Эконометрика — МИЭФ, 2023 final

МИЭФ | Эконометрика | 2023 | final

QUESTION 1

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 2

Multiple-choice test

Which of the following correctly identifies an advantage of using adjusted R^2 over R^2?

1. Adjusted R^2 corrects the bias in R^2.

2. Adjusted R^2 is easier to calculate than R^2.

3. The penalty of adding new independent variables is better understood through adjusted R^2 than R^2.

4. Adjusted R^2 can be calculated for models having logarithmic functions, while R^2 cannot be calculated for such models.

5. None of the above is correct.

QUESTION 3

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 4

Multiple-choice test

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

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

then:

1. You can get an unbiased estimate of β_3.
2. You cannot get an unbiased estimate of β_3, but can 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 5

Multiple-choice test

Which of the following correctly defines the F statistic for testing linear restrictions if R_r^2 represents the coefficient of determination from the restricted model, R_{ur}^2 represents the coefficient of determination from the unrestricted model, and q is the number of restrictions imposed?

1. $$F = \frac{(R_{ur}^2 - R_r^2)/q}{(1 - R_{ur}^2)/(n-k)}.$$

2. $$F = \frac{(R_r^2 - R_{ur}^2)/q}{(1 - R_{ur}^2)/(n-k)}.$$

3. $$F = \frac{(R_{ur}^2 - R_r^2)/q}{(1 - R_r^2)/(n-k)}.$$

4. $$F = \frac{(R_r^2 - R_{ur}^2)/q}{(1 - R_r^2)/(n-k)}.$$

5. None of the above.

QUESTION 6

Multiple-choice test

The following double-logarithmic model is estimated:

$$\log Y = \beta_1 + \beta_2 \log X_2 + u.$$

The interpretation of the coefficient β_2 is the following:

1. If X_2 increases by one unit, then Y increases approximately by $100\beta_2$ percent.
2. If X_2 increases by one unit, then Y increases approximately by $\beta_2/100$ percent.
3. If X_2 increases by one percent, then Y increases approximately by $100\beta_2$ percent.
4. If X_2 increases by one percent, then Y increases approximately by β_2 percent.
5. If X_2 increases by one percent, then Y increases approximately by β_2 units.

QUESTION 7

Multiple-choice test

An econometric model is described by the following three equations:

$$\begin{aligned} y_1&=\alpha+\beta y_3+\gamma x_1+\sigma x_3+\pi x_4+u_1, \tag{1}\\ y_2&=\delta+\varepsilon y_1+\lambda x_2+u_2, \tag{2}\\ y_3&=\mu+\theta y_1+\omega y_2+\rho x_3+\chi x_4+u_3. \tag{3} \end{aligned}$$

Here y_1, y_2, and y_3 are endogenous variables; x_1, x_2, x_3, and x_4 are exogenous variables; and u_1, u_2, and u_3 are disturbance terms, independent and satisfying the Gauss-Markov conditions. Choose the correct statement:

1. Equation (2) is exactly identified.
2. Equation (1) is overidentified.
3. Equation (3) is underidentified.
4. Equation (1) is exactly identified.
5. Equation (2) is underidentified.

QUESTION 8

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 $P_i = P_U$ or $P_i = P_L$. The student decided to estimate a Tobit model with the truncated sample, with all observations on the upper or lower bounds excluded. Please indicate the correct statement among the following ones:

1. The estimated coefficients are biased but consistent.
2. The estimated coefficients are biased and inconsistent.
3. The estimated coefficients are unbiased.
4. For the truncated sample, OLS estimation would provide unbiased estimates.
5. None of the above.

