ECON3360 Causal Inference for Microeconometrics

Tutorial 3: Stata application of IVs

Instructor: Julie Moschion

Problem I: the effects of institutions on economic development

Background The data set for this exercise comes from the paper by Daron Acemoglu, Simon Johnson and James Robinson "The Colonial Origins of Comparative Development: An Empirical Investigation", published in the American Economic Review, December 2001, vol. 91, p. 1369-1401. The paper can be downloaded from JSTOR using http://www.jstor.org/stable/2677930.

Idea They want to test the idea that 'institutions matter' for economic growth, in particular that countries with more secure protection of property rights have an advantage in encouraging capital accumulation. But a simple regression of GDP per capita on some measure of the current strength of property rights is vulnerable to the critique of omitted variables and reverse causality. They use an IV approach, instrumenting property rights with a measure of settler mortality from the 19th century. They argue that this is a good instrument because areas of the world in which settler mortality was high tended to introduce institutions designed to exploit the area's resources, while areas where the settler mortality was low instead tended to build sound institutions. And it is argued that these early institutions are correlated with current institutions.

Use the data in institution.dta for the following questions.

(1) Describe the variables in the data.

(2) How many countries are there in the sample?

(3) Draw two scatter plots: (i) between the measure of GDP (*lgdp*) and the level of protection of property rights (*prot*), and (ii) between *prot* and settler mortality (*logmort*).

(4) Regress *lgdp* on *logmort* (reduced form) and regress *prot* on *logmort*(first stage) (hint: use robust standard errors). Provide the IV estimate of the effect of property rights (*prot*) on GDP per capita (*lgdp*) using these two OLS estimates.

(5) Now use the 2SLS to obtain the IV estimate of the effect of the strength of property rights on GDP per capita.

(6) Let's calculate the IV estimate in two steps. First, regress *prot* on *logmort*. Generate the predicted value of the outcome and call this *prothat*. Second, regress *lgdp* on *prothat*. Show standard errors will be wrong.

Another researcher argues that settler mortality also had an effect on the proportion of current population who are of European descent and that the 'Neo-Europes' (those countries like the US, Canada etc that, through a combination of guns, germs and immigration, have a high proportion of their population who are of European descent) have better economic performance even given their institutions.

(7) If this argument is correct, what is the consequence for the consistency of the IV estimate of the effect of *prot* on *lgdp*?

(8) How would you deal with the problem?

(9) Implement your proposed solution and comment on the results.

Problem II: measurement error

The following exercise is based on generated data and the linear regression model presented below:

$$y = \beta_0 + \beta_1 x_1 + u_1 \tag{1}$$

where β_0=0 and β_1=1. We are interested in estimating $\hat{\beta}_1$.

(1) Set the number of observations equal to 500. Generate x_1 and u_1 from a standard normal distribution and generate y based on the regression model presented in (1).

(2) Suppose we do not observe x_1 and we only observe a variable with measurement error x such that $x = x_1 + v$ and where v is the measurement error. Generate v from a standard normal distribution and also generate x.

Now consider a regressor with measurement error (notice than we use x instead of x_1):

$$y = b_0 + b_1 x + u_2 \tag{2}$$

(3) Using OLS, estimate both equations (1) and (2) and compare the OLS estimators of β_1 and b_1 for the two regression models.

(4) Re-do instructions (1) to (3) while you set the number of observations equal to 1,000.

(5) Is there any major difference in the OLS estimate when the sample size increases? Does the bias in the OLS estimate due to measurement depend on sample size?

(6) Suppose the measurement error is larger now ($=2 * v$). Now re-define and re-generate $x = x_1 + 2 * v$. How does the OLS estimator of b_1 change when compared with the one constructed when the measurement error was equal to v.

(7) Suppose the measurement error is smaller now ($=v/2$). Now re-define and re-generate $x = x_1 + v/2$. How does the OLS estimator of b_1 change when compared with the one constructed when the measurement error was equal to v.