**Instrument Variable (IV) Approach**

# Case 1: 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. This 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 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 this is a good instrument, because parts of the world in which settler mortality was high tended to introduce institutions designed to exploit the area’s resources, while in areas where the settler mortality was low they tended not to do this and they 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 and , and (ii) between

and .

1. Regress on and regress on (hint: use robust standard errors). Provide the IV estimate of the effect of property rights () on GDP per capita () using these two OLS estimates.
2. We want to estimate the effect of the strength of property rights on GDP per capita. Implement the IV method by using 2SLS.
3. Now regress on (hint: use robust standard errors). Generate predicted value and call this . Now regress on . Show standard errors will be wrong.

Another researcher argues that settler mortality also had an effect on the population today who are of European descent and that the ‘Neo-Europe’s’ (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.

1. If this argument is correct, what is the consequence for the consistency of the IV estimate of the effect of on ?
2. How would you deal with the problem?
3. Implement your proposed solution and comment on the results.

# Case 2: IV example: The Effects of Smoking on Birth Weight

Use the data in bwght.dta for the following questions. This data include observations for pregnant women in US.

1. Describe the variables in the dataset.
2. Estimate the model by using OLS and interpret the estimates (hint: use robust standard errors).

(1)

We might worry that is correlated with other health factors or the availability of good parental care, so that and unobserved might be correlated. Now, we want to examine whether can be used as a potential instrumental variable for .

1. Suppose that we assume that and are uncorrelated. Is it a good assumption? Can you somehow test this? (hint: use robust standard errors)
2. Draw a scatter plot between on .
3. Perform a reduced form regression. Regress on (hint: use robust standard errors). Interpret the regression result.
4. Implement IV method by 2SLS.

# Case 3: Conditional IV example: The Effects of Education on Wage

Use the data in card.dta for the following questions. This data include men in US during 1976. Card (1995) used wage and education data to estimate the return to education.

1. Describe the variables in the data.
2. Estimate the model by using OLS and interpret the estimates (hint: use robust standard errors).

(2)

Education is likely to be endogenous. Card proposes that we use as an instrument for the endogenous variable .

1. Use a simple OLS regression to check whether *near*4 influences (hint: use robust standard errors). Could be correlated with factors in the error term, such as ?
2. Implement the IV method by 2SLS. Interpret the coefficient of . Do you find a different coefficient when you use robust standard errors compared to when you don’t use robust standard errors?
3. For a sub-sample of the men in the data set, an score is available. Do scores vary by whether a man grew up near a four-year college? What do you conclude from that?
4. Now use OLS to regress on , , and 1966 regional dummy variables ,...., (hint: use robust standard errors). Are and related after the geographic dummy variables have been con- trolled for?
5. What do you conclude about the importance of controlling for *smsa*66 and the 1966 regional dummies in the wage equation?
6. Now implement the IV method by 2SLS again while you control for the regional dummies (hint: use robust standard errors) and interpret the coefficient of *educ*.