Polling Place Location and the Costs of Voting *

Gaurav Bagwe† Juan Margitic † Allison Stashko‡

November 22, 2023

Abstract

We study how distance to one's polling place affects the likelihood of voting using a geographic regression discontinuity design with data from Pennsylvania and Georgia. A one-mile increase in distance to polling place reduces the likelihood of voting in person by 2 to 3 percentage points. Effects are two to four times higher among those closest to the polling place. When available, voters substitute to mail-in voting as distance to polling place increases. We estimate a discrete-choice voting model to identify turnout-maximizing polling places. Some precincts have large potential gains in turnout, even when choosing from existing buildings.

JEL Classification: D72, H70, K16

Keywords: Polling Places, Election Precincts, Costs of Voting

*We would like to thank Laurent Bouton, Garance Genicot, Arik Levinson, Martin Ravallion, Carolina Concha Arriagada, J.J. Naddeo, Arturo Romero Yáñez, Daniel Valderrama, and Timothy Armstrong. We also thank participants at the American Political Science Association Annual Meeting, Southern Economic Association Conference, and the University of Utah Political Science Department. Thank you to Alvin Tsang and Sophia Bagley for their excellent research assistance.

†Bagwe and Margitic were doctoral candidates at Georgetown University while working on this project.

‡Emory University, e-mail: allison.stashko@emory.edu.

1 Introduction

Since the 2020 U.S. presidential elections, debates surrounding election administration have intensified. In 2021 alone, 36 states enacted a total of 96 new election laws.[1] Many of these laws shift the costs and benefits of voting in person or voting by mail. While some states made mail-in ballots available to all voters, others introduced new requirements for voting by mail. Some states expanded the number of early voting days, while others reduced the number of early voting days or removed polling places altogether. It is important to know how voters respond to changes in the costs of voting to assess the impact of these laws ex-ante. This is especially true in the U.S., where voter participation is low and election law has become increasingly partisan in nature (Bentele and O'Brien 2013; Burden et al. 2017; Hasen 2012).

State and local government decisions regarding mail-in voting (Meredith and Endter 2016; Meredith and Malhotra 2011; Lockhart et al. 2020; Thompson et al. 2020), early voting, (Kaplan and Yuan 2020), and at-poll voting requirements (Highton 2017; Cantoni and Pons 2022) can each potentially have important consequences for voter participation. Even small changes to the convenience or cost of voting can determine whether or not someone votes, especially in large elections (Gomez et al. 2007; Braconnier et al. 2017).

This paper focuses on a particular cost of voting: the distance from a voter's home to their polling place. Distance to the polling place is an important determinant of voting behavior to study for three reasons. First, this cost of voting cannot be eliminated, unless we remove polling places altogether. The alternative to voting at a polling place is to vote by mail, access to which varies from state to state. Three-quarters of voters chose to vote at polls on election day in the 2018.[2] Despite the COVID-19 pandemic, 54% of voters cast their ballot in person in the 2020 election cycle.[3] Second, the distance to polling place is inherently uneven among the population of eligible voters. It is important to understand if certain populations are systematically disadvantaged by higher costs of voting, in which case politicians may ignore their interests (Avery 2015; Martin 2003). Given the history of voter suppression in the United States, it is especially important to understand if the distance to polling place, and its effect on turnout, varies by race of eligible voters. Third, the polling place location is a policy choice, in contrast to other determinants of turnout like candidate characteristics or weather. State and local lawmakers determine how to divide a state

[1] Brennan Center for Justice. Voting Laws Round-up: December 2021. Available at: https://www.brennancenter.org/our-work/research-reports/voting-laws-roundup-december-2021

[2] Election Administration and Voting Survey: 2018 Comprehensive Report. https://www.eac.gov/sites/default/files/eac_assets/1/6/2018_EAVS_Report.pdf Retrieved November 11, 2020.

[3] Pew Research Center. The Voting Experience in 2020. November 20, 2020. Available at: https://www.pewresearch.org/politics/2020/11/20/the-voting-experience-in-2020/

into voting precincts and where to locate a polling places within precincts. A precise understanding of how polling place locations affect voter participation is critical for determining the optimal allocation of polling places.

Existing estimates of the effect of distance to polling place on voting vary widely.[4] Evidence from panel data in North Carolina suggests that distance to polling place has a null effect on turnout (Clinton et al. 2020, Yoder 2018).[5] Using matching methods to study polling place consolidations in Los Angeles County, Brady and McNulty (2011) find that a one-mile increase in distance to the polling place reduces turnout by 1 percentage point.[6]

At the other extreme, Cantoni (2020) estimates that a one-mile increase in distance to polling place reduces turnout by 4 to 12 percentage points in a study of nine urban municipalities in Massachusetts and Minnesota. These findings imply that moving a polling place one mile closer to a voter would cause an increase in turnout on par with some of the most successful turnout mobilization tactics (Gerber et al. 2017, Green et al. 2013, Enos and Fowler 2018). Cantoni (2020) identifies the causal effect of distance to polling place on turnout by exploiting variation in distance to polling place among voters near election precinct borders. Intuitively, registered voters who live close to an election precinct border are likely similar in terms of potential confounding factors but are assigned to different polling places. Tomkins et al. (2023) similarly use precinct borders as a source of variation. They estimate smaller effects: a mile increase in distance to polling place reduces turnout by 1.55 percentage points, on average, across 10 states.

