Devan Arnold

Econometrics, ECO 530 (2410)

Jonathan Malacarne

University of Maine

12/4/23

# Task 1

## Prompt

In a few short paragraphs, address the following questions:

* What is the paper’s primary research question?
  + To paraphrase, the research question is: How do voters influence policy in the US – through policy compromise brought about by vote shares in general elections or by the election of the candidates who then enact the policies in their platform?
* What are the primary outcome and explanatory variables of interest?
  + Outcome variables
    - ADA Score (Americans for Democratic Action Score)
    - γ
    - Elect component of γ
    - Affect component of γ
* Why is it not sufficient to simply use ordinary least squares to answer the research question?
  + As we see in the data, there is a discontinuity at the vote share of democrats at 0.5 when compared to their ADA voting score (this would also be true if the party chosen was Republicans, since the system is essentially binary). This means that regression discontinuity design (RDD) can better estimate the reason for the jump in vote share versus ADA score in an attempt to explain if the jump is associated with the affect of voter share on policy, or the election of policies espoused by the politicians during the campaign.
* What makes the scenario a good candidate for regression discontinuity design?
  + The demographics are mostly continuous, and there is a clear discontinuity in the data. Additionally, elections go to the candidate with higher vote share and the process is (mostly) binary, which is the type of structured rule-based environment in which RDD thrives.
* What are the primary assumptions under which regression discontinuity design will result in a good estimate of the primary relationship of interest?
  + Arbitrary, rules-based environment, discontinuity present between dependent and independent variables, and no discontinuity when looking at other demographic factors

## Response

As mentioned in the introduction, this exercise will be an attempt to replicate the results of the 2004 paper “Do Voters Affect or Elect Policies? Evidence from the US House” by David S Lee, Enrico Moretti, and Matthew J Butler. This paper sets out to determine the manner in which voters influence policy in the federal government: do they directly elect politicians that then enact policies consistent with their campaign platform, or do voters instead influence how politicians vote?

To this end, the authors set out to compare the score issued by the Americans for Democratic Action (ADA) against the share of votes received during the previous election cycle. The authors utilize Regression Discontinuity Design (RDD) in their analysis to determine the ADA score gap that occurs at the 50% vote share line. Since US elections are essentially a binary choice, the 50% vote share line represents the point under which the measured party loses the election, and above which they win. The gap that occurs here is measured in the paper by gamma (γ), which itself is composed of two parts – the “elect” and “affect” portions. These portions serve as the means by which the authors answer their research question.

RDD is a good choice of estimation model for this problem for a few reasons. Firstly, as mentioned prior, US elections are in essence binary choices. The authors describe that at the +/-2% vote share level that the winner of an election is essentially random. Secondly, the authors are able to argue that there is no covariance in the demographics of the voting base that explains the discontinuity in the relationship between voting share and ADA score which strengthens the case for RDD. This randomization of outcomes in a rules-based environment and little covariance with other possible confounding influences creates a strong case for an RDD approach.

Over the next few tasks set out I will attempt to recreate the results that the authors achieved in the original paper.

# Task 2

## Prompt

Compare your estimate to the estimate reported in Column 1 of Table I

## Response

As we can see, the estimate that I generated from the data is 18.7, whereas the authors generated an estimate of 21.2. That puts my estimate approximately 1.3 standard errors (as measured by the authors) below the estimate that they generated.

# Task 3

## Prompt

Compare your estimate to the estimate reported in Column 3 of Table I

## Response

The value estimated using my model is 0.57, whereas the value estimated by the authors is 0.48. My estimate is significantly outside of the range that the authors estimated, and I am sure this is due to the estimation approach that I took. Although I reviewed the estimation methods outlined in Lee 2001 and Lee 2008 (formerly Lee 2003, but it appears that the author had the working paper published since this 2004 paper) I was unable to determine the exact estimation method given in this paper.

# Task 4

## Prompt

Discuss the purpose of this figure and how it supports (or does not) the rest of your analysis

## Response

As discussed in my summary of the paper, the lack of discontinuity present in these potential confounding factors strengthens the argument for the use of Regression Discontinuity Design in this context. This composite figure is meant to illustrate that fact.

# Task 5

## Prompt

Write a short conclusion summarizing what you did, what you found, and how well you were able to replicate the original authors’ results.

## Response

In this exercise, I took steps to review the work done by the authors, to understand the methods that they implemented, and to deconstruct their methods for replication. This required careful reading of the paper, consultation of external resources from both the course and outside of it, and I created the regression models to replicate the results of the authors.

In the end, my efforts to replicate this paper were mixed. My numerical estimations were way off base, but I was able to recreate the plots used by the authors for the most part. Personally, I think that this creates an interesting problem: economic papers have to explain what they did, why it matters, what their results are, and the implications of those results but recreating the work is hard. In other sciences, my understanding of the purpose of lab reports is to document the hypothesis of the researcher, the methods of the test, the steps the researchers took, and their results. However, in economics (and in this class as well) we have learned how direct replication and inclusion of steps for direct replication are discouraged at the institutional level. Economics has a hard time being taken seriously by other sciences (we have a perception problem), and for me this exercise highlighted that. I can see the results of the author, but explicit statement of methods is excluded from the paper.

I do not know how I could have done this task better, and while I can read the results of the authors, I cannot replicate their findings. This speaks to either my lack of skill in this area, the difficulty of reproduction via inference, a wider issue in economic literature, or some combination of the three.