QUESTION 9

Multiple-choice test

The following model of determination of the size of dividends is considered:

$$D_t^* = \gamma P_t + u_t, \tag{1}$$

$$\Delta D_t = \lambda(D_t^* - D_{t-1}) + \rho(P_t - P_{t-1}), \tag{2}$$

where D_t^* is the desirable size of the dividends, P_t is the current profits, D_t is the actual size of the dividends, and $\Delta D_t = D_t - D_{t-1}$. The following statement is correct. The model is:

1. The adaptive expectations model and can be consistently estimated in the form of the Koyck distribution model.

2. The partial adjustment model and can be consistently estimated in the form of the ADL(1,0) model.

3. The partial adjustment model and can be consistently estimated in the form of the ADL(0,1) model.

4. The error correction model and can be consistently estimated in the form of the ADL(1,1) model.

5. The error correction model and can be consistently estimated in the form of the ADL(1,0) model.

QUESTION 10

Multiple-choice test

Refer to the following model:

$$Y_t = \alpha_0 + \beta_0 S_t + \beta_1 S_{t-1} + \beta_2 S_{t-2} + \beta_3 S_{t-3} + u_t.$$

Here $(\beta_0 + \beta_1)$ represents:

1. The short-run change in Y given a temporary increase in S.
2. The short-run change in Y given a permanent increase in S.
3. The long-run change in Y given a permanent increase in S.
4. The long-run change in Y given a temporary increase in S.
5. None of the above.

QUESTION 11

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 12

Multiple-choice test

In the model based on panel data

$$Y_{it} = \beta_1 + \sum_{j=2}^{k} \beta_j X_{jit} + \sum_{p=1}^{s} \gamma_p Z_{pi} + \delta t + \varepsilon_{it},$$

random effect estimation is based on the following assumptions:

I. There are no X variables that are fixed for each individual.

II. There is some unobserved heterogeneity in the model.

III. Each of the unobserved Z_p variables is treated as being drawn randomly from a given distribution.

IV. The Z_p variables are correlated with some of the X_j variables.

V. The Z_p variables are distributed independently of all of the X_j variables.

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

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.

Working on her coursework, a student of ICEF interviewed ICEF graduates of different graduation years working in Russia. She is interested in studying their current earning, $earn_i$, in thousands of rubles per month. Explanatory variables are age_i, age of respondent in years, age squared age_i^2, and also some dummy variables: $msca_i$, equal to 1 for those graduates who have received a master's degree abroad and 0 otherwise; nfe_i, no further education, equal to 1 for those graduates who received a master's degree neither abroad nor in the country; and $male_i$, equal to 1 for male and 0 for female. She posted a questionnaire on the Internet and received answers from 41 graduates of different years of graduation from ICEF. Here are the results of estimation of two regressions using different sets of variables. Standard errors are in brackets.

$$\widehat{earn}_i = 75.18 + 2.37age_i - 0.02age_i^2, \qquad R^2 = 0.48. \tag{1}$$

Standard errors:

$$(4.18) \qquad (0.21) \qquad (0.01)$$

$$\widehat{earn}_i = 123.68 + 2.54age_i - 0.03age_i^2 + 40.21msca_i - 51.23nfe_i + 0.25male_i, \qquad R^2 = 0.63 \tag{2}$$

Standard errors:

$$(6.49) \qquad (0.34) \qquad (0.007) \qquad (4.56) \qquad (22.37) \qquad (0.16)$$

(a) (12 marks)

- How many groups of dummy variables does equation (2) contain? How many categories of education level of ICEF graduates do dummy variables $msca_i$ and nfe_i describe? What is the reference category in each of the groups of dummy variables?

- Help the student estimate the expected earnings of ICEF graduates of different categories presented in equations (1) and (2) for a person 25 years old. Why are the coefficients of equations (1) and (2) different, and what is the difference in the meaning of the estimates obtained from equations (1) and (2)?

- Are the coefficients of the variables $msca_i$, nfe_i, and $male_i$ significant? Are they jointly significant?

(b) (13 marks) The student found that the variables $msca_i$, nfe_i, and $male_i$ do not correlate with age_i and with age_i^2.

- Can the effects of age be considered independently of the values of other variables?

- What is the meaning of the coefficients of age_i and age_i^2 in equation (2)? Evaluate the marginal effect of age for $age = 25$, $age = 42$, and $age = 60$ and discuss the results. Is the influence of age on earnings significant?

