PLSC 30600: Problem Set 3

[YOUR NAME]

February 16, 2023

This problem set is due at **11:59 pm on Tuesday, February 28th**.

Please upload your solutions as a .pdf file saved as "Yourlastname_Yourfirstinitial_pset3.pdf". In addition, an electronic copy of your .Rmd file (saved as "Yourlastname_Yourfirstinitial_pset3.Rmd") must be submitted to the course website at the same time. We should be able to run your code without error messages. In addition to your solutions, please submit an annotated version of this `.rmd` file saved as "Yourlastname_Yourfirstinitial_pset3_feedback.rmd" and a corresponding PDF saved as "Yourlastname_Yourfirstinitial_pset3_feedback.pdf" noting the problems where you needed to consult the solutions and why along with any remaining questions or concerns about the material. In order to receive credit, homework submissions must be substantially started and all work must be shown. Late assignments will not be accepted. In total your submissions should consist of four files.

Problem 1

This problem will have you replicate and analyze the results from Moser and Voena's 2012 AER paper on the impact of the World War I "Trading with the Enemy Act" on U.S. domestic invention. The full citation is below

> Moser, P., & Voena, A. (2012). Compulsory licensing: Evidence from the trading with the enemy act. American Economic Review, 102(1), 396-427.

The premise of the study is to evaluate the effect that "compulsory licensing" policy – that is, policies that permit domestic firms to violate foreign patents and produce foreign inventions without needing to obtain a license from the owner of the foreign patent – have on domestic invention. Does access to foreign inventions make domestic firms more innovative? The authors leverage an exogenous event in U.S. licensing policy that arose from World War I – the 1917 "Trading with the Enemy Act" (TWEA) which permitted U.S. firms to violate patents owned by enemy-country firms. This had the consequence of effectively licensing all patents from German-owned firms to U.S. firms after 1918 (that is, from 1919 onward), allowing them to produce these inventions without paying for a license from the German-owned company.

The authors look specifically at domestic innovation and patent activity in the organic chemicals sector. They note that only some of the sub-classes of organic chemicals (as defined by the US Patent Office) received any compulsory licenses under the Trading with the Enemy Act while others did not. They leverage this variation in exposure to the "treatment" of compulsory licensing to implement a differences-in-differences design looking at domestic firm patent activity in each of these sub-classes (comparing sub-classes that were exposed to compulsory licensing to those that were unexposed).

The dataset is `chem_patents_maindataset.dta` – the code below will load it.

```
library(tidyverse)
# Read in the Moser and Voena (2012) dataset
chem <- haven::read_dta("chem_patents_maindataset.dta")
```

The unit of the dataset is the sub-class/year (471,120 observations) of 7248 US Patent and Trademark Office (USPTO) patent sub-classes over 65 years.

The relevant variables are

- `uspto_class` - USPTO Patent Sub-Class (unit)
- `grntyr` - Year of observation (year)
- `count_usa` - Count of patents granted to US-owned firms in the year
- `count_france` - Count of patents granted to French-owned firms in the year
- `count_for` - Count of patents granted to foreign-owned (non-US) firms in the year
- `treat` - Treatment indicator – Whether the patent sub-class received any German patents under TWEA (after 1918 when the policy went into effect) (Note that this is not an indicator for the overall treatment *group* (whether the unit *ever* received treatment) – it is only 1 after 1918 for units that receive treatment but is still 0 for those "treated" units prior to the initiation of treatment)

Part A

If you try to use a two-way fixed effects estimator on the dataset as it is, it will likely freeze up your computer as this is a *very large* dataset. We'll instead first aggregate the data in a way that will let you use a simple difference-in-differences estimator to estimate the treatment effect.

Generate a point estimate for the average treatment effect of receiving treatment on the average annual count of US patents using a difference-in-differences estimator (using all post-treatment (1919-1939) and pre-treatment (1875-1918) time periods. You should aggregate your data such that the outcome is the post-/pre-difference in the outcome (preferably using `tidyverse` functions like `group_by` and `summarize`) and each row is a USPTO patent sub-class (rather than a sub-class/year observation) and use a difference-in-means estimator with the differenced outcome. Again, if you use `lm_robust` or even `lm` with two-way fixed effects, your computer will likely freeze up as there are many FE parameters to estimate.