The varied findings in the existing literature underscore the need for further study of polling places in the United States and their effect on turnout. One possibility is that the range of results reflects heterogeneous effects, given that each study uses data from different locations. Moreover, we know that there are large location and

[4]We focus on recent studies that use causal identification strategies. Polling places are typically located in schools and other government buildings that tend to be centrally located within a town. It is likely that eligible voters who live closer to polling places differ systematically from those who live further away in ways that matter for turnout. Earlier observational studies on the distance to polling place have found either a negative association with turnout or null results. Dyck and Gimpel (2005) find that distance to polling place is associated with lower turnout and more mail-in voting in Clark County, Nevada. Haspel and Gibbs Knotts (2005) find a negative relationship between distance to polling place and the likelihood of voting in Atlanta, Georgia. Amos et al. (2017) find a negative association between distance to polling place and voting in person on election day, but this is offset by a positive association between distance to polling place and voting early or absentee. Their study uses data from Manatee County, Florida.

[5]Alipour and Lindlacher (2023) find similar results by studying re-precincting in Munich.

[6]From Figure 3 of Brady and McNulty (2011), a change in polling place of one mile causes an estimated 4% reduction in the likelihood of voting. They estimate that most of this effect (-1.8%) is attributable to search costs of finding a new polling place, so the remaining -2.2% is attributable to distance to polling place. Given the reported turnout rate of 55.1% in their study, a 2.2% reduction in the likelihood of voting is equivalent to 1.2 percentage points.

election-specific effects on voting behavior (Cantoni and Pons 2022). However, there is also little overlap in methodologies used across existing studies, making comparisons difficult.

To help fill this gap, we collected information about the distance to polling place and turnout for over 15 million voters in Pennsylvania and Georgia. We chose two large swing states in order to estimate effects in urban and rural areas and among socio-demographic groups of interest. Regulations for accessing mail-in ballots also differ significantly between the two states, with Georgia providing easy access while Pennsylvania imposing more restrictions. For each registered voter in these two states, we observe the location of their residence, the location of their polling place, and whether or not they abstain, vote in person, or vote by mail in the 2018 election. We compute both the travel-route distance between a voter and their polling place and the time to travel to the polling place.[7]

To estimate the causal effect of distance to polling place on voting behavior, we use a geographic Regression Discontinuity (RD) design. Building on the innovation of Cantoni (2020), we use election precinct boundaries as a source of exogenous treatment assignment. We use a one-dimensional running variable that measures the distance from a voter to a single point on the nearest precinct boundary, which we call an RD point (following Keele and Titiunik (2015) and Imbens and Zajonc (2023)). We observe 9,722 RD points across the two states. The treatment variable is the distance from a voter to their polling place. The continuous treatment variable in our setup introduces additional complexity compared to existing applications of geographic RDs, which have binary treatment variables (Keele and Titiunik 2015, Dell 2010).

First, we cannot express a voter's distance to polling place (the treatment variable) solely as a function of their distance to the border (the running variable). This is because two voters in the same precinct can be equidistant from a point on a precinct border, but differ in distance to the polling place. Consequently, our study implements a fuzzy RD framework rather than a sharp RD. In the first stage, we estimate the change in distance to polling place at the discontinuity. In the second stage, we estimate the change in the likelihood that a registered voter votes in person or by mail.

Second, the standard monotonicity condition for identification would not be satisfied in our setup without the adjustments we make to the sample. Monotonicity requires that distance to polling place increases for all voters when moving from control to treatment precincts (Card et al. 2015). However, due to the relatively small size of the precincts, distance to polling place increases for some voters and decreases

[7] Existing papers, in contrast, focus on Euclidean distance to measure the cost of voting. Because we study large states, Euclidean distance may systematically differ from travel-route distance in urban versus rural areas.

for others near the discontinuity.[8] We resolve this issue by carefully constructing a running variable and RD sample in order to satisfy the monotonicity condition. These methodological considerations are useful for future research that uses geographic RD to study the effects of a continuous treatment variable.

