

# Exercise IV1 (simple world of homogeneous treatment effects).

## Problem A

In the lecture we used the TSLS approach to motivate IV estimation. I then told you that the Wald estimator is just a simple version of IV. In this exercise, I ask you to derive the Wald-estimator directly using the potential outcomes framework. Here are some guidance:

This is the true model:

$$y_i = \alpha + \beta S_i + \varepsilon_i \quad (1)$$

You are interested in estimating  $\beta$  but you think that your conditional independence assumption might be violated.

- Using the potential outcomes framework (assuming homogeneous treatment effects), write down the expression for the ATE. Based on the resulting equation, which terms does your worry that the CIA is violated refer to?

Since knowing  $\beta$  is important, you convince a wealthy donor (who wants fame attached to her name) to fund a RCT in which  $Z$  is randomly assigned. This delivers you an treatment/instrument  $Z$  which only takes values 0 and 1 and which is correlated with  $S$ . This is the effects of  $Z$  on  $S$ :

$$E[S_i | Z_i = 1] - E[S_i | Z_i = 0] \quad (2)$$

- How is this effect called in the IV-language? Do you think you can get an unbiased estimate of this using OLS?

You notice that you can also estimate the difference in  $Y$  depending on  $Z$  directly:

$$E[y_i | Z_i = 1] - E[y_i | Z_i = 0] \quad (3)$$

- Estimating this using OLS, is this biased, and why? How is this effect called? Is this effect of interest?

Your donor reminds you that what she really wants you to find out is  $\beta$ . Plug equation (1) into (3). Hint: it might help to first do this for  $E[y_i | Z_i = 1]$  and then for  $E[y_i | Z_i = 0]$ , before taking differences. When you solve for  $\beta$  you obtain the Wald estimator of  $\beta$ . This is simply the IV estimator with a dummy instrument.

- Look at your Wald-estimator exactly. Using the IV-terminology introduced in the lecture, what are the names of the effects just looking at the numerator, or just looking at the denominator?

## Problem D

We now examine a situation with multiple instruments and TSLS-estimation in Stata. Imagine you are interested in causal determinants of rental prices, and in particular to what extent these depend on housing values. For this, you collected some data (just type `webuse hsng2`, `clear` in Stata to run the following regression:

```
ivregress 2sls rent pcturban (hsngval = faminc i.region),  
vce(robust)
```

- What are the endogenous and what are the exogenous variables? What is your assessment of the quality of the instruments?

Since you are unsure about your instruments, you now use an overidentification test based on Wooldridge's score test of overidentifying restrictions (which is based on Sargan's test for overidentification). You probably came across these tests in past Econometrics courses.

So, to implement this test with the H0 that your instruments are valid use the post-estimation command `"estat overid"` in Stata.

- Can we reject this H0 at the 5 percent level? What does this mean for your set of instruments and for the use of overidentification tests more generally?

## Examples

Read over the included papers by Angrist (veteran draft lottery), Angrist and Krueger (schooling) and Angrist and Evans (cost of children). Write out the DAGs for each paper with clear labels including the instrument they propose. In each case, how can they argue that the instrument is not a parent to the regressor of interest and the error term of the equation of interest?