Decision Memo

To: Empirical Analysts at The US Federal Election Commission
From: Gisele de Araujo and Chretien Li
Date: December 18th 2020
Subject: Evidence from the U.S. House shows that voters elect policy rather than affect policy

---

Executive Summary

Below we have analyzed Lee, Moretti, and Butler's well-cited 2004 paper, "Do Voters Affect or Elect Policies? Evidence from the U.S. House," in which they examined the question of whether voters affect or elect policy. The researchers used voting record data from the U. S. House (1946–1995) and a sharp-regression discontinuity design to arrive at their results. The results ultimately revealed that voters elect policy because electoral strength was shown to have no effect on candidate voting behavior. In this paper we replicate the main regression discontinuity figure from Lee et. al. and extend their findings by applying new technical methods available through R to their RD figure. By applying the function rdrobust, replacing N/A values with averages instead of zeros, and applying an optimal bandwidth we find evidence that their results are robust. Through our extension, we were able to support that a severe discontinuity exists among swing districts that are slight-dem and slight-rep in terms of voting policies, implying that voters do not affect how their representatives vote (partial/full convergence), rather choose to elect one of two platforms (full policy divergence)[1]. Given these findings, we suggest that the U.S. government reconsider whether current voting systems allow for the true representation of U.S. citizens.

We suggest that a study be conducted to investigate why candidates do not vote as they say they will. This will lead to actionable measures to hopefully empower voters and American democracy.

Introduction

The main figure of the Lee et. al. paper is Figure 1[2], a regression discontinuity figure which visualizes the result that voters tend to elect rather than affect policy. This is the figure we replicated. The discontinuous jump at the cut-point is labeled γ, and is made of two components: the elect and affect components. Formal analysis shows that the elect component is much larger than the affect component. This difference in magnitude suggests that the *elect* component is dominant and thus full policy divergence is the force at hand.

$$\gamma = \underbrace{\pi_0(P^{*D}_{t+1} - P^{*R}_{t+1})}_{\text{“Affect”}} + \underbrace{\pi_1(P^{*D}_{t+1} - P^{*R}_{t+1})}_{\text{“Elect”}}.$$

---

[1] Downs, A. 1957. "An Economic-Theory of Political-Action in a Democracy." Journal of Political Economy, 65(2), 135-50.
[2] Original Study Code and Data: https://eml.berkeley.edu/~moretti/data3.html

Figure I. Formal representation of LMB equation

The magnitude of the elect component means that candidates vote according to their own policy preferences (full divergence), and that it is not the case that voters and electoral pressure affect partial-convergence moderate voting. Lee et. al formally arrive at the conclusion that the elect component has a greater magnitude by subtracting the elect component ($\pi1(P^{D}$ t+1– P^{R} t+1) = 22.84) from the total gap (γ = 21.2), yielding the affect component (21.2-22.84 = -1.64, $\pi0(P^{D}$ t+1– P^{R} t+1 = -1.64). Clearly, elect (22.84) is greater than affect (-1.64).

