Introduction to econometrics - 2019

Assignment

Hyeokmoon Kweon

Answer the questions below and give for each question a clear explanation and a brief discussion of the results. Also, insert Stata outputs and Stata codes for each question.

To solve each set of questions, you will use a different data set (`dta` file). The detailed description of the data is provided in the file. Make sure that you first check this information by using the `describe` command.

If you have a question, please post your question to "Discussions" menu of Cavnas so everyone else can also see it. Questions asked via email will not be answered.

Question 1: OLS

Use `highschoolgrades_HW.dta` to answer the questions below.

a. Consider the model below:

$$GPA_i = \beta_0 + \beta_1 \text{math\_test}_i + \beta_2 \text{lit\_test}_i + \beta_3 \text{gender}_i + \beta_4 \text{class-2}_i + \beta_5 \text{class-3}_i + \beta_6 \text{class-4}_i + u_i \quad (1)$$

where class-j is a dummy variable that indicates whether the student was in class j in high school.

(i) **(2 points)** Explain what the term u_i is. Why will different students have different values of u_i?

(ii) **(4 points)** Estimate the model (1) with OLS and exactly interpret your result for $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_4$.

b. Now estimate the model below:

$$\text{GPA}_i = \beta_0 + \beta_1 \text{average}_i + \beta_2 \text{gender}_i + \beta_3 \text{class-2}_i + \beta_4 \text{class-3}_i + \beta_5 \text{class-4}_i + u_i \quad (2)$$

where $\text{average}_i = (\text{math\_test}_i + \text{lit\_test}_i)/2$

(i) **(2 points)** What does the R^2 measure? What is the unit in which R^2 is measured?

(*ii*) **(5 points)** Compare the R^2's between model (1) and model (2). Clearly explain why R^2 changes in the way that you observe.

c. Standardize average_i and use it instead of average_i in the model (2). Estimate the model.

(*i*) **(3 points)** Interpret the estimated coefficient for average_i.

(*ii*) **(2 points)** What happened to R^2 now compared to the previous result from the model (2)? Explain why you observe such a result.

d. Estimate the model below:

$$log\ \text{GPA}_i = \beta_0 + \beta_1 log\ \text{average}_i + \beta_2 \text{gender}_i + \beta_3 \text{class-2}_i + \beta_4 \text{class-3}_i + \beta_5 \text{class-4}_i + u_i \tag{3}$$

where the only difference to model (2) is that we take log of GPA_i and average_i.

(*i*) **(2 points)** Interpret the 95% confidence interval for β_2.

(*ii*) **(4 points)** Interpret your result for $\hat{\beta}_1$ and $\hat{\beta}_2$.

e. **(4 points)** Estimate the model below:

$$GPA_i = \beta_0 + \beta_1 \text{math\_test}_i + \beta_2 \text{lit\_test}_i + \delta \text{math\_test}_i \times \text{lit\_test}_i + \beta_3 \text{gender}_i + \beta_4 \text{class-2}_i + \beta_5 \text{class-3}_i + \beta_6 \text{class-4}_i + u_i$$

where $\text{math\_test}_i \times \text{lit\_test}_i$ is the interaction term between math_test_i and lit_test_i.

Compute the marginal effect of the math test scores on the GPA for female students who were in Class 3 in high school with a literature test score equal to 100.

Question 2: IV estimation

Use `wage.dta` to answer the questions below.

a. We are interested in estimating the returns to education by using the presence of a university in one's county as an instrument variable (IV) for years of education.

(*i*) **(4 points)** Explain why this IV would be relevant (note that you are not being asked to perform any test).
*Bonus point (): explain also why this IV may not be exogenous.

b. Consider the model below:

$$log\ \text{wage}_i = \beta_0 + \beta_1 \text{educ}_i + \beta_2 \text{exp}_i + \beta_3 \text{black}_i + \beta_4 \text{north}_i + \beta_5 \text{south}_i + \beta_6 \text{married}_i + u_i \quad (4)$$

(i) **(2 points)** Estimate the model by OLS. Why can we not give a causal interpretation to the result for the returns to education from the OLS estimation?

(ii) **(4 points)** Assume that the presence of a university in one's county is a valid IV for years of education. Estimate the model by IV estimation (i.e., use uni_near_i as an IV for educ_i). What can you conclude from this?

(iii) **(3 points)** Write down and explain the second-stage equation for the IV estimation in (ii).

c. **(2 points)** Test if the given IV is relevant.

d. **(4 points)** Test the endogeneity of years of education variable using the two approaches introduced in the lecture.

Question 3: time series analysis

Use `ts.dta` to answer the questions below.

a. **(4 points)** The data contains quarterly time series of GDP growth. Determine the optimal lag length of the autoregressive (AR) model for the GDP growth series using the observations from 1960 and the BIC criteria (examine up to 12 lags).

b. **(4 points)** Given the lag length determined above, conduct an Augmented Dicky Fuller test to test if the GDP growth series has a unit root.

c. **(4 points)** Estimate your AR model from `a.` by OLS using the observations before 1990. Predict GDP growth from 1990 and compute the mean squared forecast error for this prediction. Plot the actual and predicted GDP growth in a line graph.

Question 4: static panel data analysis

Use `gun_state.dta` to answer the questions below.

a. We have a state-level panel data set of the U.S. from 1977 to 1999. Over this period, some U.S states enacted a law that that allows citizens to carry concealed guns. We will investigate the effect of the gun-carrying law on violent crime rates with the baseline model below:

$$log\ \text{crime}_i = \beta_0 + \beta_1 \text{carry}_i + \beta_2 \text{density}_i + \beta_3 \text{pop}_i + \beta_4 \text{avginc}_i + u_i \tag{5}$$

   (i) **(5 points)** Given the model above, estimate the effect of the gun-carrying law (*carry*) on the violent crime (*log_crime*) by OLS. Consider also a fixed effects model where state fixed effects are added to the above model. Estimate this model by within estimation. Compare the results and give an intuitive explanation of the fixed effects in this context.

   (ii) **(3 points)** Can you give a causal interpretation to the effect of the gun-carrying law estimated from the within estimation? If not, name a potential explanation that threatens causal inference (be specific).

b. **(4 points)** Demonstrate that least squares dummy variables estimator produces the same estimates as the within estimator (except for the constant).

c. **(5 points)** Test if the state-level fixed effects do exist. State null and alternative hypothesis and explain each step of the hypothesis testing in detail. Explain every necessary piece of information that you need to do this test (*e.g.*, test statistic, distribution, critical value).
   *Bonus point if you do this test without relying on `xtreg` command.

d. **(4 points)** Consider now a random effects model that accounts for state-level heterogeneity as random effects. Estimate the model by feasible generalized least squares. Formally test if the results from this estimation were consistently estimated. What can you conclude from this?

Submission deadline: September 20 Friday, 2019, 8.30 AM