Homework

- Deadline: 03.04.2020 at 23:59, Assignments are expected to be completed by due date. Assignments submitted after the due date will not be accepted.
- 20% of total grade
- Submit to : ketevani.kapanadze@cerge-ei.cz
- Email subject must be: Econometrics-HW
- Include your name, surname and UCO
- typed submissions must be in word/pdf format. If you decide to submit handwriting version be sure that the text is readable, and scanner has good quality.

Part 1: Multiple choice questions (20 points)

Circle the correct answer clearly.

1. In the simple linear regression model, the regression slope

a. indicates by how many percent Y increases, given a one percent increase in X.

b. when multiplied with the explanatory variable will give you the predicted Y.

c. indicates by how many units Y increases, given a one unit increase in X.

d. represents the elasticity of Y on X.

2. The OLS estimator is derived by

a. connecting the Y_i corresponding to the lowest X_i observation with the Y_i corresponding to the highest X_i observation.

b. making sure that the standard error of the regression equals the standard error of the slope estimator.

c. minimizing the sum of absolute residuals.

d. minimizing the sum of squared residuals.

3. Which of the following is true of the OLS t statistics?

a. The heteroskedasticity-robust t statistics are justified only if the sample size is large.

b. The heteroskedasticty-robust t statistics are justified only if the sample size is small.

c. The usual t statistics do not have exact t distributions if the sample size is large.

d. In the presence of homoskedasticity, the usual t statistics do not have exact t distributions if the sample size is small.

4. Consider the following simple regression model: $y = \beta_0 + \beta_1 x_1 + u$. Suppose z is an instrument for x. Which of the following conditions denotes instrument exogeneity?

a. $Cov(z,u) > 0$

b. $Cov(z,x) > 0$

c. $Cov(z,u) = 0$

d. $Cov(z,x) = 0$

5. Consider the equation, $Y = \beta_1 + \beta_2 X_2 + u$. A null hypothesis, $H_0: \beta_2 = 0$ states that:

a. X_2 has no effect on the expected value of β_2.

b. X_2 has no effect on the expected value of Y.

c. β_2 has no effect on the expected value of Y.

d. Y has no effect on the expected value of X_2.

6. The following simple model is used to determine the annual savings of an individual based on his annual income and education: Savings = $\beta_0 + \partial_0 * Edu + \beta_1 * Inc + u$

The variable 'Edu' takes a value of 1 if the person is educated and the variable 'Inc' measures the income of the individual. The benchmark group in this model is:

a. the group of educated people

b. the group of uneducated people

c. the group of individuals with a high income

d. the group of individuals with a low income

7. The significance level of a test is:

a. the probability of rejecting the null hypothesis when it is false.

b. one minus the probability of rejecting the null hypothesis when it is false.

c. the probability of rejecting the null hypothesis when it is true.

d. one minus the probability of rejecting the null hypothesis when it is true.

8. To decide whether or not the slope coefficient is large or small,

a. you should analyze the economic importance of a given increase in X.

b. the slope coefficient must be larger than one.

c. the slope coefficient must be statistically significant.

d. you should change the scale of the X variable if the coefficient appears to be too small.

9. If an independent variable in a multiple linear regression model is an exact linear combination of other independent variables, the model suffers from the problem of_____.

a. perfect collinearity

b. homoskedasticity

c. heteroskedasticty

d. omitted variable bias

10. Which of the following is an example of a binary response model?

a. MA model

b. ARCH model

c. GARCH model

d. Logit model

Part 2: True/false questions (20 points)

Indicate whether the statement below is true or false, no need for explanation.

1. Increasing the sample size can lead to a more precise estimate.

2. The key assumption for the general multiple regression model is that all factors in the unobserved error term be correlated with the explanatory variables.

3. A negative t-statistic indicates that the coefficient is not significant.

4. The dummy variable coefficient for a particular group represents the estimated difference in intercepts between that group and the base group.

5. The exclusion restriction for the IV requires that the instrument is significantly correlated with the endogenous variable.

Part 3: Conceptual Questions (20 points)

Answer in detail all question:

1. Explain what happens to estimated parameters if we include an irrelevant variable in the model.

2. Describe the Linear Probability Model and state its advantages and disadvantages compared to a probit/logit model.

3. Describe the concept and the use of *p-value*.

4. Describe when/why we use the instrumental variables and state the necessary conditions that valid instruments should satisfy.

Part 4: Solve the problem (40 points)

Earnings functions attempt to find the determinants of earnings, using both continuous and binary variables. One of the central questions analyzed in this relationship is the returns to education.

a) Collecting data from 253 individuals, you estimate the following relationship

$$\ln(Earn_i) = 0.54 + 0.083 * Educ, R^2 = 0.20$$
$$\quad\quad\quad (0.14)\ \ (0.011)$$

where *Earn* is average hourly earnings and *Educ* is years of education.

What is the effect of an additional year of schooling? If you had a strong belief that years of high school education were different from a college education, how would you modify the equation? What if your theory suggested that there was a "diploma effect"?

b) You read in the literature that there should also be returns to on-the-job training. To approximate on-the-job training, researchers often use the so-called Mincer or potential experience variable, which is defined as *Exper = Age – Educ – 6*. Explain the reasoning behind this approximation.

c) You incorporate the experience variable into your original regression

$$\ln(Earn_i) = -0.01 + 0.101 * Educ + 0.033 * Exper - 0.0005 * Exper^2, R^2 = 0.34$$
$$(0.16) \quad (0.012) \qquad\qquad (0.006) \qquad\qquad\qquad (0.0001)$$

What is the effect of an additional year of experience for a person who is 40 years old and had 12 years of education? What about for a person who is 60 years old with the same education background?

d) Test for the significance of each of the coefficients in the equation from part c). Why has the coefficient on education changed so little?

e) Suppose you expect that returns to education is different for man and woman. Explain carefully how you would adjust the model to check this hypothesis (in case you need to expand the model, clearly define variable that you wish to include).

f) Suppose your friend suggests that the variable *Educ* is correlated to the regression error (that is the variable is endogenous), therefore OLS estimates are not unbiased. Explain her how would you address this issue and ensure that the variable(s) you use satisfy desired conditions