Our methodology has several advantages over existing approaches. The most closely related approach to ours is the main specification of Cantoni (2020). Cantoni estimates treatment effects by regressing voting outcomes on distance to polling place, including fixed effects for a small geographic area that contains a precinct border (a buffer zone of 0.05 miles around a segment of a border). We call this the Border Fixed Effects (FE) approach. Tomkins et al. (2023) similarly estimate a Border FE model, but they include fixed effects for neighborhood blocks.[9] An important advantage of the RD approach is that treatment effects are estimated using only *across* precinct variation in distance to polling place. The Border FE approach uses both across- and within-precinct variation. Second, the non-parametric RD method allows for unrestricted treatment effect heterogeneity, including non-linear effects, which we show is important in this setting. Third, one can use data-driven methods to select voters within a bandwidth around a discontinuity, whereas the Border FE requires ad-hoc decisions about how to define a geographic fixed effect. Finally, identifying assumptions for the RD estimator are perhaps more plausible. The RD requires continuity of confounding factors at the discontinuity, while the Border FE specification requires that distance to polling place is uncorrelated with confounding factors, after controlling for border fixed effects.

Relative to panel data methods, the RD approach allows us to isolate the effect of distance to polling place, separate from the effects of changes to polling places locations. It is difficult to estimate how distance to polling place affects voting from changes in polling places due to search costs (Alipour and Lindlacher 2023, Clinton et al. 2020, Yoder 2018, Brady and McNulty 2011,) and habit formation in voting (Fujiwara et al. 2016; Plutzer 2002). A voter's initial distance to polling place might still affect their voting decisions after a polling place changes, at least in the short run.[10] The RD

[8] This would not be an issue, for example, if one were using a fuzzy RD design to estimate the effect of distance between individuals and a location which is very far from the border. The problem here stems from the fact that precincts are relatively small, so the polling places can be located near the discontinuity.

[9] Tomkins et al. (2022) define a block as a group of voters that have similar addresses to each other, differing only in the last two digits of the building number. They only include blocks in which voters are assigned to two unique polling places. There are relatively few blocks split across a precinct border, so they pool across ten states. Estimates from pooled data across multiple states can be difficult to interpret, however, since states differ in rules about voting by mail.

[10] For example, one might under-estimate the magnitude of a negative effect of distance to polling place on turnout if variation in distance to polling place comes mainly from precinct consolidations. Voters would experience an increase in distance to polling place, but might not be sensitive to this change due to relatively low costs of voting in the recent past.

approach, which relies only on cross-sectional data, is useful for policy-makers who are interested in the long-run effect of a polling place location on turnout, after short-run effects of a change in polling place dissipate.

We estimate that there is a small negative effect of distance to polling place on the likelihood of voting in person, on average, ranging from -2.7 p.p. to -1.25 p.p. per mile in Pennsylvania and Georgia, respectively. Importantly, the average effects mask substantial heterogeneity. In particular, voters close to the polling place are more sensitive to an increase in distance to polling place than voters that live further away. In Pennsylvania, a mile increase in distance to polling place causes a 7.5 p.p. decline in the likelihood of voting in person among voters in the bottom quartile of distance to polling place, who are 0.3 miles from the polling place, on average. In Georgia, voters in the bottom quartile, who are 0.7 miles from the polling place, are 6.8 p.p. less likely to vote in person per mile increase in distance to polling place. We estimate precise null effects of distance to polling place for voters in the top quartile of distance to polling place, who are more than a mile from their polling place. Taking the non-linear effects into account, a mile increase in distance to polling place reduces the likelihood of voting in person by 3.3 p.p. to 1.8 p.p., and a standard deviation in distance to polling place reduces the likelihood of voting in person by 0.6 p.p. and 0.5 p.p. in Pennsylvania and Georgia, respectively.

Pennsylvania and Georgia voters differ in their substitution to mail-in voting. In Pennsylvania, where voters were required to provide a justification for their absence, there is no significant substitution to mail-in voting as distance to polling place increases. In contrast, in Georgia, where obtaining absentee ballots requires no justification, there is a significant increase in the likelihood of voting by mail as the distance to the polling place increases. As a result, there is a statistically significant negative effect of distance to polling place on the overall likelihood of voting in Pennsylvania, and a precise null effect in Georgia.

To reconcile our findings with previous studies, we use the methodologies from the literature, including the Border FE specification, a matching specification, and a difference-in-differences approach with a panel of Pennsylvania voters. By comparing results, we conclude that heterogeneous effects explain the discrepancies in estimated effect sizes across methodologies. Estimators that impose constant marginal effects of distance to polling place on the likelihood of voting will underestimate average effects. Location-specific factors also likely play a role. For example, we find larger effects in Philadelphia, on the same order of magnitude as those found in Boston and Minneapolis by Cantoni 2020. We then turn to heterogeneous effects analyses to better understand what factors are associated with greater sensitivity to polling place locations.

