

## Topics in Microeconomics @ ITAM, Spring 2022

### Exercise 5: Political Selection and Political Resource Curse

*This exercise is based on two articles: “Political Selection and Economic Policy” (Meriläinen 2022) and “The Political Resource Curse” (Brollo et al. 2013). Please submit your responses on Canvas by May 31<sup>st</sup> (by midnight). Note that you will earn full credit even if you do not get everything 100% correct or have time to answer everything; the most important thing is that you participate in the class and get some practical exercise in data work.*

*You are encouraged to work on this exercise in small groups as we have done in the class, although you can also work on your own if you prefer. Feel free to use the Canvas group to find a “study group” for the exercise.*

---

#### 1. Political Selection and Economic Policy

There are diverse difficulties in studying the effects that political selection has on policy outcomes. In the first part of this exercise, you will discuss these problems and potential solutions.

(a) Why is it not obvious how politician quality would map into policy outcomes?

(b) Researchers studying this kind of questions often face data limitations also in terms of what kind of dependent variables they can use. The outcomes that the author is interested in are two types of economic policies: public spending and fiscal sustainability outcomes. How does the author motivate the focus on these outcomes? Can you think of any other outcomes that would be interesting to analyze?

(c) Suppose that a researcher would just regress a policy outcome on the quality of politicians. What kind of problems would this kind of regression involve?

To deal with the endogeneity issues, the paper uses an identification strategy like the close elections regression discontinuity designs that we have already seen in this class. Why is it difficult to conduct a conventional RDD in the context of the study? Briefly summarize the idea of the identification strategy that the paper uses.

(d) The measurement of politician quality is also tricky! The second paper that was assigned as a reading for this class uses university education as one metric of politician quality. Based on your reading of the paper “Political Selection and Economic Policy”, why might education be an imperfect indicator of politician quality? What alternative measures does the paper use? What are the advantages and disadvantages of these? If there were no data limitations, what kind of measure do you think would be an ideal measure of politician quality and why?

(e) Now that we have discussed all sorts of problems, it is probably a good time to get to the business and summarize the main findings of the paper. How do different politician qualities affect economic policies? What kind of explanation does the author offer for the fact that different qualities may matter for policy in different ways?

## **2. Setting up**

We will now move on to do some data work on Stata. Create a folder that has subfolders for data, code, and results. Save the files called “AER\_smallsample.dta” and “AER\_largesample.dta” in your data folder. Start a .do file and save it in your code folder. Remember to keep saving your code frequently throughout the exercise. You should submit this .do file on Canvas by May 31<sup>st</sup>. Indicate the names of the students in your group (if you worked in a group), the number of this exercise, and today’s date on top of the file. Finally, set globals that tell Stata where your data are and where you will be storing your results. Remember to comment your code throughout the exercise. You will use this .do file to answer the questions below, and to produce the requested analyses.

## **3. The Political Resource Curse**

(a) The analyses of the paper are motivated by a theoretical model that delivers six empirical predictions. What are these predictions? Briefly explain how the paper tests each of these predictions and what are the main findings.

Note that the authors use a *fuzzy* regression discontinuity design. How is this different from a *sharp* regression discontinuity design? Why do they authors have to use this approach?

(b) We will now open the data file called “AER\_smallsample.dta” and conduct some of the analyses in the paper. Let us first produce three different versions of the first panel in Figure 2. The outcome variable is called *broad* and the running variable is called *popnorm*.

For each of the graphs, specify graph options that make the figure look nice. In particular, make sure that the axes are labeled. After plotting the graphs, save it both as a gph and a pdf file.

After producing the graphs, briefly discuss which one you prefer and why.

- (i) First, we will produce something that somewhat resembles Figure 2. Install the *binscatter* command by typing *ssc install binscatter, replace*. Have a look at the help file to familiarize yourself with the command. After you know how it works, plot the RDD graph using a linear control polynomial. Restrict the sample such that  $abs(popnorm) \leq 3500$ .
- (ii) Second, plot a similar graph using the *rdplot* command that we have used in earlier exercises. Now, use a second-order control polynomial.
- (iii) Third, we will use a “brute force” approach (using a third-order control polynomial) which will allow us to also plot the 95% confidence intervals.

First, generate binned averages:

```
gen group=-3500 if popnorm>=-3500&popnorm<=3250
forvalues i=1(1)30 {
    replace group=-3500+(i*250) if popnorm>=-3500+(i*250) & popnorm<=
    3250+(i*250)
}
egen popnorm_mean2=mean(popnorm), by(group)
egen broad_mean2=mean(broad), by(group)
egen narrow_mean2=mean(narrow), by(group)
```

```
egen fraction_broad_mean2=mean(fraction_broad), by(group)
egen fraction_narrow_mean2=mean(fraction_narrow), by(group)
```

Then, run the regression and construct predictions of the regression fit and confidence intervals:

```
reg broad treatnorm popnorm popnorm_2 popnorm_3, cluster(id_city)
predict yhat_1
predict SE_1, stdp
gen low_1 = yhat_1 - 1.96*(SE_1)
gen high_1 = yhat_1 + 1.96*(SE_1)
```

Last, plot the graph:

```
twoway (scatter broad_mean2 popnorm_mean2 if popnorm<3392) (line yhat_1
low_1 high_1 popnorm if popnorm<0, pstyle(p p3 p3) sort) (line yhat_1 low_1
high_1 popnorm if popnorm>0&popnorm<3392, pstyle(p p3 p3) sort),
xtitle(Population) ytitle(Broad corruption) legend(off) xline(0)
```

(c) Your next task is to replicate part of the results in Table 9. We will focus on the specification that pools together all thresholds.

- (i) Open the data file called “AER\_largesample.dta” and run some regressions following what is done in the paper. The specification that the authors present controls for *pop*, *pop\_2*, *pop\_3*, *i.term*, and *i.regions*. They instrument *fpm* with *fpm\_hat*. Write a loop that runs this kind of analysis using the variables *opp\_college*, *opp\_yschool*, and *reele\_inc* as dependent variables. You can use the *ivreg2* command. Lastly, the authors use standard errors clustered by city (*id\_city*).
- (ii) Discuss the estimation results. How do the regression analyses here differ from what the authors visualize in Figure 3? Can you think what kind of specification could be a more credible one and why (no need to run any regressions, though)?

- (iii) What happens if you were to run the regressions without any control variables? Do you think it is necessary to include the covariates? Why/why not?
- (d) Plot the distribution of the running variable *plotnorm* as a histogram and save it. What can this kind of figure tell us about the validity of the RDD? What other validity checks should one perform to fully reassure you that the results are solid?
- (e) The paper uses population thresholds in a regression discontinuity design setting which is also something many other papers do. If you have some time, please answer the following questions about population threshold based RDDs (having a look at the extra readings might help with this):
  - (i) What potential problems could there be with studies that exploit this kind of RDD set-ups? How have researchers dealt with this?
  - (ii) One difficulty (but not one that one could not overcome!) is that there are multiple population thresholds. There are basically two ways researchers can deal with this--what are these?
  - (iii) Can you think of any population-based policies that would apply in the Mexican context?

#### **4. Celebrate!**

This is the last problem set for this course. You have already made it very far, and you are almost done with the class. Congratulations! Grab a beverage of your choice and celebrate your accomplishment.