- What would be the consequences for evaluating equation (2) if the variable age_i^2 was excluded from it? Explain based on your knowledge of econometric theory.

- Fearing the presence of heteroscedasticity, the student runs the Breusch-Pagan test for equation (2), obtaining the value of the statistic $\chi^2 = 17.5$. For equation (1), she performs a White test with cross-terms included, obtaining $\chi^2 = 10.2$. Help the student complete the tests for heteroscedasticity and draw conclusions. Explain your answer.

QUESTION 14

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

The student decided to investigate the factors that affect expenditures on air travel in the United States. To do this, she uses data from the 25 years, 1994-2018, prior to the outbreak of the Covid-19 pandemic, on total expenditure la_t, total income ld_t, and the air travel relative price index lp_t, all taken in logarithms. She first builds the following regressions using OLS and Cochrane-Orcutt (C.O.) methods. Standard errors are in parentheses.

$$\widehat{la_t} = -12.7 + 2.1 ld_t, \qquad R^2 = 0.47, \qquad DW = 0.31 \qquad \text{OLS.} \tag{1}$$

Standard errors:

$$(0.68) \qquad (0.10)$$

$$\widehat{la_t} = -7.5 + 1.3 ld_t, \qquad R^2 = 0.98, \qquad DW = 1.40 \qquad \text{C.O.}. \tag{2}$$

Standard errors:

$$(5.9) \qquad (0.84)$$

$$\widehat{la_t} = -9.6 + 2.3 ld_t - 0.99 lp_t, \qquad R^2 = 0.99, \qquad DW = 1.46 \qquad \text{OLS.} \tag{3}$$

Standard errors:

$$(0.40) \qquad (0.05) \qquad (0.09)$$

$$\widehat{la_t} = -9.4 + 2.2 ld_t - 0.97 lp_t, \qquad R^2 = 0.99, \qquad DW = 1.88 \qquad \text{C.O.}. \tag{4}$$

Standard errors:

$$(0.54) \qquad (0.07) \qquad (0.11)$$

(a) (13 marks)

- Why can one suggest the presence of autocorrelation in some of the equations listed above? Why is this question important when evaluating regression equations? Help the student explore this question using the Durbin-Watson test.

- For what purpose does the student, along with equation (1), also calculate equations (2), (3), and (4)? Explain your opinion.

- Has she been able to achieve her goals? Is there any reason to believe that there is no autocorrelation in equation (4), or should the student be advised to take an additional test? Which one?

A student's friend advised her to use a lagged variable as the best and simple tool to make the DW statistic acceptable. The corresponding equation is

$$\widehat{la}_t = -4.68 + 1.3ld_t - 0.73lp_t + 0.41la_{t-1}, \qquad R^2 = 0.99, \qquad DW = 2.32 \qquad \text{OLS.} \quad (5)$$

Standard errors:

$$(1.31) \qquad (0.26) \qquad (0.10) \qquad (0.11)$$

- Do you agree with the advice of the student's friend? Help her to test equation (5) for autocorrelation.

(b) (12 marks) The supervisor advised the student to consider the more general model ADL(1,1),

$$la_t = \beta_1 + \beta_2 ld_t + \beta_3 ld_{t-1} + \beta_4 lp_t + \beta_5 lp_{t-1} + \beta_6 la_{t-1} + u_t,$$

and conduct a Common Factor test for this model. The corresponding estimated models are as follows.

Unrestricted model

$$\widehat{la}_t = -5.5 + 1.4ld_t + 0.11ld_{t-1} - 0.65lp_t - 0.18lp_{t-1} + 0.32la_{t-1}, \qquad R^2 = 0.99, \qquad RSS \quad (6)$$

Standard errors:

$$(1.9) \qquad (0.61) \qquad (0.67) \qquad (0.18) \qquad (0.27) \qquad (0.10)$$

Restricted model

$$\widehat{la}_t = -7.0 + 2.24ld_t + 0.11ld_{t-1} - 0.96lp_t - 0.18lp_{t-1} + 0.97la_{t-1}, \qquad R^2 = 0.99, \qquad RSS \quad (7)$$

Standard errors shown in the source:

$$(2.3) \qquad (0.07) \qquad (0.11) \qquad (0.22)$$

- Demonstrate how to obtain the restricted specification of the ADL(1,1) model from the multiple regression model

$$la_t = \alpha_1 + \alpha_2 ld_t + \alpha_3 lp_t + u_t$$

with the autocorrelated disturbance term

$$u_t = \rho u_{t-1} + \varepsilon_t.$$

- Help the student to run the Common Factor test, stating the restrictions and making a conclusion.