The analysis of the two large states allows us to explore heterogeneous effects by demographic characteristics, party affiliation, economic variables, and transportation. We find little evidence that sensitivity to distance to polling place is associated with age, sex, race/ethnicity, or political party affiliation. However, we find suggestive evidence that income and educational attainment may moderate substitution to mail-in voting. In particular, there is relatively lower substitution to voting by mail in low-income and low-education areas, as distance to polling place increases. We also find suggestive evidence that the non-linear effects are reflect differences in modes of transportation.

The estimated effects of distance to polling place on voting behavior tell us how sensitive voters are to a change in the distance to their polling place. But how important are these effects in aggregate? How would turnout respond to realistic changes in polling places? To answer these questions, we study counterfactual allocations of polling places in Pennsylvania, the only state with an estimated negative effect of distance to polling place on turnout. We numerically solve a planner's problem to maximize turnout in an existing voting precinct by choosing the location of a polling place, based on the locations of all voters as well as the estimated sensitivity of turnout to changes in distance to polling place. We find the turnout-maximizing latitude and longitude of a polling place place, and also the turnout-maximizing public building location. The latter approach takes into account existing infrastructure constraints. We find that the polling places used in 2018 in Pennsylvania tend to be located near the turnout-maximizing polling place locations in their respective precincts. Using the turnout-maximizing existing buildings, turnout would increase by 0.4 p.p. (0.5%), in aggregate. This is on par with the effects of one additional day of early voting (Kaplan and Yuan 2020) or mailings that encourage people to vote (Green et al. 2013). If each precinct were to double the number of polling places, using existing buildings, then turnout would increase by 2.6 p.p. (3.3%), which is similar to the effects of voter mobilization methods like canvassing. The counterfactual simulations can also help to find precincts with the largest potential gains to improving polling place location. Among the 92 precincts in the top 1% of gains to turnout rates, using the optimal building as the polling place would increase turnout by an average of 11.1 p.p.. This is on par with more successful voter mobilization tactics, including social pressure (Gerber et al. 2017), canvassing (Green et al. 2013), and large scale advertising campaigns (Enos and Fowler 2018).

Our findings highlight some important lessons for studies of polling places in the future. First, the importance of polling place locations for voter participation will vary depending on the context. If the affected population largely owns cars and drives to work, then a one mile increase in distance to a polling place will likely have a small effect. On the other hand, a small change to a polling place in an area where people do not tend to drive can have significant effects on turnout. While polling closures in rural areas typically draw media attention,[11] a change to polling location in an urban area might have an even larger effect on turnout and should be evaluated carefully by election commissions.

Electoral design is also likely important. With two states, we cannot determine how state-level policies affect sensitivity of voting to polling place locations. However, evidence from Georgia suggests that the availability of no-excuse voting by mail is important for mitigating the cost of traveling to a polling place. A longer distance to the polling place makes a voter in Georgia less likely to vote at polls and more likely to vote by mail. In Pennsylvania, by contrast, fewer voters substitute to mail-in voting. Importantly, the response of voting in person to distance to polling place is similar in both states. Distance to polling place could deter voters from voting at polls, regardless of whether or not there is a convenient alternative method of voting.

2 Institutional Background

We study the 2018 primary and general elections in Pennsylvania and Georgia. We focus on two large swing states with publicly available voter registration and voter history data.[12] In 2018, there were 8.6 million registered voters in Pennsylvania and 6.9 million voters in Georgia.[13] In many states, including Pennsylvania and Georgia, 2018 was a year of historically high turnout for a midterm election (58% and 53% of registered voters cast ballots compared to 43% and 37% in 2014, respectively). Turnout was much lower for the primary elections (12% for Pennsylvania and 17% for Georgia).

The primary elections took place in May and the general elections on November 6, 2018. Both states elected their governor and all state executives, as well as members of the state legislature and the U.S. House of Representatives. One U.S. Senate seat was up for election in Pennsylvania and none were in Georgia.

Pennsylvania and Georgia differ in a number of election policies.[14] In Pennsylvania,

[11] "Voting precincts closed across Georgia since election oversight lifted." *Atlanta Journal-Constitution* August 31, 2018. Available at: https://www.ajc.com/news/state--regional-govt--politics/voting-precincts-closed-across-georgia-since-election-oversight-lifted/bBkHxptlim0Gp9pKu7dfrN/.

[12] The availability of voter registration files varies from state to state. See https://www.eac.gov/sites/default/files/voters/Available_Voter_File_Information.pdf for more information. By choosing large states we have the ability to study heterogeneous effects across important dimensions like rural/urban, income, education, and race and ethnicity. By choosing swing states, we study voters who are perhaps least sensitive to distance to polling place. If the benefit of voting is low in relatively uncompetitive elections, then voters may abstain regardless of distance to polling place. While we can not test this hypothesis directly, and questions of external validity remain, it is useful that Georgia and Pennsylvania had similar levels of turnout in the 2018 elections.