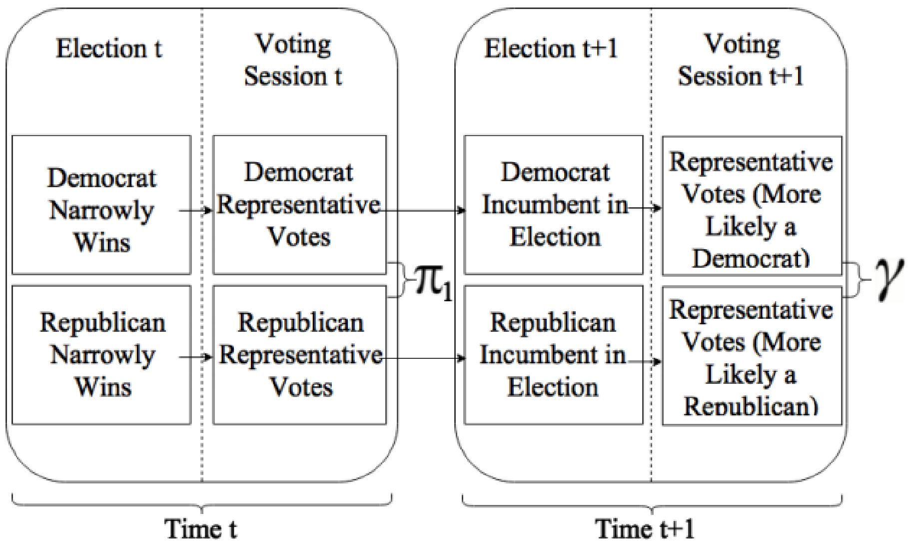


Figure II. Illustration of LMB equation. Source: http://repec.tulane.edu/RePEc/pdf/tul1518.pdf

One key assumption of this regression discontinuity figure is the initial "random assignment" of who won the 1992 election which allows for the γ gap to be attributed to the two components (elect and affect) mentioned above. Lee et. al. used quasi-experimental data from districts with extremely close election results to represent this random chance of which party was elected. This assumption accounts for the unobservable differences between districts and aims to reduce the associated bias.

By looking at the vote share of a region (electoral strength) in terms of how Democratic/Republican it is in comparison to a nominated candidate's voting record, we can begin to see whether voters affected or elected policy.

If we see that there is no discontinuity between a regression between voter shares and ADA scores, then the suggestion is that a Downsian paradigm is upheld: voters affect policy by making politicians more and more central, eventually converging their platforms. So politicians in regions with voter shares near the 50% mark should support a very similar policy platform, assuming they are honest and support what they say they will support. The alternative would be the observance of a discontinuity around 50%. This suggests that politicians do not actually vote for policies how they say they will; there is no platform convergence. Voters don't necessarily affect policies at all, only elect them. They don't affect how their representatives will vote, but instead they're only given two platforms to choose from.

We see that the ADA score (more liberal candidate voting record equals higher score) is significantly different on both ends of the cutoff (0.5, 50% democrat voter share), and that changes in shares of votes do not affect how a official votes, thus there is reason to believe that voters primarily elect policies, not affect them.

Replication[3]:

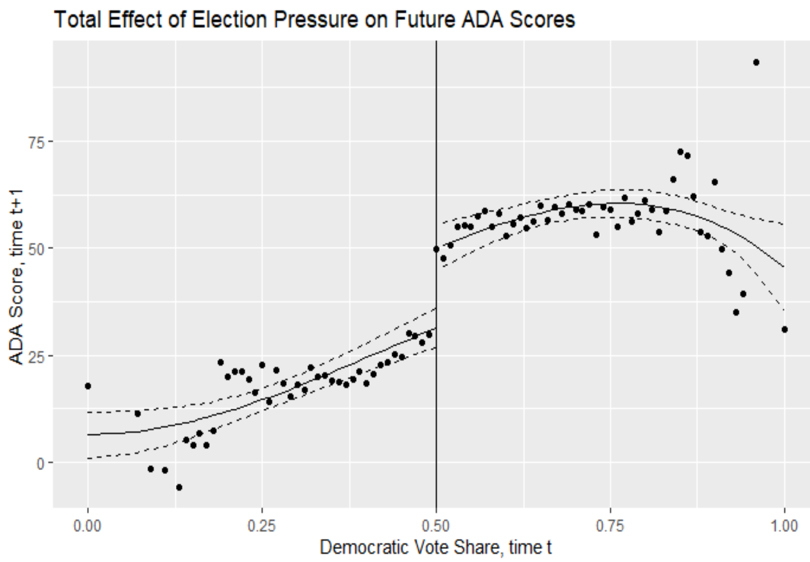
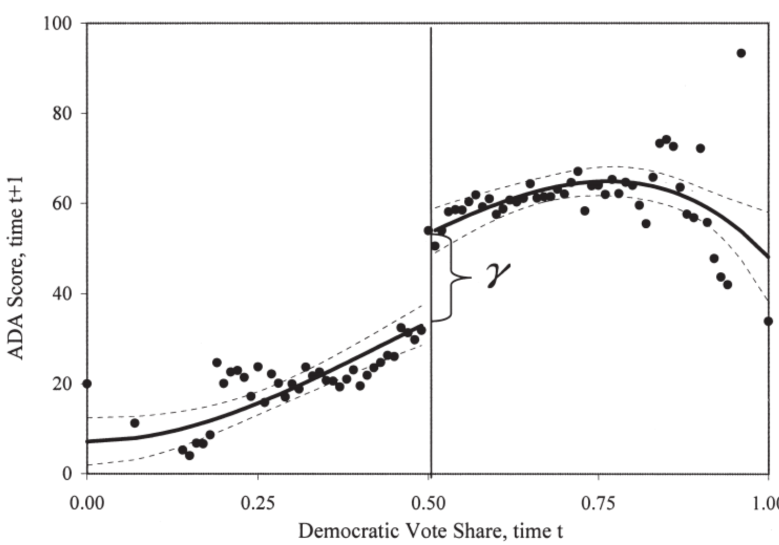


