Problem Set 3

Your id here

2025-02-24

Before attempting to answer this problem set, make sure to read carefully this guide by Felton and Stewart (2024).

1 Instrumental variable in theory

For this section, I ask you to explain the instrumental variable setting based on the modern framework, also known as LATE (Local Average Treatment Effect) or CACE (Compliers Average Causal Effect).

1.1 Write down and explain the instrumental variable assumptions in graphical terms.

This is, show the "canonical" graph where there is a valid instrumental variable, and describe the graphical conditions to verify that the candidate instrument is indeed valid.

Would these conditions give you point identification? Why?

1.2 Write down and explain the instrumental variable assumptions using potential outcomes.

Now, describe and explain the assumptions required for a valid instrumental variable, using potential outcomes. Be explicit about any extra assumption you are making (such as binary vs continuous instruments, or functional form assumptions)

Would these conditions give you point identification? Why?

1.3 First-stage

Of the assumptions above, explain what assumptions are needed to identify the "first-stage" (i.e., the effect of the instrument on the treatment)

1.4 Intent-to-treat

Of the assumptions above, explain what assumptions are needed to identify the "intent-to-treat" (i.e., the effect of the instrument on the outcome)

2 IV in practice (I)

(Adapted from Chad Hazlett)

In a well-known paper, Acemoglu, Johnson, and Robinson (2010) attempt to estimate the effect of property rights institutions on long-run economic growth. They use GDP per capita in 1995 to proxy a country's long-run economic growth. To address the endogeneity between institutions and economic growth, they use the mortality rate faced by European settlers in the 19th century as an instrument for property rights institutions in the 20th century. This problem asks you to replicate the results from 2SLS regressions in their original paper (Table 4 Columns 1 and 2). Please download `ARJ2001.txt` from Canvas. The data file include the following variables:

- `shortnam`: three letter country name
- `lat_abst`: absolute latitude of capital/90
- `avexpr`: average protection against expropriation risks
- `logpgp95`: log PPP GDP per capita in 1995
- `logem4`: log settler mortality

a) Show results from (1) the first stage regression, (2) the reduced form regression, and (3) the 2SLS estimation using the following two specifications:

$$\begin{aligned} \texttt{logpgp95}_i &= \beta_0 + \beta_1 \texttt{avexpr}_i + \epsilon_i \\ \texttt{avexpr}_i &= \gamma_0 + \gamma_1 \texttt{logem4} + \mu_i \end{aligned}$$

and

$$\begin{aligned} \texttt{logpgp95}_i &= \beta_0 + \beta_1 \texttt{avexpr}_i + \beta_3 \texttt{lat\_abst} + \epsilon_i \\ \texttt{avexpr}_i &= \gamma_0 + \gamma_1 \texttt{logem4} + + \gamma_2 \texttt{lat\_abst} + \mu_i \end{aligned}$$

Report your results in the following table (you can use canned functions, such as `ivreg`):

		(1) no covariates	(2) including latitude
First stage (dep: `avexpr`):	`logem4`	xx	xx
		(xx)	(xx)
	`lat_abst`		xx
			(xx)
Reduced form (dep: `logpgp95`):	`logem4`		
		()	()
	`lat_abst`		
			()
2SLS (dep: `logpgp95`):	`avexpr`		
		()	()
	`lat_abst`		
			()

b) Regress `logpgp95`, `avexpr`, and `logem4` on `lat_abst` ("partialling out" the effect of latitude) and re-do the 2SLS estimation using the residuals. Do you get the same result as in Column 2 in the previous question? (Don't worry about the standard errors – actually they are quite close to the right ones.)

c) For the model with no covariates, discuss the identification strategy, and whether you believe it or not. Is settler mortality a good instrument? Walk through the three identification assumptions we talked about in class and address each in the context of this paper.

3 IV in practice (II)

For this section, you need to read Doyle et al., 2015.

3.1 Draw the DAG implied by the strategy described in section III.B.

(You can use online tools such as Dagitty or excalidraw)

Justify included and omitted variables and arrows, based on the reasoning provided by the authors.

3.2 Use the DAG created above to discuss the IV assumptions

List and explain the necessary assumptions for the IV strategy to work *in this particular application*. Discuss how plausible these assumptions are *in this particular application*.

3.3 Are you generally convinced by the results? Explain

4 IV in practice (III): Extra credit

For this section, you need to replicate all the points in the previous question, but based on an article of your own choosing.

The criteria for eligible articles are the following:

- Published between 2020-2025
- Published in the *American Journal of Sociology*, *American Sociological Review*, or *European Sociological Review*
- Uses an instrumental variable analysis as the *main* identification strategy (that's it, not only as a robustness check)