Partisan Registration: A Truthful Statement or A Strategic Choice?

Fang Guo*

March 1, 2023

Abstract

Is partisan registration a truthful revelation of an individual's party identification, or is it a strategic choice for the purpose of voting in partisan primaries? Although scholars have widely used party affiliation as a proxy for party identification, no study has examined if it is a sincere expression of partisan attachment. My study offers a framework to analyze the conditions under which voters may or may not register for strategic reasons. Empirically, using a regression discontinuity design with individual-level voter file data from New York, I find no evidence of strategic party registration. My findings suggest that individuals might derive significant psychic benefit from truthfully registering with the party that best reflects their political beliefs.

*Ph.D. Student, Stanford University Graduate School of Business, California, USA. fangg@stanford.edu

1 Introduction

Voter registration in the US is a required first step before one can vote in any local, gubernatorial, or federal election. Around two-thirds of the American voters choose to register either as a Democrat or as a Republican in their voter registrations, while the rest register as independents or members of third-parties.[1] Is partisan registration a truthful revelation of an individual's party identification, or is it a strategic choice for the purpose of voting in partisan primaries? Many studies have used individual-level party affiliation to demonstrate partisan segregation and shifts in political attitudes, which have significant implications for how we study political polarization (Miller (1991); Gerber, Huber, and Washington (2010); Martin and Webster (2020), Cantoni and Pons (2020); Brown and Enos (2021)). Despite the wide use of party affiliation as a proxy for party identification, no study has examined whether there is any discrepancy between the two. Some recent studies acknowledge the possibility of strategic registration (Brown (2021)), but there is no rigorous discussion of the incentives behind strategic party affiliation and no empirical evidence of whether it exists or not.

Especially in states with closed primaries where only registered party members can vote, voters might register with a particular party not because they are politically closest to the party, but because they want to participate in and influence the partisan primaries. For example, in a deep-blue district that is dominated by Democratic voters, the Democratic candidate will be heavily favored in the general election. Voters in such districts foresee the landslide victory of the Democratic candidate, and understand that the actual election that determines the final outcome is the Democratic primary, not the general election. In fact, many candidates from the minority party of a district do not actively campaign at all, and most media coverage goes toward the partisan primary. If the primary elections in such districts are closed, then there exists a potential incentive for non-Democratic voters, especially voters who are in the middle of the ideological spectrum, to register as Democrats.

Historically, one-party dominance in the US at the local level is not a rare instance. The most extended period of one-party dominance is the "Solid South" from the late twentieth century to the Civil Rights era. Between 1880 and 1950, the Democrats won virtually every state office in every state of the former Confederacy (Hirano and Snyder Jr (2019)). Between 1878 and 1910, the Democrats won at least 90% of the vote in more than half

[1] See https://ballotpedia.org/Partisan_affiliations_of_registered_voters.

of the U.S. House elections in those states (Ansolabehere et al. (2010)). In recent years, American neighborhoods also exhibit stark partisan segregation, with the majority of voters living with virtually no exposure to voters from different political parties in their residential environment (Sussell (2013); Brown and Enos (2021)). As a result, many congressional districts could be dominated by a single party.

The possibility of strategic registration also closely relates to the discussion of the primary election system in the US. Proponents of the open primary system argue that it prevents elected officials from being pulled to the ideological extremes by allowing more moderate voters to participate in the primary elections, thereby reducing polarization (Fiorina, Abrams, and Pope (2005)). However, McGhee et al. (2014) empirically shows that the openness of a primary election has little, if any, effect on the extremism of the politicians it produces. If strategic registration exists, it could potentially explain the null difference between the open and closed systems, as moderate voters might have already registered as partisans in closed primaries.

In this paper, I offer a framework to analyze the conditions under which strategic registration might occur. Then, I use voter file data from New York in 2018 to empirically test whether strategic registration exists. In particular, by exploiting spatial regression discontinuities at congressional district borders where two sides of the border have almost the same environment except for different levels of electoral competition or party dominance, I do not find evidence of voters registering strategically in order to vote in partisan primaries. My results are robust to various settings and specifications, suggesting that individuals might derive significant psychic benefit from truthfully registering with the party that best reflects their political beliefs.

2 Simple Framework

The classic instrumental voting theory posits that rational voters respond to the change in the probability of being pivotal (Riker and Ordeshook (1968)). Specifically, an individual's decision to vote is affected by four factors: the probability of being pivotal (P), the benefit derived from the election of her preferred candidate over the closest competitor (B), the psychic benefit of voting (V), and the cost of voting (C). Under this framework, a voter will cast a vote if an only if:

$$PB + V > C \tag{1}$$

This model implies that competitive elections increase voter turnout (Downs et al. (1957); Kelley, Ayres, and Bowen (1967); Wattenberg (2002); Franklin et al. (2004); Arceneaux and Nickerson (2009)).[2]

Similar to the theory of voter turnout, I present a simple framework of strategic party affiliation. Suppose there is a one dimensional ideological space spanning from -1 to 1, with -1 being very Democratic, and 1 being very Republican. Every voter i has an inherent ideological position in this space $\theta_i \in [-1, 1]$. Suppose that there are two arbitrary thresholds $\underline{\theta}$ and $\overline{\theta}$, where $\underline{\theta} < 0 < \overline{\theta}$, that voters implicitly use to identify with a particular party. Specifically, voter i identifies as a Democrat if $\theta_i \leq \underline{\theta}$, as an Independent if $\underline{\theta} < \theta_i < \overline{\theta}$, and as Republican if $\theta_i \geq \overline{\theta}$. Denote her true party identification as $z_i \in \{D, I, R\}$.

The district has closed primaries where only registered partisans can vote, and there are two parties in this district: the Democratic party and the Republican party. An individual may choose to register as a Democrat, Republican, or Independent before the primary election, and I denote her choice as x_i. Depending on her choice of x_i, she may also need to decide whether to vote in a primary or not and whom to vote for (denote as y_i). A voter's registration (x_i) may or may not align with her true party identification (z_i). When party registration does not align with one's inherent identification, I name this behavior as *strategic registration*. The voter derives a psychic benefit W from registering with the party that she identifies with. During the primary election stage, the voter also derives a psychic benefit V from voting and incurs a cost C of voting.[3]. For each voter, there is a benefit $B^i(A)$ of electing candidate A as the partisan nominee. Since the analysis is at the primary election, the benefit $B^i(A)$ is not for electing candidate A as the representative, so the determinants of $B^i(A)$ could be flexible. For some voters, $B^i(A)$ could be decreasing in the distance between θ_i and the candidate's ideological platform θ_A. For others, $B^i(A)$ may not necessarily correlate with ideological distance, but instead depends on how likely candidate A will help the voter's party win the general election.

Strategic registration may arise in various scenarios. One scenario is that when the vast majority of the voters in a district identify as partisans from one party, the independent

[2]Some recent studies cast doubt on this claim, see Enos and Fowler (2014) and Moskowitz, Schneer, et al. (2019).

[3]I assume there is no registration cost, so voters are always registered.

voters and voters who identify with the other party might have an incentive for strategic registration. Suppose that there is a district that is dominated by Democratic voters and it needs to elect a single representative. Each party first holds a closed primary election to select a nominee, and the two nominees compete in the general election. Suppose that two Democratic candidates D_1 and D_2, and one Republican candidate R_1 are running for this election.[4] The two Democratic candidates have $\theta_{D_1} < \theta_{D_2} < \underline{\theta}$, and Republican candidates has $\theta_{R_1} > \overline{\theta}$. Given that the Democratic voters make up a majority in this district, both D_1 and D_2 have a high chance of defeating R_1 in the general election.

Since any registered voter, regardless of her partisanship, can vote in the general election, I focus my analysis on the primary election stage. For example, there is a voter i in this district who internally identifies as a Independent ($\underline{\theta} < \theta_i < \overline{\theta}$). Suppose that for voter i, her benefit function $B^i(\cdot)$ depends on the ideological distance, and we have $B^i(D_2) > B^i(D_1) > B^i(R_1)$. She mainly faces two choices: register as an Independent but not vote in any primary, or register as a Democrat and vote for D_2 in the Democratic primary. She will choose to register strategically as a Democrat if:

$$P(B^i(D_2) - B^i(D_1)) + V - C > W \tag{2}$$

where P is the probability that i casts a decisive vote in the Democratic primary. If i is a Republican voter, she also faces a similar situation as in equation 2. Essentially, the voter's party affiliation depends on the values of P, $B(D_1)$, $B(D_2)$, V, W and C. For example, if the psychic benefit W of registering with the party that a voter identifies with is sufficiently small, or the Democratic party has fairly competitive primaries, then strategic registration may arise.

Alternatively, there could be another form of strategic registration. Suppose that the district still has the same three candidates, D_1, D_2, and R_1, but the district is not as heavily dominated by the Democrats as the previous case, so both parties have a chance to win the general election. If a Republican voter i cares about helping R_1 win the general election, she could theoretically register as a Democrat and vote for the more extreme candidate D_1 despite being closer to D_2, because if the more extreme Democratic candidate becomes the nominee, the Republican nominee might have a higher chance at winning the general election.

[4]My model does not focus on how candidates decide to run, and assume that the candidates and their policy platform are exogenous.

3 Empirical Study

Empirically, I test for the possibility of strategic registration in one scenario from the simple framework when the district has closed partisan primaries, one party dominates the district in the sense that voters expect that party to win the general election with high probability, and only the dominant party has a contested and potentially competitive primary. In addition to the three conditions, the relative magnitude of the utilities derived from pivotal voting, participation, partisan identification, and the cost of voting together determine whether a voter registers strategically or not.

Specifically, I compare neighboring districts where one district satisfies the three key conditions for strategic registration but the other one doesn't. For individuals who live close to the district boundary, they live in a similar environment except for the difference in congressional districts, and their individual characteristics should vary smoothly at the boundary. Therefore, I can use a regression discontinuity design (RDD) to estimate the extent of strategic registration.

3.1 Voter and Election Data

I use individual-level voter file data from New York in September 2018, collected from *L2-Data*. In addition to the basic voter information, L2 also includes the longitude and latitude coordinates of voters' residential addresses. Therefore, I can precisely locate the congressional district that each voter lives in, and the distance from each residence to the district boundaries. I have also collected vote share data for the 2018 U.S. House primary and general elections from *CQ Press*.

New York offers an ideal setting to test strategic party registration. First, New York has closed primaries. Second, the New York metropolitan area is densely populated, with many observations close to the district borders. More importantly, New York has many deep-blue districts where the Republican party rarely holds a contested primary. For example, the entire New York metropolitan area consists of 17 congressional districts (Figure 1), but only District 11 had a Republican primary with more than one candidate in 2018. In all other districts, registered Republican voters did not have any chance to participate in a contested primary. On the other hand, eight districts had Democratic primaries with more than one candidate, and the outcomes of some primaries were fairly close (see Table 1).