Figure IIIa. Replication of Fig.1in "DO VOTERS AFFECT OR ELECT POLICIES? EVIDENCE FROM THE U. S. HOUSE" by Lee et. al. (2004). Solid line is regression with dotted line being 95% confidence interval[4]



[3] Replication & extension code: https://github.com/Chrecci/CS112-Final-Research-Replication-Do-Voters-Elect-or-Affect-Policy-

[4] #dataviz: We analyze Figure 1, the regression discontinuity figure, from Lee et. al. and create a sophisticated replication.

Figure IIIb. Original figure by Lee et. al. (2004). Gamma is the total effect of election on ADA scores. Gamma consists of two components, the "elect" and "affect" components. See "Introduction" section for more.

Our replication of Lee et. al. Figure 1. has a few key differences in approach compared to the original that are important to call out. Firstly, the original authors' work is done entirely through STATA, whereas our replication translates their code into R. We opted for this option predominantly because (1) as authors we are much more familiar with programming in R, and (2), a wide set of new techniques available to regression-discontinuity analysis is, in our eyes, better implemented in R. However, this will likely lead to some variability between our answers and the original[5]. Second, we don't have access to the original source data[6], and so have no way of validating the accuracy of data available; however, we trust the integrity and accuracy of the original research design, and we have no reason to believe otherwise.

Considering the age of the research, it is not too shocking to find that the original approach taken by Lee et. al. seems rather archaic, with nearly two decades of code and research design improvements having passed by. However, their approach is still legitimate. After filtering the data, higher-order variables of the independent variable, voter share, are manually constructed. Then, the dataset is collapsed by the hundredth of a vote share. A 4th order polynomial regression is then run across the entire dataset. Finally, after splitting the regression along the cut-off point, 0.5, a graph is constructed illustrating the treatment effect to be $\gamma = 21.1$.

We follow this exact procedure to the best of our capabilities. To our delight, our replication yields a treatment effect of 21.1, well within an acceptable margin of error. One thing that remains unclear is how the original paper calculates the treatment effect. Most modern tools utilize the limit of the regression as it approaches either side of the cutoff point; however, this isn't perfectly evident for Lee et. al. Regardless, we tested a few different methods to calculate the treatment effect, all of which yielded a similar result around the order of 20.

Overall, the replication was a success, reaffirming the conclusions drawn by Lee et. al. in the original research. A significant, and expected, discontinuity is observed in voter scores (ADA). Whether this discontinuity exists predominantly due to the elect or affect component determines what conclusions can be drawn concerning the Downsian paradigm. This work is done in Figure IIa and Figure IIb of the original paper, and although it falls outside the scope of our replication, we confirm that their findings in both those figures are indeed in line with our own. The conclusion that voters merely "elect" policies appears to be supported.

