

ECO 102: Topics in Economics

Ecole Polytechnique, Spring 2022

Professors: Geoffroy Barrows, Benoit Schmutz

Teaching Assistants: Arnault Chatelain, Maddalena Conte

TD4: Instrumental Variables

Today, we will follow Albouy (2012)'s critic of Acemoglu et al. (2001)' Instrumental Variable (IV). As usual, download the folder TD4.zip, type your answers directly in the Tex script and your Stata commands in a do file.

Exercise

1. A refresher on IVs:
 - (a) What is an Instrumental Variable? Why do we use them?
 - (b) In the case of Acemoglu et al. (2001), can you recall what they study, what variable they use as an IV and why?
 - (c) What are the two conditions that should be met when using an IV?
2. Go to Moodle and download the dataset `ajrcomment.dta`. Let us start with a short comment on the use of log in economics.
 - (a) In both papers the authors use log GDP per capita in 1995. Create a variable `gdp` that contains the GDP per capita in 1995. (We are referring to the natural logarithm here).
 - (b) Create a scatter plot of mortality rates (y-axis) against GDP per capita in 1995 (x-axis). Do the same with log mortality rate and log GDP. Compare the two. What do you observe?
 - (c) What is the benefit of using log values in tables?
3. Let us now reproduce the results of Acemoglu et al. (2001).
 - (a) Write down the two equations you will use in your 2 Stage Least Square (2SLS) regression (include a single control for Latitude).
 - (b) Run the first stage equation. Does the instrument seem valid?
 - (c) Using the command **predict**, generate a variable `riskhat` containing the fitted values of this first stage.
 - (d) Run the second stage regression. Comment on the results. Are the standard errors correct here?
 - (e) Run the IV regression using the **ivregress** command this time. Compare with your previous results.
4. We can now turn to Albouy (2012)'s critic. Albouy's first critic concerns Acemoglu et al. (2001)' standard errors.
 - (a) Albouy notices that our authors make conjectures about mortality rates for some countries. Notably, using mortality rate for some countries they extrapolate the mortality rate of neighbouring countries - in fact they even reuse the same rates sometimes. This violates one of the basic assumptions of OLS regression. Which one?

- (b) To make up for this, it is possible to cluster data, that is to consider that different groups of data (say continents) are independent but the observations composing them (countries here) are not. Albouy also suggest to run an regression robust to heteroskedasticity. Do you know what this is? Can you give an example of what it corresponds to?
 - (c) Rerun the first stage equation with standard errors (SE) that are robust (*i.e.* allowing for heteroskedasticity) and clustered at the mortality rate level. How does it change?
5. Albouy's second critic concern the validity of the data for some countries. Not only is the data for some countries simply extrapolated from neighbouring countries but also all data sources are not comparable, some rates concerning soldiers living in barracks, some concerning soldiers during campaign and some concerning forced labor.
- (a) Retaining only countries for which the mortality rate is not extrapolated and including dummies for data sources (campaign and slave variables) and continents rerun a 2SLS regression. Comment on the first stage. What value do you find for the expropriation risk coefficient?
 - (b) Compute the GDP ratio of Mexico on the US. Now using the previous coefficient what would be the new value of this ratio were Mexico to have the same property right as the US? Comment on the result.

Bibliography

- Acemoglu, D., Johnson, S., and Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American economic review*, 91(5):1369–1401.
- Albouy, D. Y. (2012). The colonial origins of comparative development: an empirical investigation: comment. *American economic review*, 102(6):3059–76.