

UNIVERSITY OF BAYREUTH

Causal Inference

Assignment 3

Due noon on 18 July 2022

Word limit: 4000 words

Exercise 1: Institutions and Economic Development

The assignment is based on the famous Acemoglu, Johnson & Robinson (AJR) 2001 study on the importance of inclusive institutions on economic development.¹ AJR argue that institutions leave a long imprint on countries' economic activity. They distinguish between inclusive and extractive institutions. The former serve to diffuse economic returns along different strata, whereas the latter serve to appropriate wealth among few.

To find evidence for the importance of the distinction between good and bad institutions, they turn to the colonial structures of the 19th and early 20th century. Their identification strategy comes from the quasi-random variation in geography which determined whether colonizers would establish inclusive or extractive institutions. In those areas in which settlers encountered high mortality rates they built extractive institutions, without long-term planning. In areas with low mortality rates, they built inclusive institutions. To the extent that the geographic determinants of mortality rates are as good as randomly assigned, they provide a very important instrument in order to test the economic returns of institutions.

The data can be found on the website as `AJR.dta`. The key variables are the following:

- `logpgp95`: the logged GDP per capita measured in 1995. (Logging GDP is a very typical way to bring its distribution closer to the normal, while changing the interpretation from levels to

¹ Acemoglu, Johnson, & Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review*, 91(5): 1369-1401, available [here](#).

percentages – if you have more questions about what logarithmic transformations do and how they are interpreted, please ask me). This is Y , the dependent variable of interest.

- **avexpr**: Average protection against expropriation risk (1985—1995). This is an index of how extractive a given set of institutions is. This is our D , which is of course not randomly assigned.
- **logem4**: Logged settler mortality rates. This is Z , our instrument.

You can also use other variables from the dataset if you think this helps in any part of your analysis.

Your tasks are to answer the following questions:

- Is there a link between institutions and economic development? This is not a causal question, we are asking is if there is any association between the two. Provide a scatterplot to show this is the case.
- Is this relationship causal? How do mortality rates help in answering this question?
- Provide a graph for the reduced form equation. Estimate the ITT.
- Is the first-stage assumption valid in this example?
- Estimate the LATE, using both a Wald estimator and a 2SLS estimator. Does conditioning on observables change the estimates?
- After presenting their main results, the authors try to address criticisms about the validity of the instrument (practically about exclusion). Read this section of the paper (available [here](#)). Now evaluate the exclusion restriction: Do you conclude the authors' IV design is valid? Why or why not?

Exercise 2: Incumbency Advantage

The assignment is based on Broockman’s 2009 study on the presence of “reverse coattails”.² Broockman examines whether the outcome of the Congressional race in a district affects the outcome of the next Presidential race in the same district. The data he uses are available on the website. The data are sorted by state and election year (`statesabbrev` & `t2year` respectively). No district identifier is used and you will not need one for your analysis. The variables of interest are the following:

- `dv_c_t1`: Democratic Congressional Vote share at Time $T = 1$ (margin of victory).
- `dv_c_t2`: Democratic Congressional Vote share at Time $T = 2$ (margin of victory).
- `dv_p_t2`: Democratic Presidential Vote share at Time $T = 2$ (margin of victory).
- `t2_ismidterm`: a dummy that switches on for midterm election years.

You can also use other variables from the dataset if you think this helps in any part of your analysis. Remember, the unit of analysis (the row in your dataset) is a congressional district in an election year. Note that the main goal here is *not* to replicate Broockman’s study.

Your aim is to find out if there is an incumbency advantage in US Congressional elections. In other words, does the outcome of a Congressional election at time $T = 1$ affect the outcome of a Congressional election at time $T = 2$? [Note that this question has nothing to do with Broockman’s topic, which is the relationship between Congressional and Presidential election outcomes.] To address this question, take the following steps:

1. Using the `rdplot` command in the `rdrobust` package, produce an RD plot showing the relationship between Congressional outcomes at $T = 2$ and Congressional outcomes at $T = 1$.
2. Restrict attention to races decided by 5 percentage points or less and estimate the incumbency effect using a linear regression. Report the estimated effect.
3. Report the estimated incumbency effect produced using `rdrobust`.
4. Show how the estimated incumbency effect varies with the bandwidth you choose. Visually present the estimate and confidence interval for at least 5 bandwidths. (The bandwidth

² Broockman, David E. 2009. “Do Congressional Candidates Have Reverse Coattails? Evidence from a Regression Discontinuity Design,” *Political Analysis*, 17(4):418-34, available [here](#).

should be shown on the horizontal axis; the point estimate and confidence interval should be shown on the vertical axis.)

5. Still using `rdrobust`, conduct a falsification test (placebo test): does the outcome at time $T = 2$ affect the outcome at time $T = 1$?
6. Sorting can be a serious problem in RD designs. Show both visually and statistically that sorting does not invalidate the results of this study.