[13] Voter Registration Statistics from Pennsylvania's Department of State and Voter Registration Statistics from Georgia's Secretary of State Office

[14] Election policies were retrieved from Pennsylvania and Georgia Secretary of State websites and from state election law: Pennsylvania Statutes Title 25 and Georgia Title Code 25.

voting in person happened only on election day in 2018; there was no early voting. Voters also required an excuse to vote by absentee ballot at this time in Pennsylvania.[15] In contrast, early voting began in Georgia three-weeks before election day and any voter could request a mail-in ballot up to 180 days before the election.[16]

Both states are divided into voting precincts by local government authorities (either county election commissions or municipal or county heads of government). An accessible location within each precinct, typically a school, library, police station, or church, is chosen as the polling place location by local authorities. Barring emergencies, polling places must be announced no less than 60 days prior to an election in Georgia and 20 days prior to an election in Pennsylvania. This means that voters may register to vote before knowing exactly where their polling place will be located. Neither state uses same-day voter registration, so voters must register several weeks in advance of the election date. Importantly for our identification strategy, if voting in person, each voter may only vote on election day at the polling place for the precinct in which they reside. Election day polls are open from 7am to 8pm in Pennsylvania and from 7am to 7pm in Georgia. If a registered voter has voted before in Pennsylvania, they do not need to bring identification. Georgia requires voters to show photo identification when voting in person.

We do not have the ability to assess the effect of these electoral policies with only two states and one cross-section of the data. However, the policies are potentially important for understanding the substantive differences in our findings for Pennsylvania and Georgia.

3 Data

From the Pennsylvania Department of State and Georgia Secretary of State, we obtain voter registration files, which include a unique voter identification, address, and voting precinct for each registered voter within the state in 2018. We merge this information with the voter history files, which records whether or not a registered voter voted in each election as well as their method of voting (at polls or by absentee ballot).[17]

We find the locations of polling places from Georgia's Secretary of State's website and from Pennsylvania's state-run polling place look-up website.[18] Next, we determine

[15]Pennsylvania introduced early voting and no-excuse mail-in voting in 2019, after the sample period, with Act 77.

[16]Georgia introduced no-excuse absentee voting in 2005 with House Bill 244. In 2021, the state added an ID requirement to vote by absentee ballot with Senate Bill 202.

[17]For Pennsylvania, we also observe whether the registered voter participated in previous elections. We also were able to observe voter locations in 2016 in Pennsylvania, but not in Georgia. We analyze the panel of voters in Pennsylvania in an alternative empirical specification in Section 4.2.

[18]https://www.pavoterservices.pa.gov/Pages/PollingPlaceInfo.aspx. and https://sos.ga.gov/index.php/

the latitude and longitude of polling place locations and registered voter addresses.[19] We compute the travel-route distance, which is the shortest distance by road in miles between a voter and their assigned polling place.[20] The travel route distance is preferred to Euclidean distance, since it takes road infrastructure into account. This is especially important since we have urban, suburban, and rural areas in our study, and road density might vary systematically with other factors that determine turnout. We also measure the Euclidean distance between voters and polling places and the travel time to the polling place, which is measured in minutes by car. To limit noise from imprecise geocoding, we exclude individuals that are more than 10 miles from their polling place (2.44 standard deviations from the mean in Pennsylvania and 3.09 standard deviations from the mean in Georgia).We also exclude individuals where there is a discrepancy between the precinct that they are geocoded to, using Census voting tabulation districts, and the precinct that they are assigned to in voter registration files. All together, the observations that we exclude as likely geocoding errors account for 3% of registered voters in PA and 2% of registered voters in GA.

Finally, we assign voters to Census blocks using their geolocation, so that we can include additional information from the 2010 Census and the American Community Survey, including income, education, and mode of travel to work (see Online Appendix A for details). We also observe the voting age population of each block, which allows us to test whether distance to polling place affects the share of the voting-eligible population that is registered to vote (Section 5.2).[21]

4 Identification Strategy

We estimate the effect of distance to polling place on the likelihood of voting at the poll, voting by absentee ballot (i.e., voting by mail), and voting by either method. Estimating this causal effect presents several challenges because polling places are not located randomly. Local election officials are supposed to choose convenient and accessible locations for polling places, often using schools and other public buildings. Voters who live close to polling places likely differ systematically from voters who tend to live far away from polling places in ways that might affect turnout decisions.

elections. Retrieved October 2018 and July 2020.

[19]To geolocate voters and polling places, we use the Address Locator provided by ArcGIS. This address locator uses interpolation to locate addresses, meaning it has the latitude and longitude of the endpoints of every street. It then interpolates the latitude and longitude of the specific address based on the street endpoints.

[20]Travel route distance and travel time were computed by the DIGIT Lab at the University of Utah's Department of Geography, which uses ArcGIS.

