and why?

- To test equation (2) for heteroscedasticity, the student uses the Breusch-Pagan test and obtains $R^2 = 0.38$ for the auxiliary regression. Complete the test, describe the procedure, and state the result.

On the advice of a friend, the student estimates the model in logarithms:

$$\ln E_i = 9.63 - 0.37 \ln Y_i + 1.37 \ln P_i + e_i, \qquad R^2 = 0.95,$$

with standard errors

$$(0.34) \qquad (0.09) \qquad (0.09). \tag{3}$$

For equation (3), the auxiliary regression has $R^2 = 0.16$ for the Breusch-Pagan test and $R^2 = 0.36$ for the White test with cross terms.

- Why might using logarithms help eliminate heteroscedasticity? Did it help according to the Breusch-Pagan and White tests? Complete both tests. Why do their results differ, and which test should be trusted more?

- What would you advise the student to do to eliminate heteroscedasticity in equation (2)?

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

QUESTION 15

Written Question 3 — 25 marks

A student in ICEF's econometrics course uses data on a sample of 100 students to study which factors determine the score Y_i, out of 100 points, on the winter econometrics exam. Since econometrics relies heavily on statistics, one possible factor is the student's knowledge of statistics Z_i:

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

Direct measurement of Z_i is not possible. The available variable is S_i, the score, also out of 100 points, obtained in the second-year statistics exam. Because students were nervous during this exam, the student assumes a measurement error:

$$S_i = Z_i + w_i,$$

where w_i is independent of Z_i and u_i, with

$$E(w_i) = 0, \qquad \mathrm{Var}(w_i) = \sigma_w^2.$$

Using OLS, she obtains

$$\widehat{Y_i} = -8.26 + 0.80 S_i, \qquad R^2 = 0.42,$$

with standard errors

$$(6.00) \qquad (0.09). \tag{2}$$

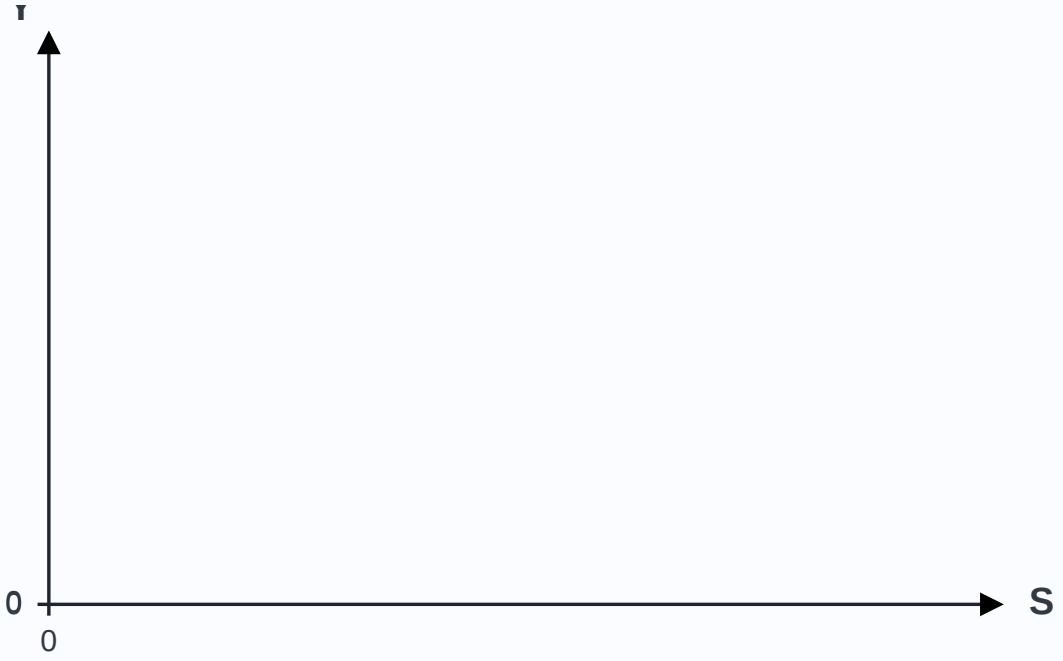
(a)

- What are the consequences of measurement error in the regressor when estimating β_2 by OLS?
- A friend points out that the statistics exam was graded very harshly, with students' grades lowered, so it may be more appropriate to assume $E(w_i) = \mu_w < 0$. What additional consequences for estimation of β_2 by OLS does this assumption have?