Extension:

Upon closer examination, we found a few disheartening techniques employed by the original paper. The most glaring one is a failure to implement any bandwidth at all. This paper precedes when

[5] Errors may arise in 2 ways. (A) Interpretive errors where human errors lead to a non-fully accurate translation of code from one language to the other. (B) Algorithmic errors. Due to the nature of the source code of both languages, certain operations may be carried out with slight variation. Although we ensure a rigorous validation and checking of our work, it is nearly inevitable that some margin of error bleeds through our replication.

[6] We have the dataset that Lee et. al. created after their own data processing; however, we do not have access to the original source data from which Lee et. al. received their data from

bandwidths started becoming nearly universally applied to RD designs, but it is important to have nonetheless. RD's fundamental premise that allows it to infer a treatment effect, is that at the cut off point, where treatment is administered, observations falling near the cutoff point should be similar across covariates. However, this assumption is not sound the further we deviate from our cutoff point. A bandwidth is essentially a bound of acceptable observations around the cutoff point that is appropriate to run either a local linear[7] or global polynomial regression. Having too large of a bandwidth is prone to introduce bias, and too narrow introduces variance; hence, bandwidth setting is often referred to as having a bias-variance trade off (Cattaneo et. al., 2019). This is one potential risk of the original research. By not setting a bandwidth, that is, regressing across all observations, even those that are likely to be dissimilar to one another, a high bias can be introduced very easily.

The second issue is an arguable mishandling of data. In order to shorten and simplify the dataset, Lee et. al. collapses the data by voter share. So each observation with say voter share equalling 0.55 gets aggregated into one observation, and the ADA score of this new observation being the average ADA score of all observations that had a voter share of 0.55. The issue lies in the handling of missing values. According to the source code of STATA's collapse function, missing values are automatically replaced with 0's when calculating for the average. Lee et. al fail to account for this fact. Given that null values are evenly distributed among evenly distributed observations, the replacement of null values with 0 shouldn't reflect too significantly of a change in our final results; however, in order to produce a more robust and accurate finding, we argue a better strategy ought to be devised.

Our extension seeks to produce new results while accounting for these criticisms[8].

It's rather intuitively unlikely that the voter score of a congressional representative in a district with missing data is 0. In fact, this is a rather rare occurrence. If the underlying assumption of the research is that districts with a similar distribution of voter shares should also have candidates that vote similarly on policies, then we argue a better handling of missing values is imputing the average voter score based off of available data.