Figure 1: Congressional Districts in the New York Metropolitan Area

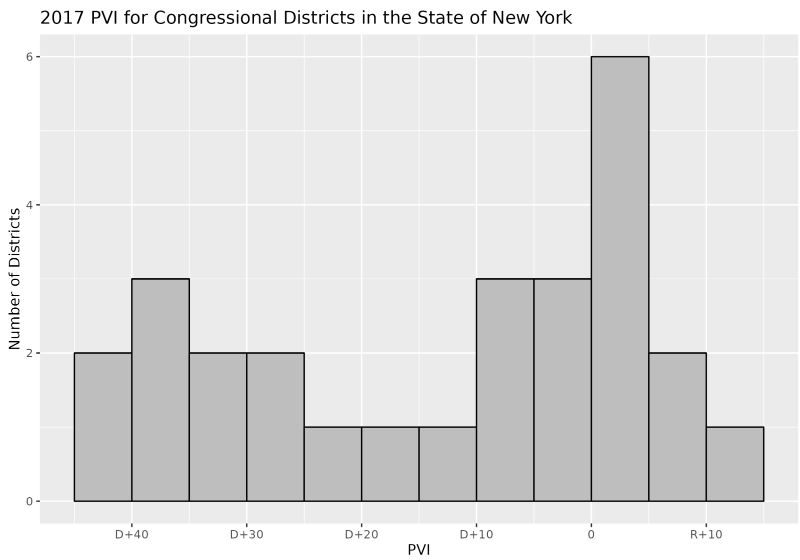
3.2 Measuring Partisan Dominance

Partisan dominance is a key condition for strategic registration to occur. Theoretically, a party is dominant in a district if any candidate from that party could win the general election with a probability of one. In reality, it means that an overwhelming majority of the voters support that party. Past election vote share offers a good measure of partisan dominance. Specifically, I use the 2017 Partisan Voting Index (PVI) from *Cook's Political Report*. The PVI describes the extent to which a given congressional district favors a Democratic candidate or a Republican candidate relative to the national average vote share in the 2012 and 2016 presidential elections. The PVI formula is the following:

$$PVI_i = \frac{(D_{i,2012} - A_{2012}) + (D_{i,2016} - A_{2016})}{2} \tag{3}$$

where $D_{i,2012}$ is the Democratic share of the two-party vote in the 2012 presidential election for district i, and A_{2012} is the national average Democratic share of the two-party vote in the 2012 presidential election. $D_{i,2016}$ and A_{2016} are similarly defined for the 2016 presidential election. Basically, the PVI takes the average of the mean-deviations in the Democratic presidential vote share over two election cycles. A PVI of 0 indicates that the district is evenly split between supporters of both parties. A PVI of D+10 (R+10) indicates that there are on average 10% more Democrats (Republicans) than the national average. While the PVI is derived from the presidential vote, it is highly correlated with congressional

Figure 2: The Distribution of PVI among Congressional Districts in New York



election win margins (Moskowitz, Schneer, et al. (2019)). In Figure 2, I plot the distribution of PVI among the congressional districts in the state of New York. A fair number of districts exhibit clear partisan preference, with some districts reporting a PVI as high as D+40.

In order for voters to believe that a district is dominated by a particular party, the PVI should be fairly large. I define a district as being dominated by the Democrats (Republicans) if it has a PVI score of D+15 (R+15) or more. If the national average of the Democratic vote share is 50%, then a district with D+15 on average has 65% Democratic votes and only 35% Republican votes over two presidential elections. A 30% difference should be large enough for voters to see the Democratic party as the dominant party in the district.

Applying the D+15/R+15 criteria to the districts in New York, there are 11 districts that are dominated by the Democratic party: District 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16. These districts cover the major areas of Bronx, Brooklyn, Manhattan, and Queens. In the 2018 U.S. House general elections, Democratic candidates had landslide victories in all of these districts, with some winning margins as large as over 50% (See Table 1 and Table 2).

3.3 Regression Discontinuity

My main outcome variable is a binary indicator of whether a voter registered as a Democrat or not in 2018. Treated districts satisfy the three necessary conditions for strategic registration (described at the beginning of section 3). Control districts border the treated districts but do not meet all the criteria.[5] I use the classic regression discontinuity specification:

$$y_i = \beta W_i + f(D_i) + \gamma X_i + \alpha_{b(i)} + \epsilon_i \tag{4}$$

where y_i is the indicator variable of democratic party affiliation for voter i in 2018. W_i is a binary indicator of whether the individual resides in the treated district. $f(D_i)$ is the locally linear regression of the running variable D_i, the distance to the boundary, which is negative for voters who live outside of the treated district, and positive for those who live inside. I follow Calonico, Cattaneo, and Titiunik (2014) for optimal bandwidth selection and bias correction. X_i includes individual characteristics such as age, gender, race, marital status, and household size, which are available in the L2 data. I also include boundary fixed effects $\alpha_{b(i)}$ to ensure that the specification is comparing voters who live close to the same boundary in cases where the treated district is surrounded by several control districts.

In the subsequent three sections, I discuss three methods to define treated and control districts based on 1) whether there was a contested primary (with more than one candidate), 2) whether the primary was competitive, and 3) whether a district was dominated by a party in the long term.

3.4 Contested Primaries

In my first approach to categorize treatment and control status, both the treated and control districts are dominated by the Democratic party. The main difference is that in 2018, the treated district had a contested Democratic primary (with more than one candidate) and an uncontested Republican primary (only one candidate), while the control district did not have a contested primary in either party. The main test here is to see whether more individuals in the treated districts register as Democrats in order to vote in contested primaries.

According to these criteria, there are five treated districts and six control districts. I pool together pairs of adjacent districts where one is treated and the other is control. Each treated district might have one or more neighboring control districts as the boundaries are two dimensional. In Table 1 and Table 2, I report the PVI scores and the actual election vote shares in both the general election and the partisan primaries in 2018. Consistent

[5] Empirically, I offer three approaches to define treated and control districts based on the key conditions, which will be explained in the following three sections 3.4, 3.5, and 3.6.

Table 1: Vote Shares in the 2018 U.S. House Election for Treated Districts

District	PVI	General Election		Democratic Primary		Republican Primary	
		Democratic	Republican	Nominee	Second	Nominee	Second
5th	D+37	100%	0%	82%	10%	0%	0%
9th	D+34	89%	10%	53%	47%	100%	0%
12th	D+31	86%	12%	60%	40%	100%	0%
14th	D+29	78%	14%	57%	43%	100%	0%
16th	D+24	100%	0%	74%	16%	0%	0%

Table 2: Vote Shares in the 2018 U.S. House Election for Control Districts

District	PVI	General Election		Democratic Primary		Republican Primary	
		Democratic	Republican	Nominee	Second	Nominee	Second
6th	D+16	91%	0%	100%	0%	0%	0%
7th	D+38	93%	0%	100%	0%	0%	0%
8th	D+36	94%	0%	100%	0%	0%	0%
10th	D+26	82%	18%	100%	0%	100%	0%
13th	D+43	95%	5%	100%	0%	100%	0%
15th	D+44	96%	4%	100%	0%	100%	0%

with the large PVI scores, the Democratic party in these districts won the general elections by large margins (more than 50%). In addition, all districts had completely uncontested Republican primaries, and some did not even have a Republican candidate. Specifically, in the control districts, there was no contested Democratic primary either, so voters did not have any incentive to affiliate with any party in order to vote in a partisan primary. In the treated districts, there were contested Democratic primaries, and some of them were fairly competitive in terms of actual vote shares from an ex-post perspective. Notably, in all the treated districts, the margin of victory in the general election is always larger than that in the Democratic primary, which means that a voter's probability of being pivotal in the Democratic primary is larger than that in the general election. As a result, the Democratic primary plays a more important role in determining the election outcome than the general election.

A key assumption in an RDD is that all relevant factors besides the treatment should vary smoothly at the boundary. This assumption is necessary to establish that voters who live just outside of a treated district serve as an appropriate counterfactual for voters who live just inside the treated district. To ensure that this assumption holds, I first check whether the congressional district boundaries overlap with other boundaries. In particular, I exclude treated-control pairs from my sample if the district boundaries coincide with school district boundaries. I also exclude pairs where the two districts are separated by a big park

Table 3: Treated and Control District Pairs

Treated	Control
5th	6th
	7th
	8th
9th	7th
	8th
	10th
12th	10th
14th	6th
	15th
16th	13th

(e.g. Central Park) because in this case, there are no observations close to the boundary.[6] The remaining qualified district pairs are shown in Table 3. In addition, I perform a balance check on individual-level demographics variables such as age, gender, race, marital status, and household size in Appendix A Table 7. No variable exhibits sizable difference between the treated and control districts.[7]

If voters have an incentive to register as Democrats to vote in a contested primary, we should expect more individuals who live just inside the treated district to register as Democrats. In Table 4, I show the RDD estimates based on Equation 4. In all specifications, the signs of the estimates do not support strategic registration and none of the estimates is significant. I also visually show in Appendix A Figure 4 that there is no significant change at the district boundary. I further show in Appendix A Figure 5 that the results are generally robust to various choices of bandwidths. My findings imply that it is unlikely for voters to strategically register for the purpose of voting in a contested primary.

3.5 Competitive Primaries

My previous analysis shows that individuals do not strategically register to vote in contested primaries, suggesting that the psychic benefit of voting in a primary is not enough to motivate individuals to affiliate with an opposite party. In my sample, although the treated districts had contested primaries, these primaries were not always competitive. In some dis-

[6] Specifically, I dropped the 12th-13th boundary, 14th-13th boundary, half of the 12th-10th boundary, and the 16th-13th boundary.

[7] Although some variables have statistically significant differences, the average values in the treated and control groups are very close. For example, the mean age in the treated districts is 48.918 and the mean age in the control districts is 49.594. Even though the difference of 0.6 year of age is significant, it is too small to be practically meaningful.

Table 4: Regression Discontinuity Design Finds No Significant Difference in Democratic Party Registration Between District With and Without Contested Democratic Primaries in New York in 2018

	(1)	(2)	(3)	(4)
Contested Primary	0.000	-0.009	-0.012	-0.011
	(0.016)	(0.014)	(0.013)	(0.016)
Bandwidth	65.615	91.306	88.017	73.973
N	64,957	89,409	86,764	74,776
Covariates	No	No	Yes	Yes
Kernel	Uniform	Triangular	Uniform	Triangular

Notes: All columns use regression discontinuity under local linear regression with bias correction and optimal bandwidth as in Calonico, Cattaneo, and Titiunik (2014). Robust standard errors are reported in parenthesis. Bandwidths are measured in meters. Column (3) and (4) include boundary fixed effects, as well as voters' age, gender, race, marriage status, and household size as covariates.