Provide a 95% robust confidence interval and interpret your point estimate. Do we reject the null of no treatment effect at the $\alpha = .05$ level?

Part B

A colleague suggests that you should instead just compare the average differences in the count of US patents in the post-1918 period between exposed and unexposed sub-classes to estimate the treatment effect. Based on what we observe in the pre-1919 period, is ignorability of the treatment likely to hold under this strategy? Discuss why or why not – what do you observe in the patent counts in the pre-treatment period between exposed and unexposed subclasses.

Part C

The authors implement a test of their identification assumptions by also estimating the effect (using the differences-in-differences design) of the Trading with the Enemy Act on patents granted by French firms, which the authors note "could not license enemy patents under the TWEA." Describe what sort of a diagnostic strategy this is. What do the authors expect to find if their parallel trends assumption holds?

Estimate the effect of TWEA exposure on the count of French firm patents using a difference-in-differences design and provide a a 95% robust confidence interval. Are the results consistent with what the authors expect if their design assumptions hold?

Part D

We might be concerned that there are differential trends in pre-treatment patenting between those sub-classes exposed to the treatment and those exposed to control. Estimate the difference in the trend in US patents between exposed and unexposed sub-classes from 1918 to 1917, 1916, 1915, and 1914 (four estimates in total: 1918-1917, 1918-1916, 1918-1915, 1918-1914). Provide a 95% robust confidence interval for each of these

estimates and interpret your results. Do we reject the null that any of these differ from 0 (at $\alpha = .05$)? If the outcome trends were evolving in parallel between the treated and control groups, what would we expect these estimates to be? What do your results suggest for the validity of the parallel trends assumption?

Part E

The authors adjust for covariates out of concern for possible parallel trends violations. One possible confounder that might be driving a parallel trends violation is the overall amount of foreign patenting in the sub-class and its change over time – reflecting general technological differences that might differ between the patent sub-classes. Since the treatment does not affect the amount of foreign patenting, this is a valid control.

Create a variable for the change between the post- and pre-treatment count of foreign patents in the USPTO subclass. Use the Abadie (2005) weighting estimator to estimate the ATT under a conditional parallel trends assumption (fit a model for the propensity score using this covariate). Do we reject the null of no treatment effect at the $\alpha = .05$ level? Compare your results to your estimate from Question A and discuss why they might differ.

Problem 2

Consider the standard instrumental variables set-up in the treatment non-compliance setting, Z_i denotes the assignment to treatment, D_i is the actual treatment that is taken and Y_i is the outcome of interest. Assume that all of the instrumental variables assumptions hold (treatment is ignorable, exclusion restriction, non-zero first stage, monotonicity/no-defiers). The specific example you'll examine in this problem is the JOBS II randomized trial which evaluated the effect of a job training program on reducing depression among those who recently experienced job loss (among other health outcomes). This specific dataset on 502 "high-risk" individuals from this experiment comes from:

> Little, Roderick J., and Linda HY Yau. "Statistical techniques for analyzing data from prevention trials: Treatment of no-shows using Rubin's causal model." Psychological Methods 3.2 (1998): 147.

The code below will load the data into R

```
## Load the JOBS II dataset
jobs2 <- read_table2("wjobs.tab", na= ".")
## Make the treatment and instrument variable
jobs2$Z <- jobs2$Tx
jobs2$D <- as.integer(jobs2$Tx == 1 & jobs2$c1 == 1)
```

The relevant columns are

- `Z` - Assignment to the job training program
- `D` - Actual participation in job training
- `depress` - Change in depression from baseline (higher values = more depression)
- `risk` - Pre-treatment mental health risk score
- `educ` - Number of years of education completed
- `age` - Age
- `single` - Marital status: single

Part A

Examine the data and explain why monotonicity **is guaranteed** to hold for this particular design.