[21]For the block-level analysis, we assign each block to a unique precinct using the centroid of the block. The geometry of Census blocks and precincts do not overlap perfectly, but we confirm that each block in our sample is assigned to a unique precinct using the 2010 Census Block Assignment Files.

For example, adults who choose to live close to a school may tend to have school-age children, and therefore belong to a demographic group associated with a relatively low turnout rate (Wolfinger and Raymond 2008). These differences might be unobservable or not adequately measured due to the aggregated nature of some of the covariates.

To identify treatment effects, we use variation in distance to polling place created by election precinct borders. Election precincts are small geographic areas that assign voters to polling places. Intuitively, neighbors who live close to a voting precinct border should be comparable in ways that affect voting behavior, but have different costs of voting in person because they are assigned to different polling places. We focus only on a subset of precinct borders where this logic applies. For example, precincts on either side of a school district border would make a poor natural experiment, because individuals might choose to live on one side or the other based on family structure or income. Instead, we want to use variation in distance to polling place created by boundaries that are inconsequential, except for on election day. We therefore restrict attention to the parts of precinct boundaries that do not overlap with important borders. We define a border segment as the length of a precinct border that separates two precincts with unique polling places, and which does not overlap with the borders of school districts, towns, counties, state legislative districts, or federal congressional districts.[22] Figure 1 shows the voting precinct border segments that are included and excluded in the samples for Georgia and Pennsylvania. There are 9,921 border segments in Pennsylvania and 3,945 border segments in Georgia.

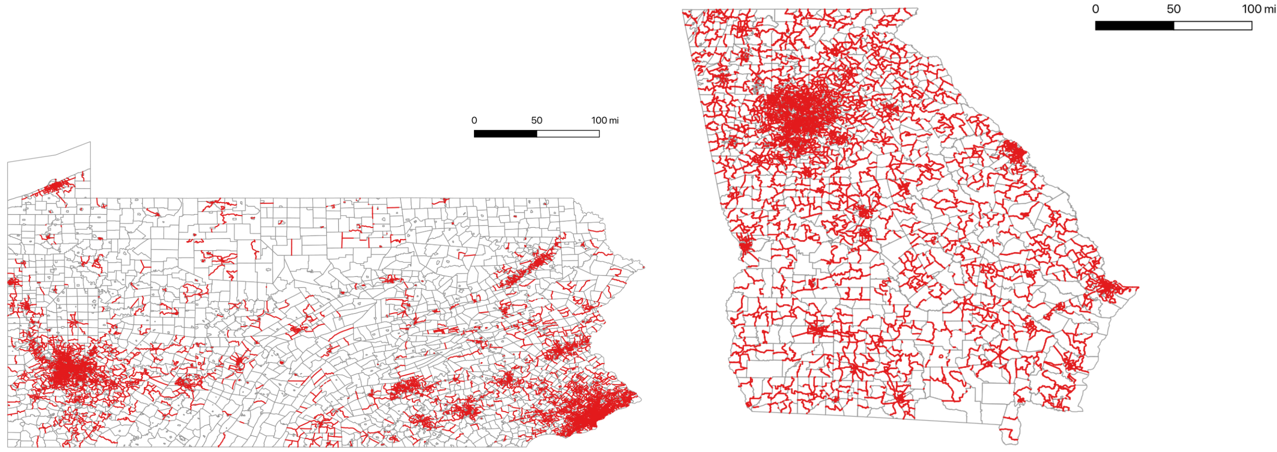
We use two empirical strategies to estimate treatment effects using variation in distance to polling place near precinct border segments. In Section 4.1, we introduce our preferred empirical specification, a geographic Regression Discontinuity (RD). We define the running variable as the distance between a registered voter and a single point on the nearest border segment, called an RD point. Thus, treatment effects are estimated by comparing voters in close proximity to each other, but on opposite sides of a border segment. Figure 2, Panel A shows an example of a precinct boundary, with the locations of voters, the RD point, and polling places indicated. Voters in the shaded area are within the bandwidth used for RD estimation.

We complement the RD specification with a border fixed effects (FE) regression approach (Section 4.2). This allows us to compare our results to the existing literature. Here we follow closely the methodology of Cantoni (2020). We select a sample of voters close to a border segment, then regress voting outcomes on distance to polling place

[22]We obtain 2010 voting tabulation district boundaries from the Census. There may be some differences between boundaries reported to the Census in 2010 and implemented by the states in 2018. To account for this, we only include a border in the sample if voters on either side of the border are assigned to polling places with different geographic locations in 2018. This also accounts for the fact that some precincts share polling places.