Obs.	Ada Score	Voter Share
1	NA	0.55
2	40	0.55
3	50	0.55

Obs.	Ada Score	Voter Share
1	0	0.55
2	40	0.55
3	50	0.55

Obs.	Ada Score (mean)	Voter Share
collapse1	30	0.55

Obs.	Ada Score	Voter Share
1	NA	0.55
2	40	0.55
3	50	0.55

Obs.	Ada Score	Voter Share
1	45	0.55
2	40	0.55
3	50	0.55

Obs.	Ada Score (mean)	Voter Share
collapse1	45	0.55

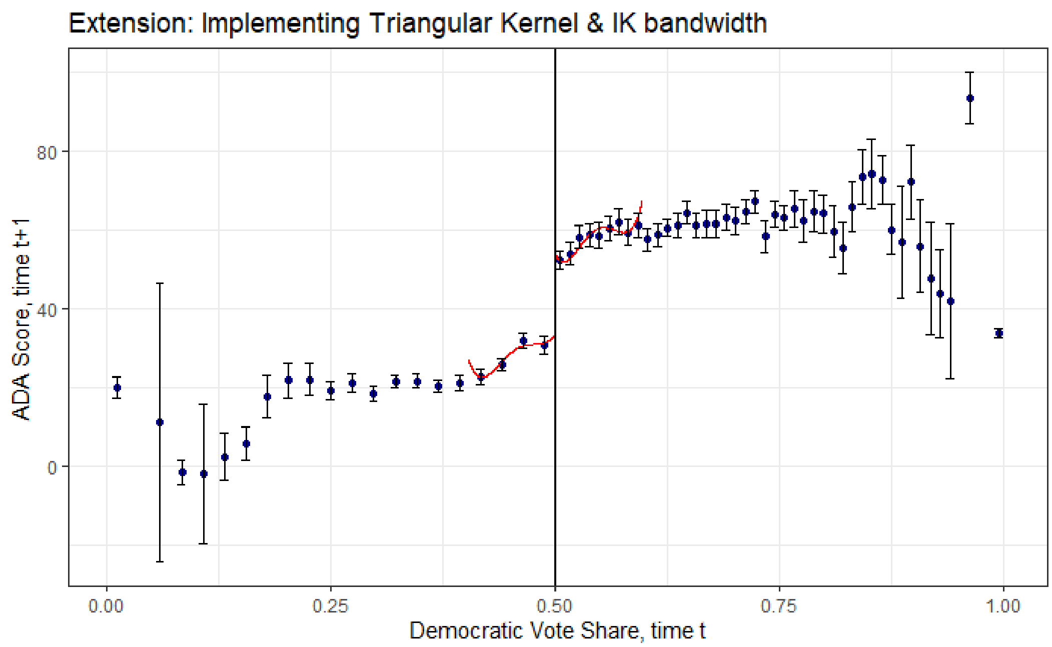
Figure IV. Handling of missing values. Top path is taken by original researchers, where missing values are replaced with 0's. Bottom is our proposed approach, with imputation of mean

[7] Modern literature also seems to suggest a local linear regression ought to be preferred for RD designs. This is because local linear estimators are rate optimal and have attractive bias properties (Porter, 2003)

[8] Another subtle difference to note, is that our extension ignores polynomial order. An appropriately chosen bandwidth will adjust to the chosen polynomial order anyway, still resulting in a reliable linear approximation to the unknown regression functions (Cattaneo et. al. 2019)

Surely enough, we find the mean of ADA scores changes from 38.71 to 41.57. Again, intuitively, the closer the average ADA score is to 50 (balance of left and right leaning voters in congress), the more plausible the scores are.

Our main extension is adding a kernel and bandwidth selection process, both of which were not employed in the original study. As mentioned earlier, a bandwidth is the area of interest around a cutoff point. The selection of an appropriate bandwidth is critical. The kernel sets a weight assigning method for the construction of the local-polynomial estimator (Cattaneo et. al., 2019). Essentially, the kernel determines the regression that takes place around the cutoff. We opted for a triangular kernel over a rectangular/uniform kernel[9]. As for our bandwidth selection, we felt it most appropriate to use an IK bandwidth as proposed by Imbens & Kalyanaraman (2009). Not only was this method specifically derived for another similar RD-design study[10], it also makes practical sense for our situation. The IK bandwidth selection method also minimizes MSE, but doing so only for immediate points around the cutoff. This is preferred over other bandwidth selection methods, such as cross-validation bandwidth approaches, that calculate across the entire data range, since "further away" observations are too different across covariates, increasing bias[11].



[9] Triangular assigns weight only to values within bounds, with increasing weight closer to the cutoff point. Uniform kernel assigns equal weights to all values within bounds. Triangular is argued to be better, especially since when paired with a bandwidth estimator that optimizes MSE, it leads to a point estimator with optimal properties.

[10] Lee, D. S. (2008). Randomized experiments from non-random selection in u.s. house elections. Journal of Econometrics, 142(2), 675–697. https://doi.org/10.1016/j.jeconom.2007.05.004

[11] #optimization: our optimization problem is to minimize our MSE, mean squared error. The intent here is to hopefully arrive at a bandwidth that gives us the most accurate and relevant result. The problem is a rather complex and challenging one. We identified the bias-variance tradeoff at play and carefully evaluated a numerous set of options that could help us reach our optimal bandwidth. Ultimately a conclusion was reached, as the variables considered by IK bandwidth are simply more relevant appropriate. In order to fully maximize the effectiveness of IK bandwidth, a triangular kernel and local linear regression ought to be employed as well, both of which we included for best results.