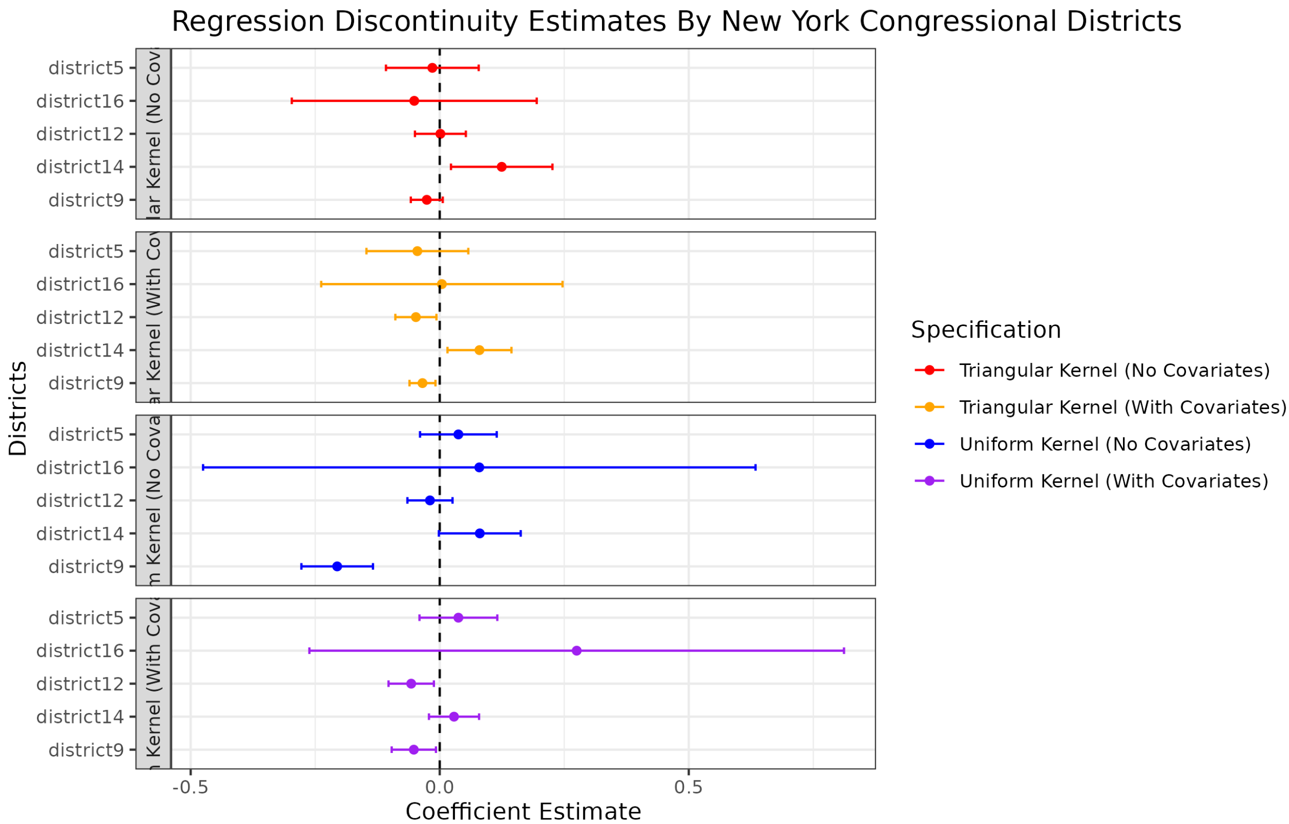
tricts, the Democratic nominees won the primaries with a sizable margin of victory. If voters also care about their pivotality in an election, we might not see a significant effect for all districts with contested primaries, but only for those with relatively competitive primaries.

I re-run the same analysis separately for every treated district in my sample. From an ex-post perspective, the smaller the vote share difference between the winning candidate and the closest competitor in a partisan primary, the more likely that the primary was competitive. For example, according to the vote share data in Table 1, District 9 had the smallest margin of victory, so if individuals strategically register to vote in competitive primaries, the estimates should be the largest in District 9.

Figure 3 reports the RDD results for every treated district under various specifications of kernels and covariates. In each specification, from the top to bottom, I rank the districts in a descending order in terms of the margin of victory in the Democratic primaries in 2018. District 5 has the largest margin (72%), which is likely to be the least competitive primary ex-ante, while District 9 has the smallest margin (6%), which is likely to be the most competitive primary ex-ante. If strategic registration exists, then districts with more competitive Democratic primaries should have larger and more significant estimates. However, the sign and the magnitude of the results in Figure 3 do not align with the ordering of competitiveness.[8] In addition, I run balance checks and create regression discontinuity plots separately for each district in Appendix B, where I also find no evidence of strategic registration.

[8] Although in the specifications with triangular kernel, District 14 shows a significant and positive effect, District 9 shows a significantly negative result, which can not be explained by strategic registraion.

Figure 3: Regression Discontinuity Estimates By Districts Find No Evidence of Strategic Registration



Notes: All estimations use regression discontinuity under local linear regression with bias correction and optimal bandwidth as in Calonico, Cattaneo, and Titiunik (2014). In each specification from top to bottom, the margin of victory in the Democratic primaries in 2018 goes down, indicating more competitive primaries. Covariates include boundary fixed effects, as well as voters' age, gender, race, marriage status, and household size.

Table 5: Treated and Control Districts

State	Treated	Treated PVI	Control	Control PVI
NY	8th	D+36	11th	R+3
	9th	D+34	11th	R+3
	10th	D+26	11th	R+3

3.6 Long-term Dominance

My previous analyses mostly focus on one specific election. In practice, voters may not change their party affiliation just for one election, but instead treat it as a long term decision. For example, if an individual lives in a very democratic district, she might decide to register as a Democrat, not because there is a competitive Democratic primary in a particular year, but because in the long run, if there is ever going to have a competitive election, that election will likely be the Democratic primary.

In this case, my RDD should compare a district that is dominated by a particular party to a battleground district. Although the New York metropolitan area is overall democratic, District 11 is a particular case that leans Republican. District 11 covers Staten Island and parts of Brooklyn. The Brooklyn section of District 11 borders three other districts (8, 9, and 10) that are all dominated by the Democrats (Table 5). For voters who live close to the district boundaries but are in District 8, 9, or 10, they rarely see any competitive general elections, while for voters who are just inside District 11, they elected a Republican representative in 2016 but a Democratic representative in 2018.[9] Therefore, strategic registration could possibly occur in District 8, 9, and 10 but not in District 11.

Similarly, I subset to individuals who live close to the boundaries between 8, 9, 10 (treated districts) and district 11 (control district) in 2018. In my RDD analysis (Table 6), all estimates are small and insignificant, suggesting that individuals do not strategically register with the dominant party as a long term decision. Again, I run balance checks and create regression discontinuity plots in Appendix C, which also show no evidence of strategic registration.

4 Conclusion

In this paper, I offer a framework to analyze the conditions under which strategic registration might occur. Then, I use voter file data from New York in 2018 to empirically

[9]https://en.wikipedia.org/wiki/New_York%27s_11th_congressional_district

Table 6: Regression Discontinuity Design Finds No Significant Difference in Democratic Party Registration Between Heavily Democratic and Battleground Districts in New York in 2018

	(1)	(2)	(3)	(4)
Democratic District	-0.013	0.031	0.024	0.030
	(0.026)	(0.032)	(0.042)	(0.042)
Bandwidth	116.845	110.686	92.208	98.256
N	29,658	28,187	22,946	24,381
Covariates	No	No	Yes	Yes
Kernel	Uniform	Triangular	Uniform	Triangular

Notes: All columns use regression discontinuity under local linear regression with bias correction and optimal bandwidth as in Calonico, Cattaneo, and Titiunik (2014). Robust standard errors are reported in parenthesis. Bandwidths are measured in meters. Column (3) and (4) include boundary fixed effects, as well as voters' age, gender, race, marriage status, and household size as covariates.

test whether strategic registration exists. In particular, by exploiting the spatial regression discontinuities at the congressional district borders where two sides of the border have almost the same environment except for different levels of electoral competition or party dominance, I do not find evidence of voters registering strategically in order to vote in partisan primaries. My results are robust to various settings and specifications, suggesting that individuals might derive significant psychic benefit from truthfully registering with the party that best reflects their political beliefs.

References

Ansolabehere, Stephen et al. (2010). "More democracy: The direct primary and competition in US elections". In: *Studies in American Political Development* 24(2), pp. 190–205.

Arceneaux, Kevin and David W Nickerson (2009). "Who is mobilized to vote? A re-analysis of 11 field experiments". In: *American Journal of Political Science* 53(1), pp. 1–16.

Brown, Jacob (2021). "Partisan Conversion Through Neighborhood Influence: How Voters Adopt the Partisanship of their Neighbors and Reinforce Geographic Polarization". In.

Brown, Jacob and Ryan D Enos (2021). "The measurement of partisan sorting for 180 million voters". In: *Nature Human Behaviour*, pp. 1–11.

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik (2014). "Robust nonparametric confidence intervals for regression-discontinuity designs". In: *Econometrica* 82(6), pp. 2295–2326.

Cantoni, Enrico and Vincent Pons (2020). *Does Context Outweigh Individual Characteristics in Driving Voting Behavior? Evidence from Relocations within the US*. Tech. rep. National Bureau of Economic Research.

Downs, Anthony et al. (1957). "An economic theory of democracy". In.

Enos, Ryan D and Anthony Fowler (2014). "Pivotality and turnout: Evidence from a field experiment in the aftermath of a tied election". In: *Political Science Research and Methods* 2(2), pp. 309–319.

Fiorina, Morris P., Samuel J. Abrams, and Jeremy Pope (2005). *Culture war? The myth of a polarized America.*

Franklin, Mark N et al. (2004). *Voter turnout and the dynamics of electoral competition in established democracies since 1945*. Cambridge University Press.

Gerber, Alan S, Gregory A Huber, and Ebonya Washington (2010). "Party affiliation, partisanship, and political beliefs: A field experiment". In: *American Political Science Review* 104(4), pp. 720–744.

Hirano, Shigeo and James M Snyder Jr (2019). *Primary elections in the United States*. Cambridge University Press.

Kelley, Stanley, Richard E Ayres, and William G Bowen (1967). "Registration and voting: Putting first things first". In: *American Political Science Review* 61(2), pp. 359–379.

Martin, Gregory J and Steven W Webster (2020). "Does residential sorting explain geographic polarization?" In: *Political Science Research and Methods* 8(2), pp. 215–231.

McGhee, Eric et al. (2014). “A primary cause of partisanship? Nomination systems and legislator ideology”. In: *American Journal of Political Science* 58(2), pp. 337–351.

Miller, Warren E (1991). “Party identification, realignment, and party voting: Back to the basics”. In: *American Political Science Review* 85(2), pp. 557–568.

Moskowitz, Daniel J, Benjamin Schneer, et al. (2019). “Reevaluating competition and turnout in US house elections”. In: *Quarterly Journal of Political Science* 14(2), pp. 191–223.

Riker, William H and Peter C Ordeshook (1968). “A Theory of the Calculus of Voting”. In: *American political science review* 62(1), pp. 25–42.

Sussell, Jesse (2013). “New support for the big sort hypothesis: an assessment of partisan geographic sorting in California, 1992–2010”. In: *PS: Political Science & Politics* 46(4), pp. 768–773.

Wattenberg, Martin P (2002). *Where have all the voters gone?* Harvard University Press.