Figure 1: Election precinct borders

A. Pennsylvania B. Georgia

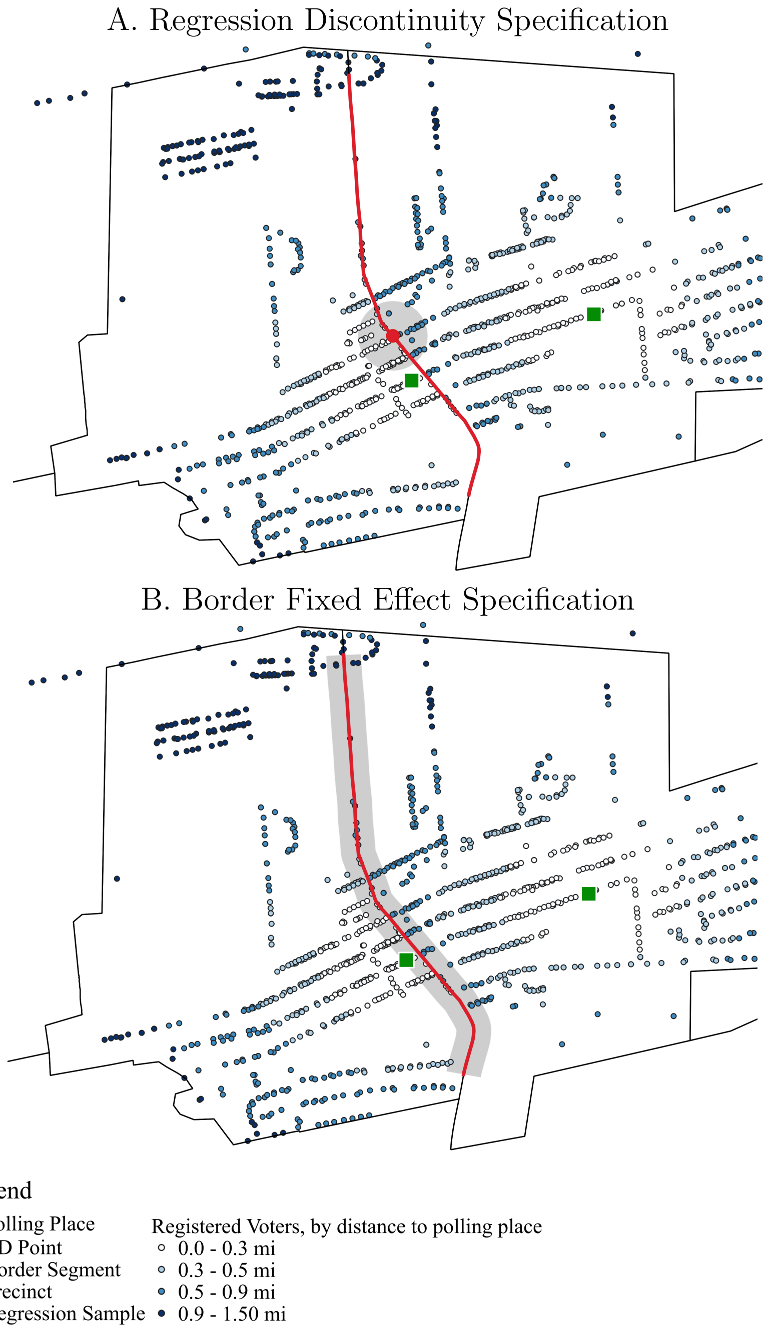


Note: These maps show election precinct borders in Pennsylvania and Georgia. Only precinct border segments highlighted in red are used in the empirical analysis. The red borders do not overlap with the borders of school districts, towns, counties, state legislative districts, or federal congressional districts.

with border segment fixed effects. Figure 2, Panel B shows the same precinct boundary with the shaded area indicating the sample of voters that belong to the same border segment.

The key distinction between the geographic RD and the border FE approach is that the geographic RD uses only *across*-precinct variation in distance to polling place, whereas the border FE approach uses both within and across precinct variation. Note that, since polling places themselves can be located close to the border, there is substantial variation in distance to polling place along one side of the border in Figure 2. This highlights the importance of choosing the correct geographic area to define a fixed effect. Cantoni (2020) uses a buffer zone of 0.05 miles around a border, while Tomkins et al. (2022) use voters with similar addresses. If the fixed effects are not properly selected, for example by choosing an appropriate length of a segment or bandwidth around it, then there may be bias from confounding factors. By comparison, in the RD approach, we only require continuity of confounding factors at a single point on a precinct border. Moreover, we can use existing methodologies to choose bandwidths and present graphic evidence in a transparent, data-driven manner. Finally, a key advantage of the non-parametric RD estimation is that we impose no restrictions on treatment effect heterogeneity. While the parametric border FE approach gives more precise estimates, the linear model may be misspecified.

Figure 2: Geographic discontinuity and empirical specifications



Note: The maps show the locations of voters and polling stations along one of the border segments included in both the regression discontinuity (RD) and border fixed effects (FE) specifications. This border is located in Union county, PA and is 1.2 miles long. The RD sample includes all voters located within the 0.10 mile bandwidth of the RD point (Panel A). The Border FE sample includes all voters located within 0.05 miles of the border segment (Panel B).