Part B

Using the Wald estimator, estimate the local average treatment effect of participation in the job training program among the compliers. Assuming asymptotic normality, provide a 95% confidence interval and

conduct a hypothesis test for the null of no LATE at $\alpha = .05$.

Part C

Estimate the first stage effect of assignment to training on participation in the program. Would you consider assignment to the job program tobe a strong instrument for participation?

Part D

Which subset of your sample consists exclusively of never-takers? Suppose you were to take the mean of one of your covariates among this group – explain why this would identify the mean of the covariate among the never-takers.

Part E

Using the group you identified in D, examine the pre-treatment covariate averages among the never-treated respondents. Compare these to the overall means in the sample. Discuss whether there are any notable differences and what this would imply for the representativeness of the LATE you identified in Part B.

Problem 3

In this problem you will analyze the data from David Broockman's paper examining whether congressional incumbents help boost the vote shares for their party's presidential candidate in their respective congressional districts. The citation is here

> Broockman, David E. (2009) "Do Congressional Candidates Have Reverse Coattails? Evidence from a Regression Discontinuity Design," Political Analysis 17(4): 18–434.

Below is the code to import the dataset into R

```
### Load in the Broockman (2009) dataset
congress <- read_csv("congress.csv")
```

This paper is interested in examining whether there exist "reverse coattails" in US elections. The "coattail" effect refers to the observation that popular presidential candidates tend to boost the performance of their party's candidates in down-ballot races (e.g. U.S. House and Senate races). This paper investigates whether the *reverse* also might exist – whether congressional incumbents are able to boost the performance of co-partisan Presidential candidates in their districts.

The study analyzes the effect of congressional incumbency on district Presidential vote share using a Regression Discontinuity Design. The dataset consists of 8994 observations of congressional districts over time. For each observation, we observe the Democratic party congressional candidate's vote share margin of victory in the election at time t along with the Democratic party congressional candidate's vote share margin of victory in time $t + 1$ and the Democratic party presidential candidate's vote share margin of victory in time $t + 1$.

Candidates who have a margin of victory greater than 0 in time t go on to be the incumbent in time $t + 1$, so the discontinuity in treatment assignment (incumbency) is at 0.

- `dv_c_t1` - Democratic candidate *congressional* vote-share margin of victory in time t
- `dv_c_t2` - Democratic candidate *congressional* vote-share margin of victory in time $t + 1$
- `dv_p_t2` - Democratic candidate *presidential* vote-share margin of victory in time $t + 1$
- `t2_is_midterm` - Whether the election at time $t + 1$ is a midterm election (no Presidential election)

Part A

In words, what are the main assumptions that we would need to make in order to identify the causal effect of Democratic congressional incumbency in time $t + 1$ on Democratic presidential vote share in time $t + 1$ using

a regression discontinuity design?

Part B

The study first replicates the incumbency findings in Lee (2008). Considering only years with Presidential elections (no midterm elections), estimate the effect of democratic incumbency on democratic party *congressional* vote share margin of victory in time $t + 1$ using a local-linear RD estimator with a bandwidth of .25. Provide a 95% confidence interval and interpret your estimate.

Part C

Create a binned scatterplot for your analysis in Question 2 and overlay the local linear regressions above and below the discontinuity. Does the linear model seem like a reasonable approximation for the conditional expectation function above and below the discontinuity for the original bandwidth choice (.25)?

Part D

Now estimate the effect of democratic congressional incumbency on democratic *presidential* candidate vote share margin of victory. Again, use a local linear regression with a bandwidth of .25. Provide a 95% confidence interval for your estimate and interpret your results. Provide a binned scatterplot with the local linear regressions overlaid on top and discuss whether the local linear models are a reasonable approximation to the conditional expectation functions.

Part E

Consider choosing a number of different bandwidths for estimating the RD effect on *presidential* vote share using a local linear regression. Estimate the effect using three different bandwidth choices: .05, .1, and .4. For each estimate, provide a 95% confidence interval and interpret your results. Discuss the extent to which the results differ (both the point estimates and standard errors).

Given your analyses, would you conclude that there is evidence of a "reverse coattails" effect?