Appendices

A RD Contested

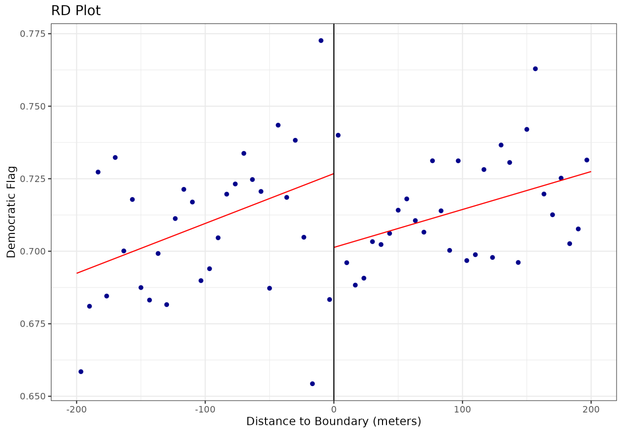
Table 7: Balance Check Between Treated and Control Districts

Variable	Mean Treated	Mean Control	Difference	Estimate	P-Value
Age	48.918	49.594	-0.675	2.558	0.001
Female	0.553	0.547	0.006	0.030	0.121
White	0.350	0.352	-0.002	-0.493	0.000
Black	0.251	0.232	0.019	0.463	0.000
Asian	0.087	0.091	-0.003	-0.010	0.481
Hispanic	0.203	0.215	-0.012	0.023	0.029
Married	0.214	0.211	0.003	0.029	0.025
Household Size	1.717	1.735	-0.018	0.235	0.000

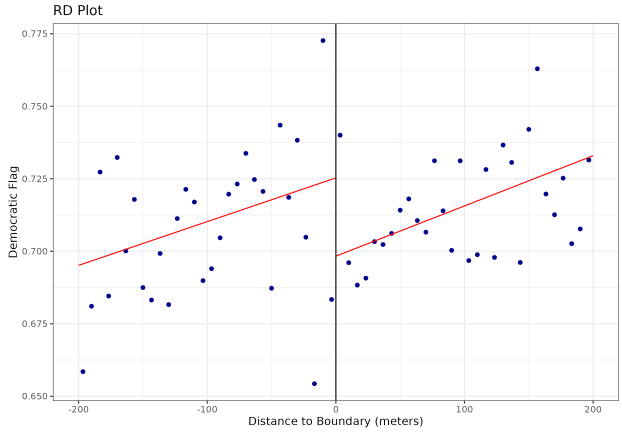
Notes: I compute the mean of each covariates among treated and control voters who live close to the district boundaries (within 200 meters). I report the pvalue of the regression discontinuity estimate for each control variable under local linear regression with bias correction and optimal bandwidth as in Calonico, Cattaneo, and Titiunik (2014). Robust standard errors are reported in parenthesis.

Figure 4: Regression Discontinuity Plots

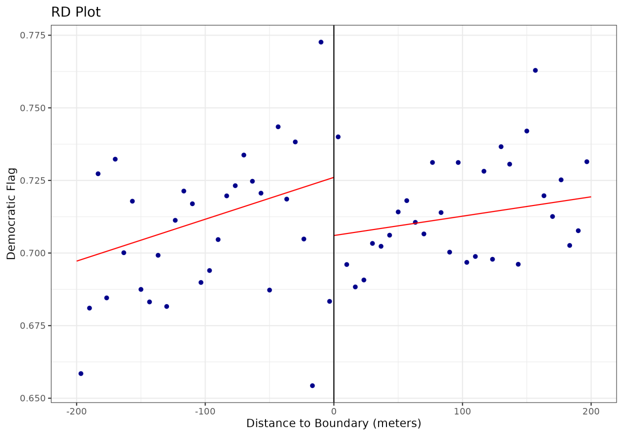
(a) Uniform Kernel No Covariates



(b) Triangular Kernel No Covariates



(c) Uniform Kernel With Covariates



(d) Triangular Kernel With Covariates

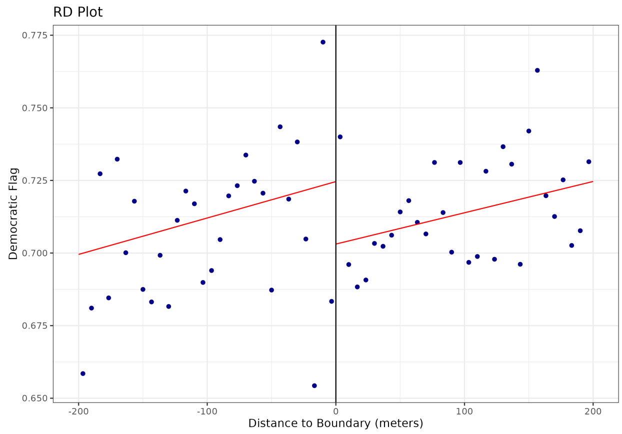
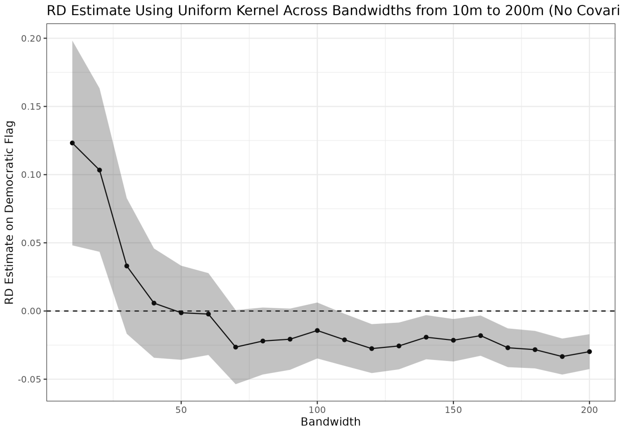
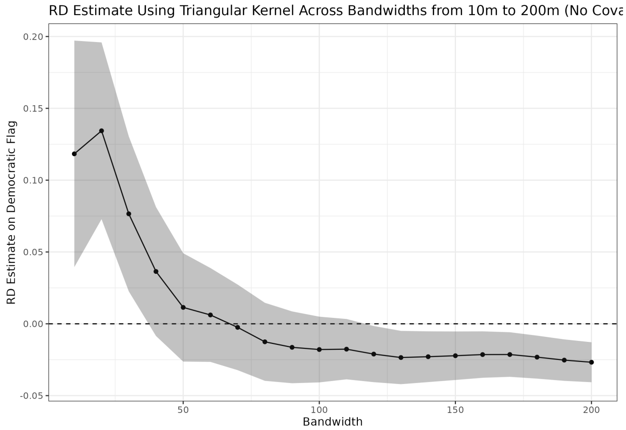


Figure 5: Regression Discontinuity Estimates Across Bandwidths from 10m to 200m

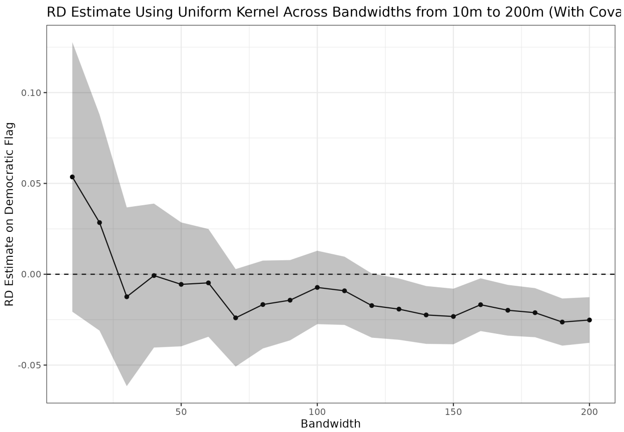
(a) Uniform Kernel No Covariates



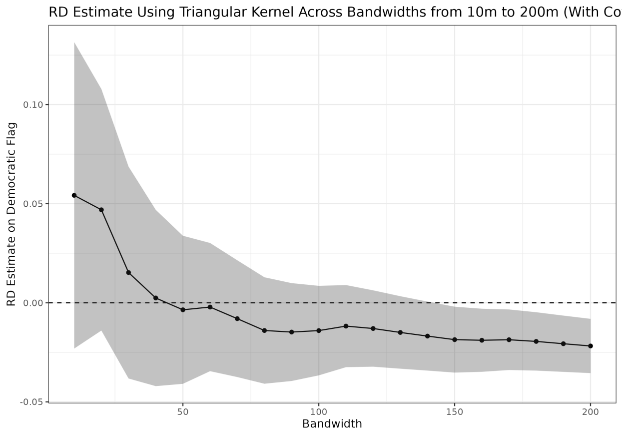
(b) Triangular Kernel No Covariates



(c) Uniform Kernel With Covariates



(d) Triangular Kernel With Covariates



B RD Competitive

B.1 District 5

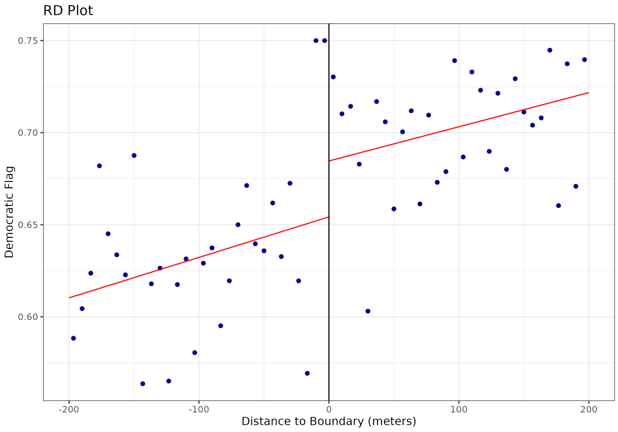
Table 8: Balance Check Between Treated and Control Districts

Variable	Mean Treated	Mean Control	Difference	Estimate	P-Value
Age	48.703	48.495	0.207	6.164	0.011
Female	0.535	0.512	0.023	0.099	0.019
White	0.197	0.261	-0.064	0.222	0.000
Black	0.163	0.045	0.118	0.068	0.000
Asian	0.138	0.154	-0.016	-0.077	0.059
Hispanic	0.290	0.295	-0.005	-0.157	0.000
Married	0.223	0.238	-0.015	0.050	0.126
Household Size	1.808	1.870	-0.062	-0.209	0.031

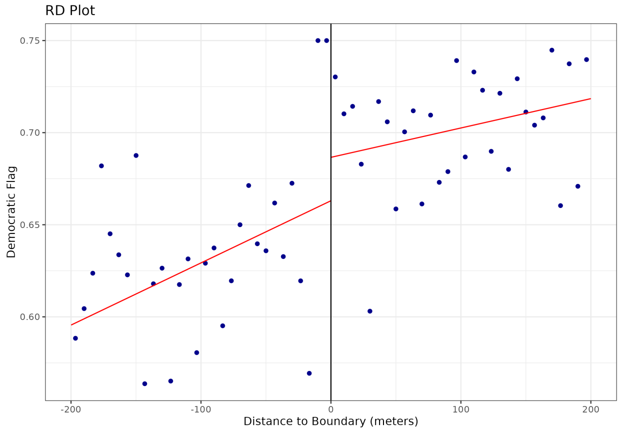
Notes: I compute the mean of each covariates among treated and control voters who live close to the district boundaries (within 200 meters). I report the pvalue of the regression discontinuity estimate under local linear regression with bias correction and optimal bandwidth as in Calonico, Cattaneo, and Titiunik (2014). The outcome variable is each covariate. Robust standard errors are reported in parenthesis.