(b)

- What are the consequences of measurement error when estimating the intercept β_1 by OLS under the assumption $E(w_i) = 0$?

- Illustrate graphically the result obtained in the previous question for the estimation of β_1.



Effect of attenuation bias on the fitted intercept

(c) The student also has grades in other subjects: M_i for mathematics, B_i for banking, L_i for linear algebra, and others. She assumes that these variables are not subject to measurement error. She regresses S_i on all these variables, saves the residuals E_i, and includes them in the equation

$$\widehat{Y_i} = -20.69 + 1.00 S_i - 0.47 E_i, \qquad R^2 = 0.46,$$

with standard errors

$$(7.64) \qquad (0.12) \qquad (0.19). \tag{3}$$

- Comment on the aim and logic of this procedure and on the obtained results.

- During the winter econometrics exam, students were also nervous, introducing measurement error v_i into Y_i, with $E(v_i) = 0$ and $\mathrm{Var}(v_i) = \sigma_v^2$. How does this assumption affect the properties of the estimate of β_2 in equation (2)?

- Would your conclusions change if, because the econometrics exam was just before New Year, graders were instructed to resolve all controversial cases in favor of students, so that $E(v_i) = a > 0$? No rigorous derivation is required.

QUESTION 16

Written Question 4 — 25 marks

After graduating from university, a student joins a consulting firm dealing with the promotion of candidates and the organization of online elections. Her first assignment is to advise potential candidate A for the position of head of the student organization.

Candidate A is young and inexperienced. He can spend only $2,000 on advertising, but respondents rank his attractiveness as 5. Candidate B is more experienced and plans to spend $5,000 on advertising, with an attractiveness rank of 2. Each candidate receives 3 free appearances on local television; additional appearances cost $799 each.

The student has data from 200 past elections:

- V: number of votes cast for the candidate;
- E: binary variable equal to 1 if the candidate was elected;
- AD: amount spent on promotion, in thousands of US dollars;
- TV: number of appearances on television special events;
- APP: personal appeal of the candidate, on a scale from 1 to 5.

Using these data, she estimates the following models. Standard errors, or their counterparts, are in parentheses. For the probit model,

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

is the standard normal probability density function.

$$\widehat{V_i} = -41.60 + 25.15 AD_i + 32.64 TV_i + 21.60 APP_i, \qquad R^2 = 0.66,$$

with standard errors

$$(17.84) \qquad (2.08) \qquad (2.75) \qquad (4.72). \qquad \text{(1-OLS)}$$

$$\widehat{E_i} = -0.74 + 0.14 AD_i + 0.18 TV_i + 0.11 APP_i, \qquad R^2 = 0.51,$$

with standard errors

$$(0.13) \qquad (0.016) \qquad (0.02) \qquad (0.04). \qquad \text{(2-OLS)}$$

$$\widehat{E}_i = -5.64 + 0.64AD_i + 0.75TV_i + 0.54APP_i, \qquad \text{McFadden } R^2 = 0.49,$$

with standard errors

$$(0.94) \qquad (0.13) \qquad (0.12) \qquad (0.19). \qquad \text{(3-Probit)}$$

(a)

- Explain the meaning of regression (1). Compare the candidates' chances based on model (1), assuming that a higher expected number of votes is an indicator of success.
- Explain the meaning of regression (2) and its coefficients.
- Explain the logic of model (3), including the mechanism used to obtain its regression results.

(b) According to model (3), what are the chances for each candidate to be elected? Compare them with the results of model (2). Which model can be trusted more?

Recall the candidate data:

- Candidate A: $AD = 2$, $TV = 3$, $APP = 5$;
- Candidate B: $AD = 5$, $TV = 3$, $APP = 2$.

(c) Using the marginal effects of advertising and television appearances in model (3), advise Candidate A how to reallocate his funds between advertising and television appearances to close the gap with Candidate B or overtake him. One additional television appearance costs $799, and the candidate has no additional funds. Show with calculations that the proposed reallocation can improve Candidate A's chances.