Figure V. Regression Discontinuity design incorporating a triangular kernel and IK bandwidth selector and 95% confidence intervals. The IK bandwidth is set to 0.09674781[12]

Our replication yielded a treatment estimate of 21.1. Following our extension, the estimate dropped slightly to 19.9. As is apparent from *Figure V*, the discontinuity is clearly visible, with the confidence intervals narrowing as they near the cutoff point. We believe this to be a more robust and accurate estimate of the true treatment effect; however, the similarity of results between our study and the one by Lee et. al. done nearly two decades ago stands as a testament to their solid research design.

Conclusion

Through our replication and extension, our findings support Lee et. al's conclusion that U.S. voters elect rather than affect policy. The results support that even in borderline elections, candidates tend toward full divergence, voting according to their personal policy preferences rather than converging on their platforms in response to voter share. The explanation best supported by the literature is that politicians have a credibility problem founded in their inability to commit to their promises to be more moderate, and this credibility problem drowns out any observable Downsian effects. These are troubling results because the implication is that while population voting preferences change, their representatives' policy behaviors do not. While voters are told they have the power to affect policy, in reality they merely choose from the (often two) platform options presented. If you agree that these results inspire concern that the U.S. political system is not working as it should be, we suggest that a study be conducted to investigate why candidates do not vote as they say they will. This will lead to actionable measures to hopefully empower voters and American democracy.[13]

---

[12] #dataviz: We made improvements to the original RD figure by incorporating a triangular kernel, IK bandwidth selector, and 95% confidence intervals.

[13] #algorithms: We effectively implemented working code in our replication which was challenging because we had to translate the code from STATA to R, but we did so successfully! Being able to translate this code required a thorough understanding of each algorithmic step that went into creating the RD figure. By creating an accurate replication and supplementing it with new techniques and methods available in R, we were able to demonstrate a very strong application of our algorithmic skills.

References

Button, P. (2018). A replication of "do voters affect or elect policies? evidence from the u.s. house" (quarterly journal of economics, 2004). *Public Finance Review*, *46*(5), 886–893. https://doi.org/10.1177/1091142117721739

Cattaneo, M., Idrobo, N., & Titiunik, R. (2020). *A Practical Introduction to Regression Discontinuity Designs: Foundations* (Elements in Quantitative and Computational Methods for the Social Sciences). Cambridge: Cambridge University Press. doi:10.1017/9781108684606

Downs, A. 1957. "An Economic-Theory of Political-Action in a Democracy." Journal of Political Economy, 65(2), 135-50.

Gelman, A., & Imbens, G. (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, *37*(3), 447–456. https://doi.org/10.1080/07350015.2017.1366909

Imbens, G., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. Review of Economic Studies, 79(3), 933–959.

Lee, D. S. and T. Lemieux. 2010. "Regression Discontinuity Designs in Economics." Journal of Economic Literature, 48(2), 281-355.

Lee, D. S., Moretti, E., & Butler, M. J. (2004). Do voters affect or elect policies? evidence from the u. s. house. The Quarterly Journal of Economics, 119(3), 807–859. https://eml.berkeley.edu/~moretti/final.pdf

Porter, J., (2003), "Estimation in the Regression Discontinuity Model," Working Paper, Harvard University Department of Economics, draft date September 25, 2003.

Code:
https://cran.r-project.org/web/packages/rdrobust/rdrobust.pdf
https://www.stata.com/manuals13/dcollapse.pdf

Author Contributions:
Since we are both in the same location, we were luckily able to collaborate closely on each step of this assignment from brainstorming to visiting Hadavand's office hours to analyzing the paper and creating the replication and extension. Chretien took the lead on translating STATA code into R for the replication and extension figures, and Gisele joined to help apply the RD concepts learned in class to our replication and extension. We both iteratively wrote and edited this paper together.

Exceptionally good replication that successfully recovered the original result and discussed shortcomings in the original paper.

5 #cs112-decisionreview