Figure 6: Regression Discontinuity Plots

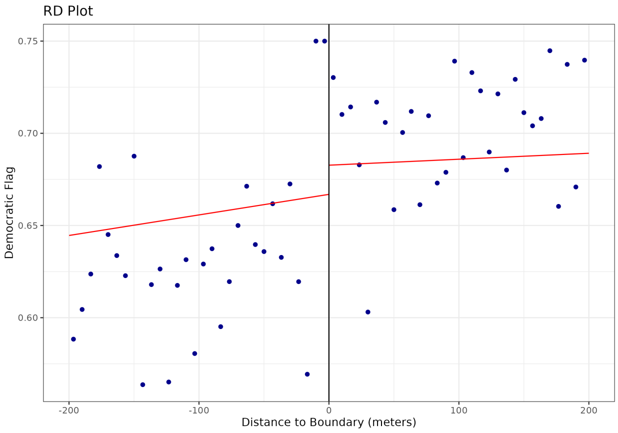
(a) Uniform Kernel No Covariates



(b) Triangular Kernel No Covariates



(c) Uniform Kernel With Covariates



(d) Triangular Kernel With Covariates

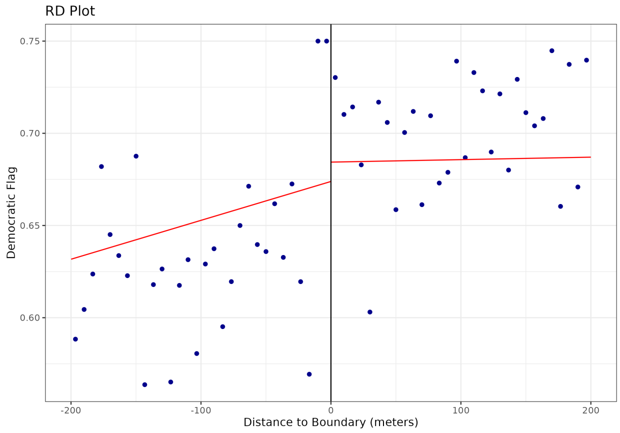
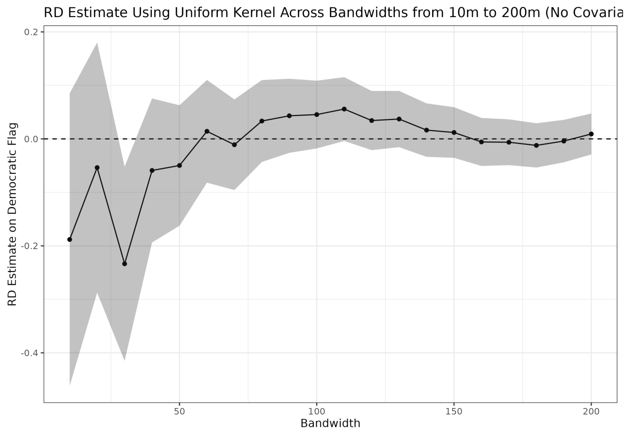
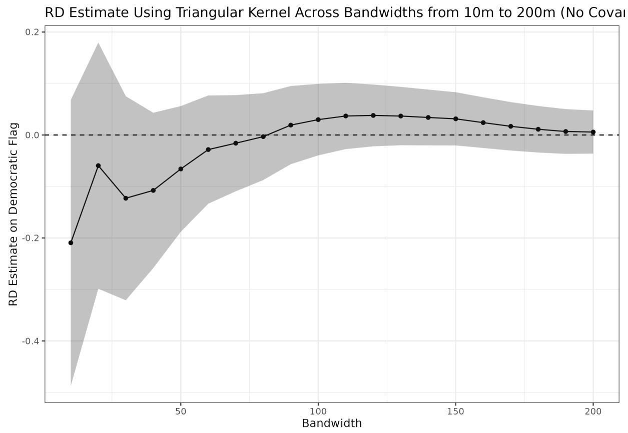


Figure 7: Regression Discontinuity Estimates Across Bandwidths from 10m to 200m

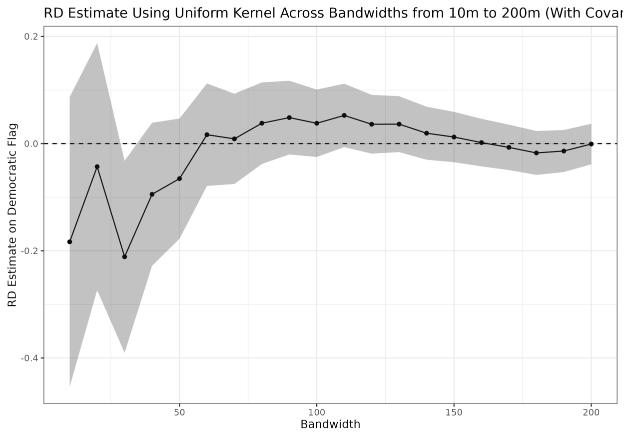
(a) Uniform Kernel No Covariates



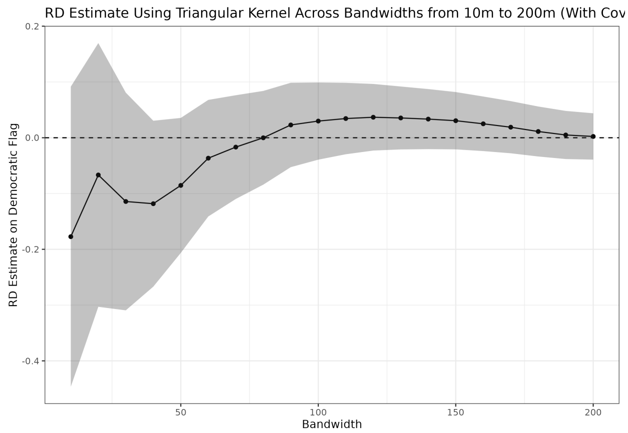
(b) Triangular Kernel No Covariates



(c) Uniform Kernel With Covariates



(d) Triangular Kernel With Covariates



B.2 District 9

Table 9: Balance Check Between Treated and Control Districts

Variable	Mean Treated	Mean Control	Difference	Estimate	P-Value
Age	48.681	49.294	-0.613	1.481	0.189
Female	0.575	0.569	0.007	-0.018	0.535
White	0.357	0.320	0.037	0.028	0.479
Black	0.387	0.416	-0.028	-0.177	0.000
Asian	0.043	0.041	0.001	0.015	0.080
Hispanic	0.117	0.121	-0.004	0.039	0.145
Married	0.225	0.219	0.005	0.030	0.018
Household Size	1.784	1.782	0.001	0.110	0.008

Figure 8: Regression Discontinuity Plots

(a) Uniform Kernel No Covariates

(b) Triangular Kernel No Covariates

(c) Uniform Kernel With Covariates

(d) Triangular Kernel With Covariates

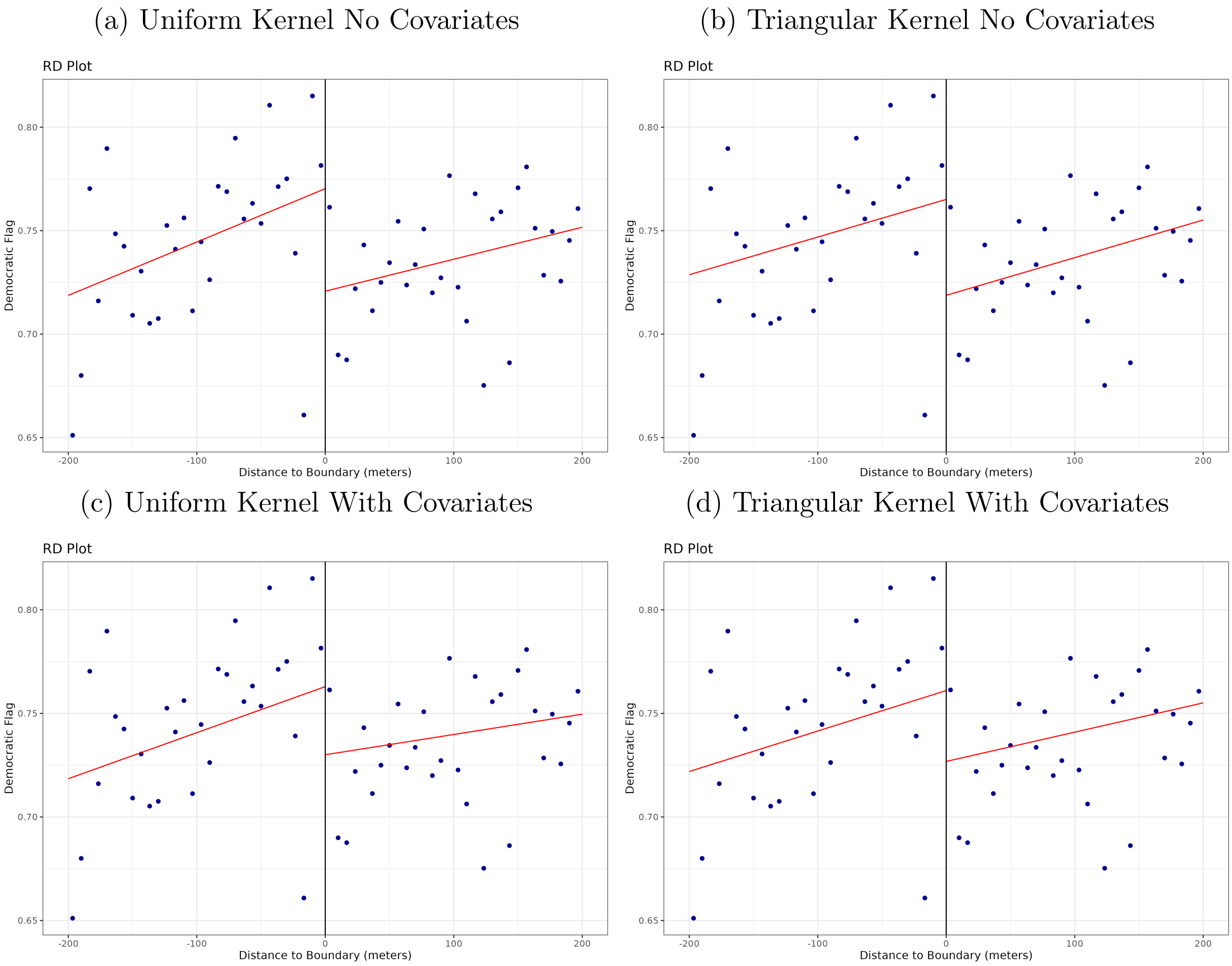
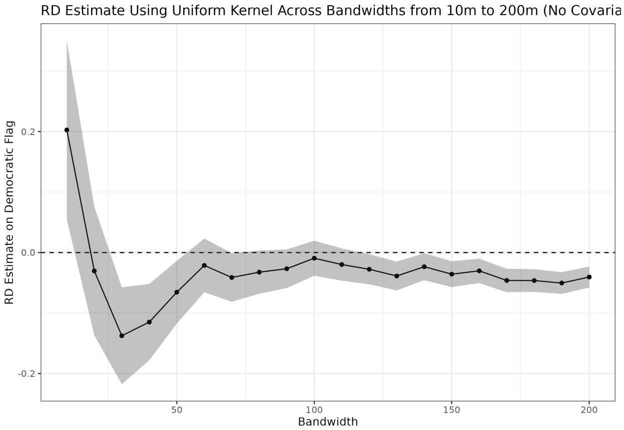
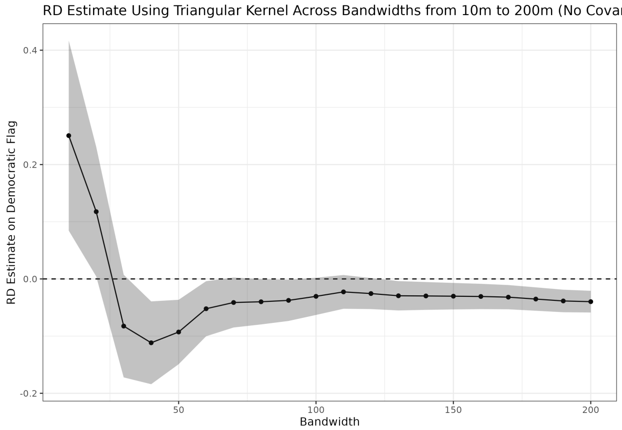


