BUEC 333-D200, Test 2

July 27, 2016, 14:30-17:20

1 Chapter 6

From Dougherty (2011, Chapter 6, p. 254-255). Using data on the educational attainment of $n = 1000$ individuals, their score on a knowledge test, and the educational attainment of their parents, we estimate the following regression:

$$\widehat{S} = 5.42 + 0.13ASV + 0.12SP$$
$$(0.01) \qquad (0.03)$$

where S is the individual's number of years of schooling, SP is the number of years of schooling of the parents, and ASV is the individual's score on a knowledge test (between 0 and 10).

1. Interpret the coefficient estimate of 0.12 on SP. **We estimate that the increase in the conditional expectation of your education level is higher by 0.12 years if, ceteris paribus, the education level of your parents goes up by 1 year.**

2. Is there sufficient statistical evidence that the education level of your parents has an effect on your education level? **Yes, the t-value is 4>2.58>1.96. We would reject the null of beta2=0.**

Next, we estimate the regression again, but now omitting ASV. We find that

$$\widehat{S} = 10.0 + 0.31SP.$$
$$(0.03)$$

3. Is there any evidence of omitted variable bias? Explain! **Yes, there is a big change in the coefficient on SP.**

4. Based on the results above, do you think that the relationship between SP and ASV is positive, negative, or zero? Explain, using the formula for omitted variable bias. **Use the OV formula. The bias is positive, so Cov(SP,ASV)>0.**

5. Explain in words, using your answer to (4), why the coefficient of SP changes from 0.31 to 0.12 when you include ASV. **People with more highly educated parents are more likely to score well on ASV, which measures underlying intelligence. That intelligence increase your expected education level. In the second regression, that effect is falsely attributed to your parents' education level.**

2 Chapter 5+8

Suppose that we have data on BUEC 333 test scores (Y_i), duration for which student i studies for the test (X_i), and the major of the student (D_i), where

$$D_i = \begin{cases} 1, \text{ if economics major} \\ 0, \text{ if non economics major.} \end{cases}$$

Consider the following model:

$$\ln(Y_i) = \beta_0 + \beta_1 X_i + \beta_2 D_i + u_i \tag{1}$$

where Assumption 1 holds:

$$E(u_i | X_i, D_i) = 0. \tag{2}$$

Y_i is the score between 10 and 100. X_i is the duration studied in hours, between 1 and 100.

1. What is the interpretation (in words) of β_1? **The expectation of your test score goes up by beta1*100% if you study for one additional hour, c.p.**

Now suppose that the OLS estimate for β_1 is 0.08, and its standard error is 0.04.

2. Test whether $\beta_1 = 0$, at the 5% significance level. **t=0.08/0.04=2>1.96. Reject.**

3. Do the same test, but at the 1% significance level. **t<2.58, do not reject.**

Finally, we try a different model:

$$Y_i = \beta_0 + \beta_1 \ln(X_i) + \beta_2 D_i + \beta_3 D_i \ln(X_i) + u_i \tag{3}$$

4. What is the interpretation of β_3 in this model? **This is about the unit change in the conditional expectation of test scores for a percentage increase in the hours studied, c.p. This change is higher, by 0.01*beta3, for econ majors than for non-econ majors.**

3 Chapter 9

Solutions in the book's Chapter 9.

1. What is the difference between internal and external validity?

2. Give an example of a study that is **not** externally valid. You can make one up. Explain!

4 Chapter 12

In a famous example of instrumental variables estimation, Acemoglu, Johnson, and Robinson (2001, AER) estimate the effect of institutions on economic growth. They use the following model:

$$GDP_i = \beta_0 + \beta_1 INST_i + u_i. \tag{4}$$

In this case, GDP_i is the gross domestic product of country i, and $INST_i$ is a measure of the quality of the institutions in that country. In particular, it may measure the strength of the legal system, and how well property rights are protected.

1. In this model, what does u_i capture? Give some real-world examples of quantities captured by u_i. **Other factors that influence a country's economic development. For example, the presence of natural resources, the health and education outcomes of its population, etc.**

2. The authors believe that u_i may be correlated with $INST_i$. Do you agree? Explain! **Yes. Excellent institutions would help with a country's health and education levels, for example.**

3. If u_i is correlated with $INST_i$, is the OLS estimator for β_1 unbiased? Explain! **It is biased, since Assumption 1 fails.**

The authors use an instrument $MORT_i$ that is defined as follows. First, they restrict the sample to former colonies, i.e. countries that were colonized by e.g. Portugal, the Netherlands, and Spain during the 15th through 18th century. For each of the remaining countries, they have a measurement for the proportion of settlers that died. For example, if Spain sent ships with 1000 men to Brazil, and only 700 of those are alive after 10 years, then $MORT_i = 0.3$ for Brazil.

4. What are the two conditions that this instrumental variable must satisfy? **It should be (i) correlated with INST, and (ii) uncorrelated with ui.**

5. In the model above, Y is GDP, X is $INST$, and Z is $MORT$. If it is given that

$$\begin{aligned} s_{XY} &= 1, \\ s_X^2 &= 2, \\ s_{ZX} &= -1, \\ s_{ZY} &= -1. \end{aligned}$$

   then what is $\hat{\beta}_1$, the OLS estimate? **beta1-hat = sxy / s2x = 0.5**

6. What is $\hat{\beta}_1^{TSLS}$? Why is it different from $\hat{\beta}_1$? **beta1-hat-2sls=szy/szx=1. It is different from the OLS estimate because Assumption 1 is violated, and therefore the OLS estimate is biased. If the conditions in (4) hold, then the TSLS estimator is consistent.**