Hertie School
The University of Governance in Berlin

Prof. Dr. Simon Munzert
Dr. Lukas F. Stoetzer

GRAD-C6-2001: Statistics II

Replication Project

Instructions

Spring Semester 20

General information

In the replication project, you are supposed to replicate the main findings of a recently published study. In doing so, you will apply and extend your statistical and coding skills learned in class and in the lab. The papers we chose are all from different research journals, covering different areas.

We will grade your submission based on (1) the completeness and correctness of the presented results, (2) the quality of your R code, which should be commented well and formatted such that any reader can understand what you were doing, and (3) the quality of your extension of the replication project. Regarding the last criterion, you are asked to use the available data to explore an alternative hypothesis or relationship that comes into your mind and to report and interpret the results. While this should encourage you to make creative use of the data, your data explorations should be guided by reason. For instance, adding or dropping a control variable is likely insufficient. Choosing another modeling strategy, testing an interactive relationship, subsetting the data in a theoretically motivated way or exploring alternative dependent or explanatory variables could be reasonable options. The replication project counts toward 30% to the final grade.

Instructions

1. Choose **one** of the papers below. Read it and download the replication materials.
2. The replication paper should comply with the following **structure**:
    a) Header that reports course title, your name, submission date, and the paper chosen for replication.
    b) Brief summary of the original study in your own words. This should cover the following issues: What it is about, which data are used, which empirical strategy was implemented (including a brief discussion of whether you think it is valid to identify the causal effect of interest or not), and what the findings are.
    c) Analytic part along the lines of the paper-specific instructions.
    d) Bibliographic references if any.
3. We want to see **all code and output in the knitted file**. The document should start with importing the packages necessary for your analysis and loading the data.
4. Offer a **step-by-step description** of what you do. For example, first describe the model you replicate, then replicate the model (in an R code chunk), then describe the results. Comments in the code should be limited to very brief descriptions, e.g. “re-code forcing variable”, “run model”, or “plot figure 1”.
5. **Table output** should be nicely formatted in HTML. We don’t want to see the raw model output. Use `stargazer` or alternative packages. That said, when replicating tables of the original paper, some of the additional information provided (such as statistical tests

for the validity of an instrument) are not straightforward to replicate with R. We do not expect you to replicate these if they are not part of the standard output of the estimation in R.

6. **Graphs** you replicate do not have to look exactly like the graphs in the original paper. The should carry the same (or more) information though. If you feel the original graphs could or should be improved, please do so and explain why. Also, take care of proper labels, axes, etc. The graphs should look publishable. You can use colors.
7. There is **no limit with regards to length**, but text, code and output (figures, tables) combined should probably be no more than 8 pages when printed. Focus on concise, clear writing, efficient code, and appealing output.
8. The replication paper has to be submitted as a **knitted Rmd in HTML format** on GitHub **until May 25, 2020, 11.59pm CET**.

Replication papers

Paper 1

Bibliographic information. Broockman, David E. (2014). Do female politicians empower women to vote or run for office? A regression discontinuity approach. *Electoral Studies*. https://www.sciencedirect.com/science/article/pii/S0261379413001583.

Abstract. Persistent gender gaps in political officeholding and mass political participation jeopardize women's equal representation in government. This paper brings new evidence to the longstanding hypotheses that the presence of additional female candidates and officeholders helps address these gaps by empowering other women to vote or run for office themselves. With a regression discontinuity approach and data on 3813 US state legislative elections where a woman opposed a man, I find that the election of additional women in competitive US state legislative elections has no discernible causal effects on other women's political participation at the mass or elite levels. These estimates are precise enough to rule out even substantively small effects. These results stand in stark contrast to a number of findings from India, suggesting that although electing the first women in a society can have these empowering effects, remaining barriers to women's inclusion in American democracy go beyond what further increases in female officeholding can themselves erode.

Data. https://doi.org/10.7910/DVN/JTSYV3

Core modeling strategy. Sharp RDD

Paper-specific instructions.

- Replicate the three models in Table 1 without the State x year FEs. Also, replicate Figure 3.
- Replicate Table 2 and Figure 4.
- Use the available data to probe the robustness of the results, or to explore an alternative hypothesis/relationships that comes into your mind; report and interpret the results.

Paper 2

Bibliographic information. Cortina, Jeronimo B and Brandon Rottinghaus (2019). Vote centers and turnout by election type in Texas. *Research & Politics*. https://doi.org/10.1177%2F2053168019864224.

Abstract. The use of vote centers—specific locations in a county where all voters will vote—is on the rise nationwide, as more than a dozen states used this process by 2018. More states are moving toward using voting centers to remedy the problem of low voter turnout, with the assumption that the centralization of voting to several core county locations will increase voter accessibility. What we have less clear information about is the effect of vote centers on turnout in individual elections across several cycles. Using a natural experiment in Texas -a state that has three fixed election cycles - we find vote centers have a small positive impact on traditionally lower turnout elections but no effect on higher turnout elections. The cumulative impact of vote centers has a small effect on turnout over time. These results suggest a more cautious assessment is needed when considering the use and impact of vote centers.

Data. https://doi.org/10.7910/DVN/1EHQQS

Core modeling strategy. Difference-in-differences, propensity score matching

Paper-specific instructions.

- Replicate Tables 1, 2, and 3.
- Create a graph that allows you to assess the parallel trend assumption.
- Use the available data to probe the robustness of the results, or to explore an alternative hypothesis/relationships that comes into your mind; report and interpret the results.

Paper 3

Bibliographic information. Ritter, Emily H. and Courtenay R. Conrad (2016). Preventing and Responding to Dissent: The Observational Challenges of Explaining Strategic Repression. *American Political Science Review.*
https://doi.org/10.1017/S0003055415000623

Abstract. Although scholarly consensus suggests that dissent causes repression, the behaviors are endogenous: governments and dissidents act in expectation of each other's behavior. Empirical studies have not accounted well for this endogeneity. We argue that preventive aspects of repression meaningfully affect the relationship between observed dissent and repression. When governments use preventive repression, the best response to dissent that does occur is unclear; observed dissent does not meaningfully predict responsive repression. By contrast, governments that do not engage in ex ante repression will be more likely to do it ex post. We follow U.S. voting scholarship and propose a new instrument to model the endogeneity: rainfall. We couple rainfall data in African provinces and U.S. states with data on dissent and repression and find that dissent fails to have a significant effect on responsive repression in states that engage in preventive repression.

Data. https://doi.org/10.7910/DVN/T1JZ0L

Core modeling strategy. Instrumental Variable; Matching.

Paper-specific instructions.

- Replicate Table 1.
- Come up with a descriptive plot illustrating the core relationship of interest.
- Use the available data to probe the robustness of the results, or to explore an alternative hypothesis/relationships that comes into your mind; report and interpret the results.