Figure 9: Regression Discontinuity Estimates Across Bandwidths from 10m to 200m

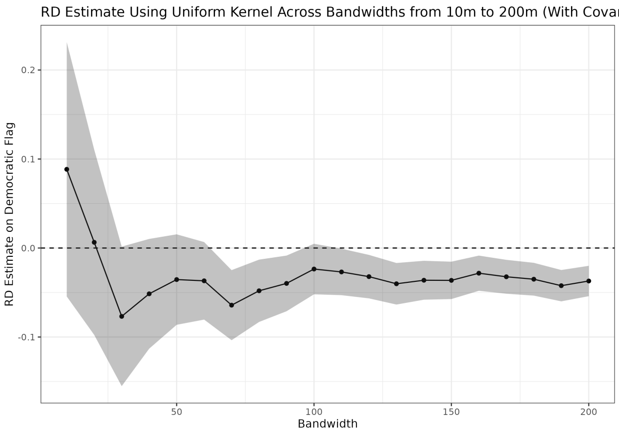
(a) Uniform Kernel No Covariates



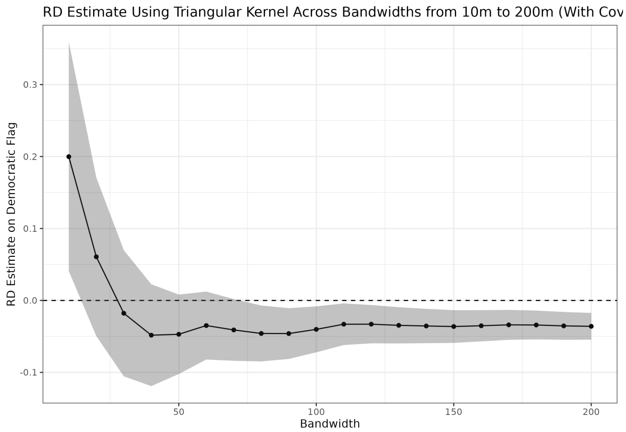
(b) Triangular Kernel No Covariates



(c) Uniform Kernel With Covariates



(d) Triangular Kernel With Covariates



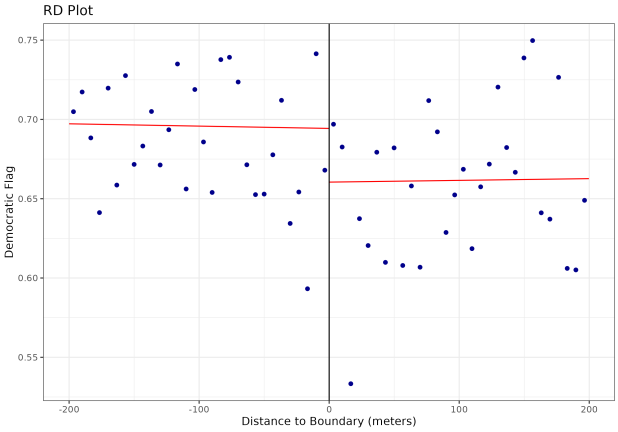
B.3 District 12

Table 10: Balance Check Between Treated and Control Districts

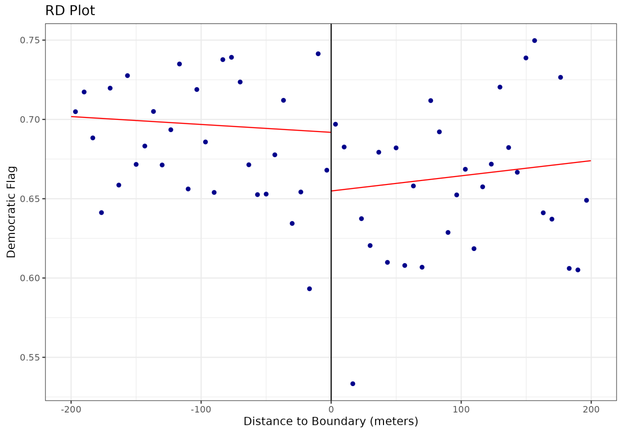
Variable	Mean Treated	Mean Control	Difference	Estimate	P-Value
Age	49.485	51.410	-1.925	13.178	0.000
Female	0.511	0.503	0.008	-0.030	0.295
White	0.713	0.709	0.003	-0.312	0.001
Black	0.024	0.013	0.010	0.044	0.000
Asian	0.092	0.086	0.006	-0.057	0.031
Hispanic	0.114	0.138	-0.024	0.032	0.245
Married	0.206	0.200	0.006	0.003	0.921
Household Size	1.465	1.514	-0.048	-0.191	0.000

Figure 10: Regression Discontinuity Plots

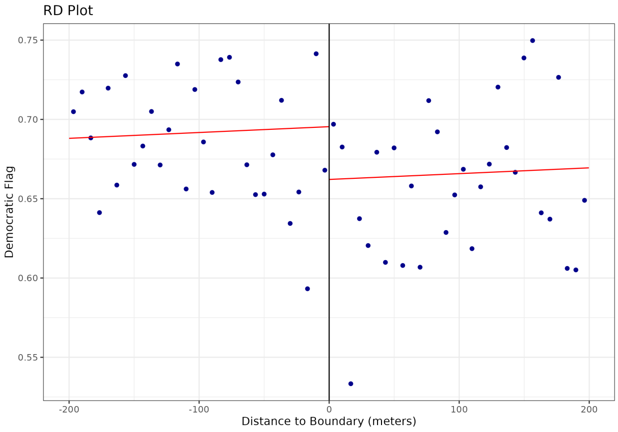
(a) Uniform Kernel No Covariates



(b) Triangular Kernel No Covariates



(c) Uniform Kernel With Covariates



(d) Triangular Kernel With Covariates

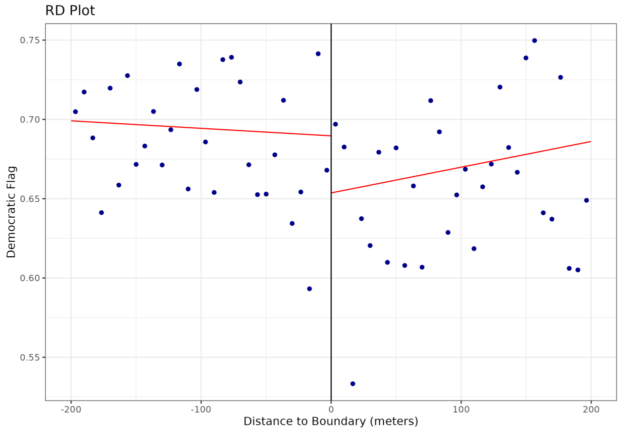
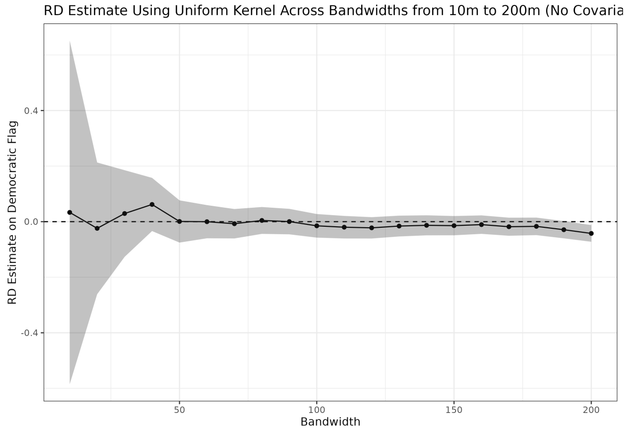
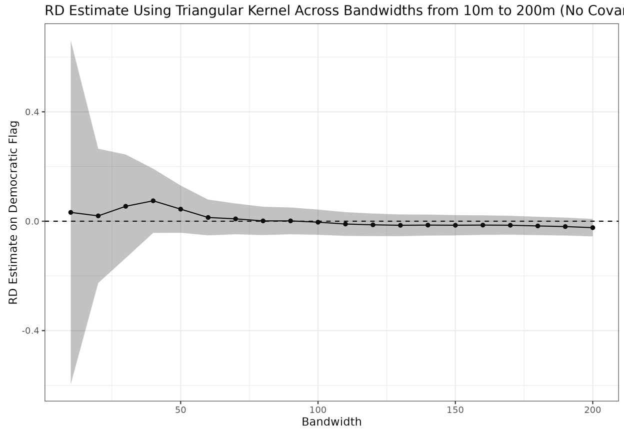


Figure 11: Regression Discontinuity Estimates Across Bandwidths from 10m to 200m

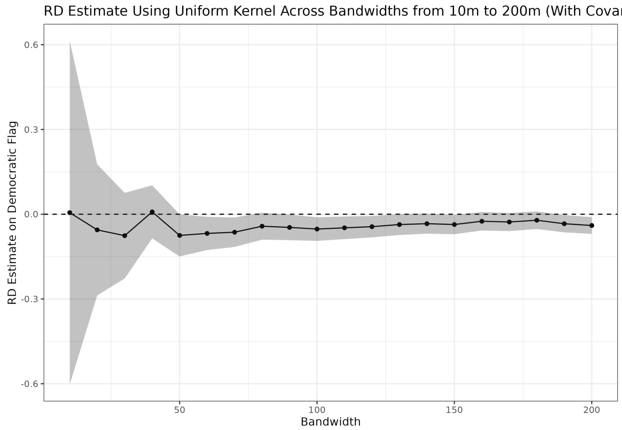
(a) Uniform Kernel No Covariates



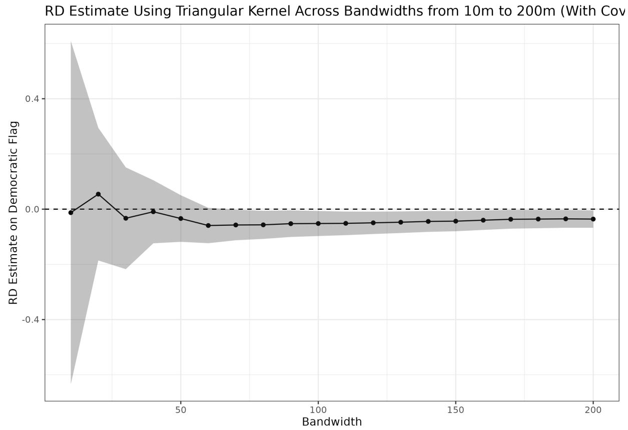
(b) Triangular Kernel No Covariates



(c) Uniform Kernel With Covariates



(d) Triangular Kernel With Covariates



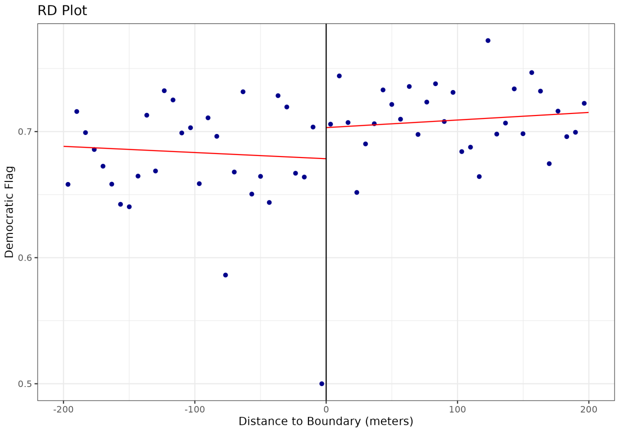
B.4 District 14

Table 11: Balance Check Between Treated and Control Districts

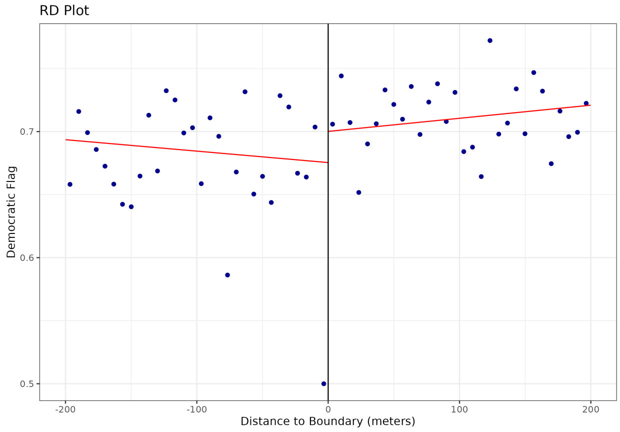
Variable	Mean Treated	Mean Control	Difference	Estimate	P-Value
Age	48.893	49.479	-0.586	0.622	0.726
Female	0.540	0.551	-0.011	-0.018	0.735
White	0.127	0.176	-0.049	0.041	0.233
Black	0.153	0.082	0.071	0.033	0.419
Asian	0.171	0.183	-0.012	-0.424	0.001
Hispanic	0.446	0.464	-0.018	0.007	0.903
Married	0.181	0.181	-0.001	0.056	0.140
Household Size	1.669	1.718	-0.049	0.074	0.251