- Under what conditions is the Common Factor test valid? Explain how the student can test that these conditions are met.

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) (10 marks)

- Explain what is meant by a stationary time series and a non-stationary time series. How to understand if a time series is stationary?

- What is detrending of a time series? What is differencing of a time series? Explain what you understand by difference-stationary and trend-stationary time series. What is the difference in the impact of random shocks on difference-stationary and trend-stationary time series?

- Demonstrate that the time trend

$$X_t = \alpha_0 + \alpha_1 t + u_t$$

is a trend-stationary time series. There is no need to prove that this series is non-stationary. We assume

$$E[u_t] = 0, \qquad Var(u_t) = \sigma^2, \qquad E[u_t u_s] = 0 \quad \forall s \neq t.$$

- Demonstrate that the random walk

$$X_t = X_{t-1} + u_t$$

is a difference-stationary time series. There is no need to prove that this series is non-stationary. Use the same assumptions about u_t.

(b) (7 marks) Consider the following non-stationary process:

$$y_t = \gamma_0 + \gamma_1 t + u_t, \qquad u_t = \rho u_{t-1} + \varepsilon_t, \tag{1}$$

where ε_t is i.i.d. $(0, \sigma^2)$.

- Explain the source or sources of non-stationarity of y_t. Indicate at what values of the parameters process (1) turns out to be difference stationary and at what values it is

trend stationary.

- Investigate the implications of detrending process (1) under the assumption $|\rho| < 1$.

- Investigate the consequences of applying the differencing transformation to process (1) under the assumption $\rho = 1$.

(c) (8 marks)

- Show that you can rewrite model (1) as

$$\Delta y_t = \beta_0 + \beta_1 t + \beta_2 y_{t-1} + \varepsilon_t. \tag{2}$$

Clearly indicate the one-to-one relation between $(\gamma_0, \gamma_1, \rho)$ and $(\beta_0, \beta_1, \beta_2)$.

- How can equation (2) be used to test time series (1) for stationarity?

QUESTION 16

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

The researcher investigates the effect of having vocational training available in high school on the probability of currently living in poverty for the population of men who grew up with a disadvantaged background. Let pov be a dummy variable equal to one if a man is currently living below the poverty line and zero otherwise. The variable age is age, and edu is total years of schooling. Let voc be an indicator equal to unity if a man's high school offered vocational training. Using a random sample of 850 men, the researcher obtains

$$\widehat{\Pr}(pov = 1 \mid age, edu, voc) = F(0.453 - 0.016age - 0.087edu - 0.149voc), \qquad (1)$$

where

$$F(z) = \frac{\exp(z)}{1 + \exp(z)}$$

is the logit function.

(a) (10 marks)

- Why is model (1) estimated by maximum likelihood and not OLS? Explain the meaning of the maximum likelihood method. What properties do estimates obtained by the maximum likelihood method have?

- Discuss the benefits and drawbacks of using the logit regression model when trying to explain a binary variable pov.

- Equation (1) contains information only about the estimated coefficients of the model. What additional information is needed to be able to judge the statistical quality of econometric model (1)? What tests can be carried out for this purpose?

(b) (7 marks)

- Use the direct comparison of two probabilities of living in poverty calculated by the logit function to evaluate the effect of having vocational training available in high school for a 40-year-old man with 12 years of education. Give details and interpret the results.

(c) (8 marks)

- Now do the same estimation of the marginal effect of vocational education as in (b) using derivatives.

- What percentage is the calculated marginal effect of the maximum possible?