Figure 12: Regression Discontinuity Plots

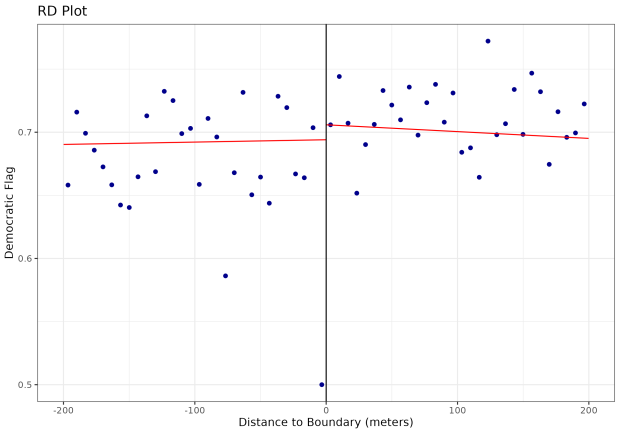
(a) Uniform Kernel No Covariates



(b) Triangular Kernel No Covariates



(c) Uniform Kernel With Covariates



(d) Triangular Kernel With Covariates

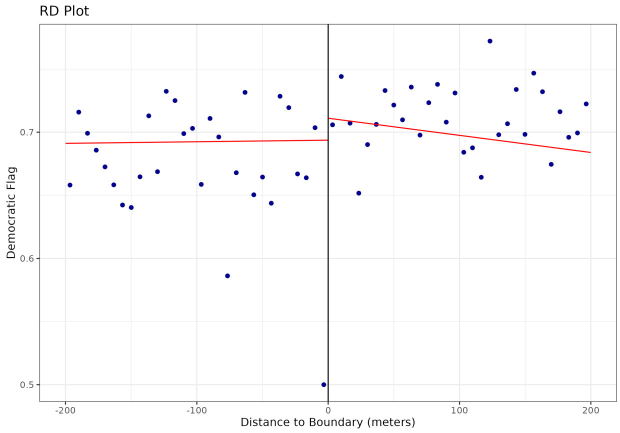
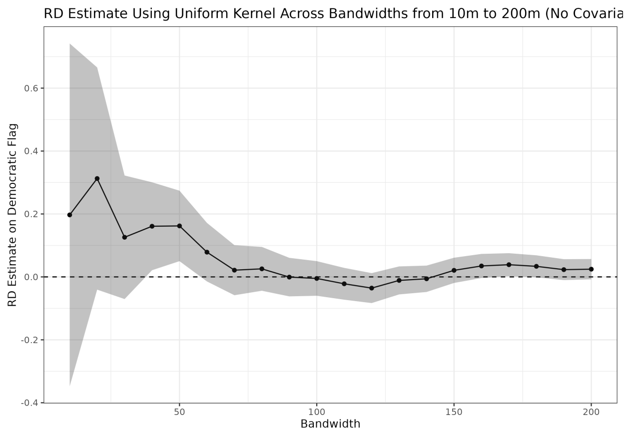
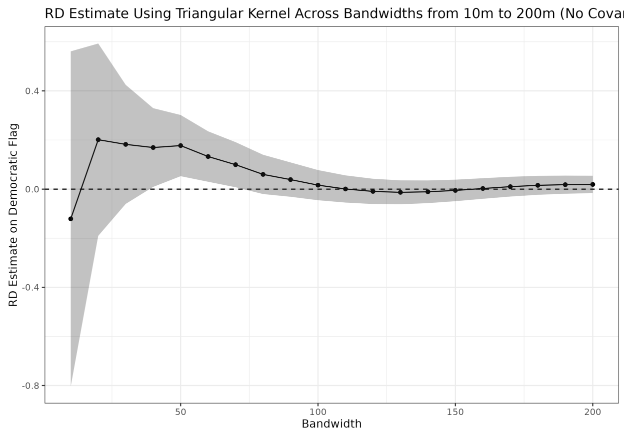


Figure 13: RD Finds No Significant Effect Across Bandwidths

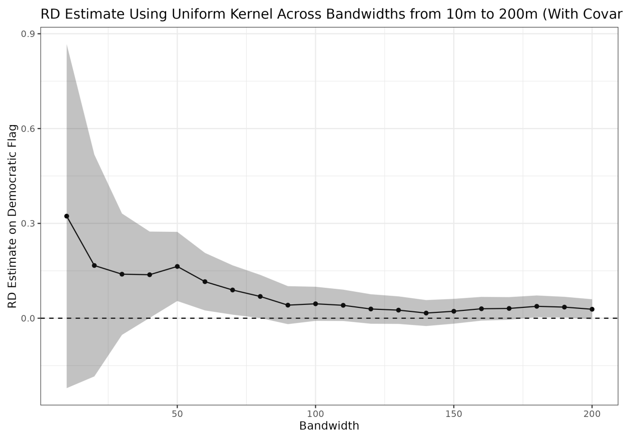
(a) Uniform Kernel No Covariates



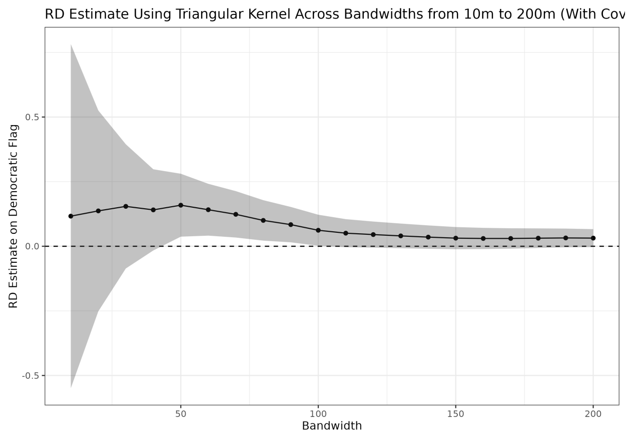
(b) Triangular Kernel No Covariates



(c) Uniform Kernel With Covariates



(d) Triangular Kernel With Covariates



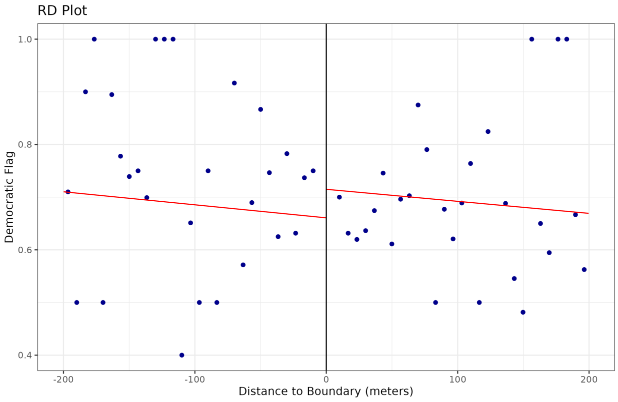
B.5 District 16

Table 12: Balance Check Between Treated and Control Districts

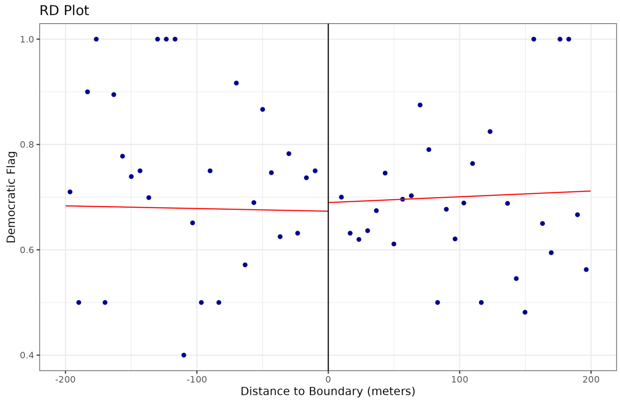
Variable	Mean Treated	Mean Control	Difference	Estimate	P-Value
Age	53.433	50.324	3.109	1.451	0.916
Female	0.579	0.564	0.015	-0.089	0.570
White	0.452	0.401	0.051	-0.303	0.011
Black	0.029	0.035	-0.006	0.065	0.138
Asian	0.072	0.059	0.012	0.183	0.007
Hispanic	0.353	0.460	-0.107	0.030	0.829
Married	0.229	0.227	0.002	-0.113	0.713
Household Size	1.716	1.687	0.029	2.291	0.001

Figure 14: Regression Discontinuity Plots

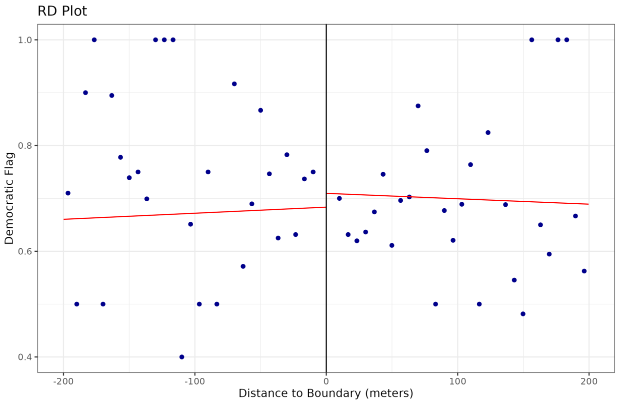
(a) Uniform Kernel No Covariates



(b) Triangular Kernel No Covariates



(c) Uniform Kernel With Covariates



(d) Triangular Kernel With Covariates

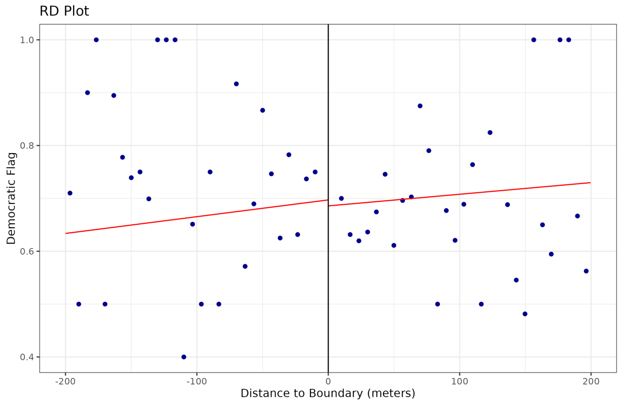
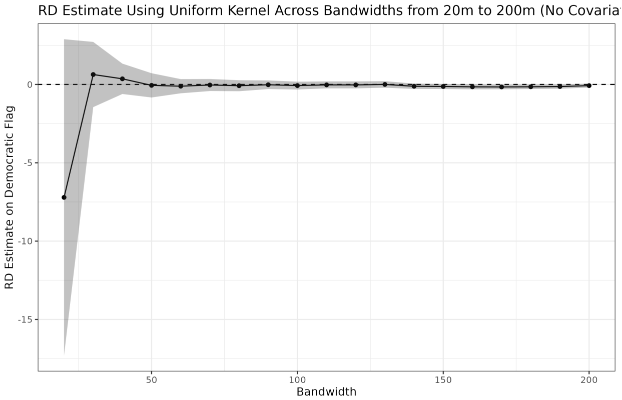
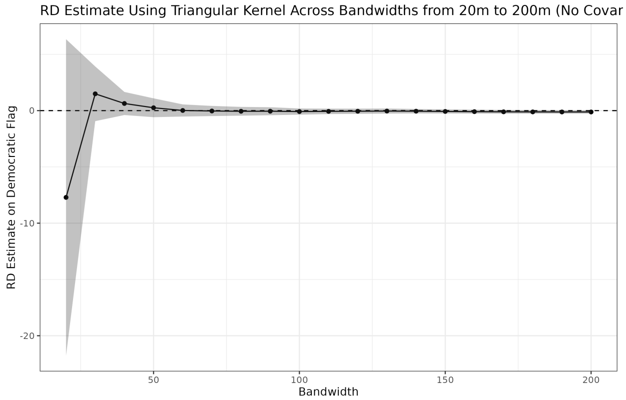


Figure 15: Regression Discontinuity Estimates Across Bandwidths from 20m to 200m

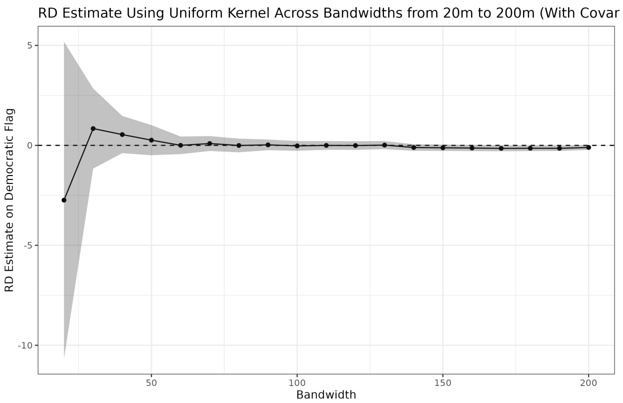
(a) Uniform Kernel No Covariates



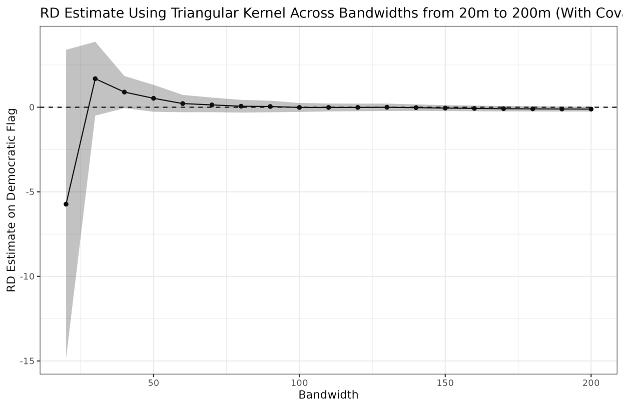
(b) Triangular Kernel No Covariates



(c) Uniform Kernel With Covariates



(d) Triangular Kernel With Covariates



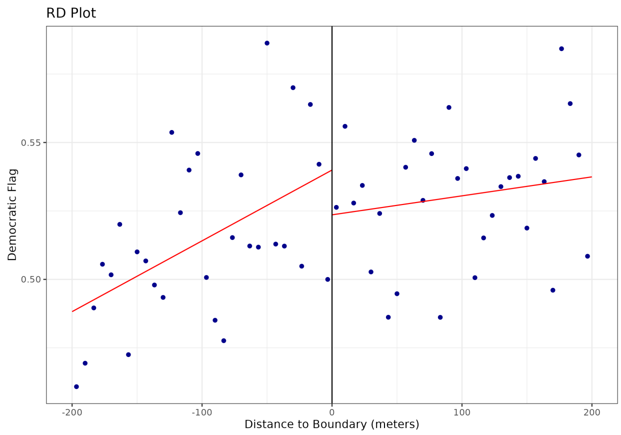
C RD Long Term

Table 13: Balance Check Between Treated and Control Districts

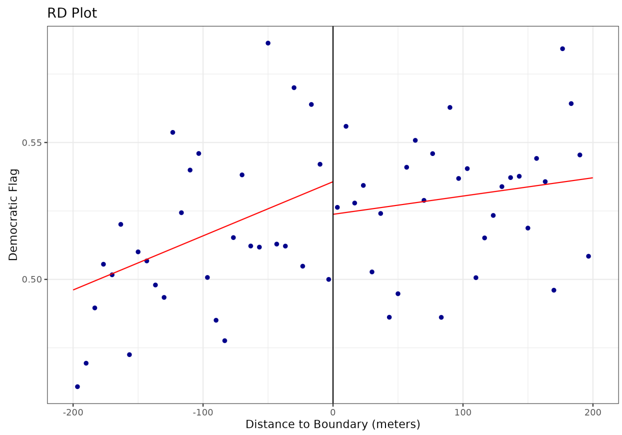
Variable	Mean Treated	Mean Control	Difference	Estimate	P-Value
Age	53.433	50.324	3.109	1.451	0.916
Female	0.579	0.564	0.015	-0.089	0.570
White	0.452	0.401	0.051	-0.303	0.011
Black	0.029	0.035	-0.006	0.065	0.138
Asian	0.072	0.059	0.012	0.183	0.007
Hispanic	0.353	0.460	-0.107	0.030	0.829
Married	0.229	0.227	0.002	-0.113	0.713
Household Size	1.716	1.687	0.029	2.291	0.001

Figure 16: Regression Discontinuity Plots

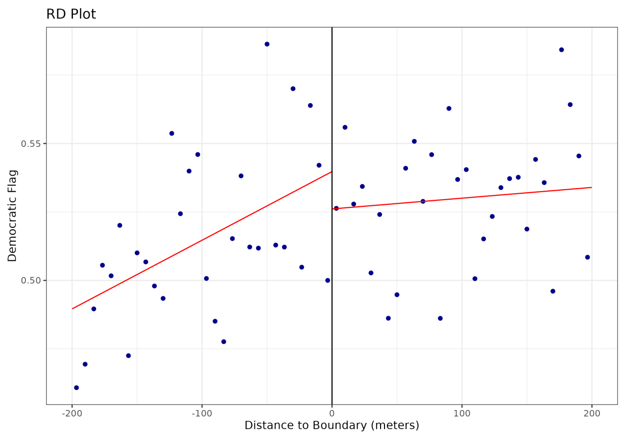
(a) Uniform Kernel No Covariates



(b) Triangular Kernel No Covariates



(c) Uniform Kernel With Covariates



(d) Triangular Kernel With Covariates

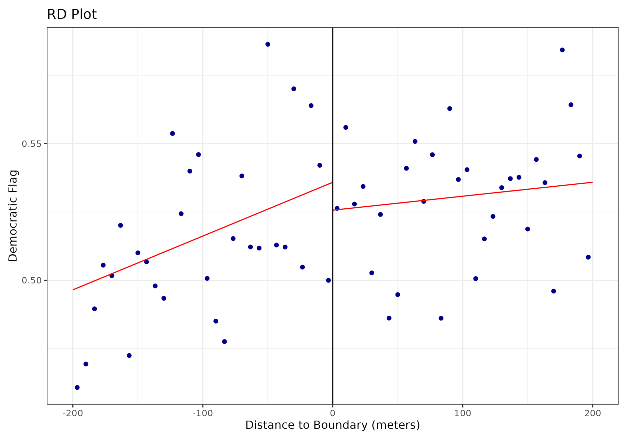
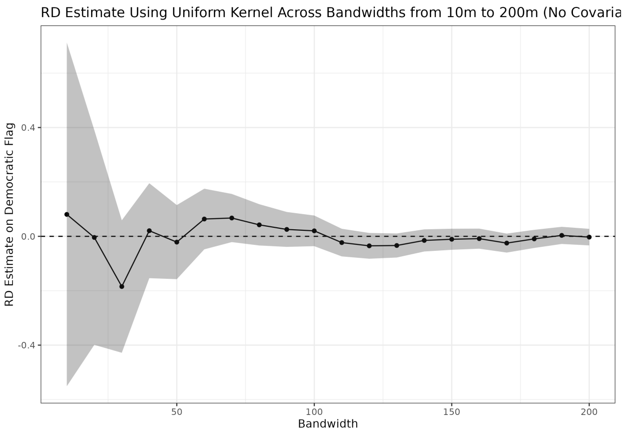
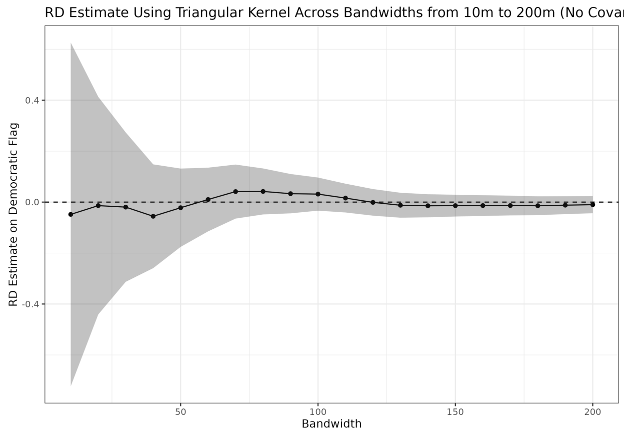


Figure 17: Regression Discontinuity Estimates Across Bandwidths from 10m to 200m

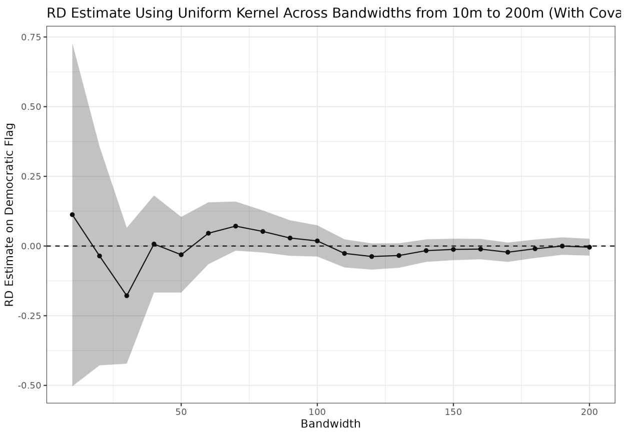
(a) Uniform Kernel No Covariates



(b) Triangular Kernel No Covariates



(c) Uniform Kernel With Covariates



(d) Triangular Kernel With Covariates