4.1 Regression Discontinuity

In the RD specification, we use the exogenous jump in distance to polling place that occurs at the border of an election precinct to identify the effect of distance to the polling place on voter turnout. Consider the following model of voting decisions:

$$Y_i = y(d_i, v_i, u_i), \tag{1}$$

Where Y_i is one of three outcome variables that indicate: whether an individual voted in person, whether an individual voted by mail, and whether an individual voted by either method. For ease of interpreting small coefficients, we have Y_i=100 if voter i votes and $Y_i = 0$ if voter i does not vote. Voting is a function of the distance between a voter and the polling place, d_i, a running variable, v_i, and an error term, u_i. This framework allows for a non-linear relationship between the outcome and distance to polling place and for treatment effect heterogeneity. We are interested in estimating the average marginal effect of distance to polling place on the likelihood of voting, $\partial y/\partial d_i$, called the local average response (Altonji and Matzkin, 2005) or local average treatment effect (LATE).

The choice of the running variable, v_i, is critical for identifying the LATE. Note that, by nature of the geographic discontinuity, we have a two-dimensional treatment assignment rule (i.e., latitude and longitude of the voter). We simplify a two-dimensional RD to a one-dimensional RD, in line with the recommendations of Imbens and Zajonc (2011) and Keele and Titiunik (2015). We construct v_i as the Euclidean distance between a voter i and a single point on the border segment nearest to voter i, which we call an RD point.[23] We choose the midpoint of the border segment as an RD point. We compute the running variable for all voters within 0.5 miles of the RD point.[24] The running variable takes negative values for voters in a control precinct and the positive values for voters in a treatment precinct.

The simplification from a two-dimensional RD to a one-dimensional RD means that we can rely on existing identification results for continuous treatment variables and we can present visual evidence of the discontinuity. It also comes at some cost because we can not write the treatment variable (distance to polling place) as a function of the

[23] As in Keele and Titiunik (2015), we avoid using the shortest distance between a voter and the geographic border as the running variable because voters might be equidistant to the border yet far from each other. Intuitively, we want to draw inference from voters that are located near to each other, not just to a border, which might be long and irregularly shaped. It is likely that confounding factors are continuously distributed in the neighborhood of a single point on the border, though this might not be true for the entire length of a border.

[24] We could, in principal, compute the distance from every voter to every border point. We avoid this due to computational challenges. The data-driven optimal bandwidths for the RD estimator are also substantially smaller than 0.5 miles.

running variable alone. To see why, note that two voters might be equidistant from a point on the border, but differ in distance to polling place (see voters A and B in Figure 3). The fact that we can not write d_i as a function of a single-dimensional v_i means that we have a fuzzy RD rather than a sharp RD (see Appendix B for details). As we have a fuzzy RD framework with a continuous treatment variable, we rely on the identification results of Card et al. (2015).

Intuitively, the fuzzy RD framework applies because distance to polling place will tend to increase for voters as we move from a control precinct to a treatment precinct, but the size and direction of the jump in distance to polling place is not determined by the running variable alone. The last step in defining the running variable is to choose a rule for determining the treatment and control precinct for each RD point. A natural starting point would be to use the precinct with the higher average distance to polling place as the treatment precinct. This is problematic, however, since we are really concerned with the change in the treatment variable for voters near the RD point. Since precincts are relatively small geographic areas, the average distance to polling place can mask substantial *within*-precinct variation in distance to polling place. For example, in Figure 3, Panel A, Precinct 1 has a higher average distance to polling place, but a relatively lower distance to polling place among voters nearest to the RD point. Instead, we want to use voters nearest to the RD point to define treatment status. In the language of a Fuzzy RD framework, we want to ensure that the necessary monotonicity condition is satisfied.

Monotonicity requires that the first stage effect, the change in distance to polling place, goes in the same direction for voters sufficiently close to the discontinuity. In the limit, the distance between the voter and the polling place is exactly equal to the distance between the RD point and the polling place. We therefore use the distance between the RD point and the polling place to define treatment status for each pair of precincts. That is, the treatment precinct is the one with the larger distance between the RD point and the polling place. Let w_i be the Euclidean distance between the RD point and the polling place of voter i. For voters i and j on opposite sides of a border, if $w_i < w_j$ then i is in the control precinct ($v_i < 0$) and j is in the treatment precinct ($v_j > 0$). Using w_i to assign treatment status ensures that there is a discontinuous increase in distance to polling place for voters sufficiently close to the RD point.

While the definition of the running variable ensures monotonicity in the limit, there may be defiers in our finite sample. Typically, in a Fuzzy RD, the outside factors that affect treatment status are unobservable, so monotonicity is an assumption. Here, we observe all factors that determine treatment (the locations of the voter and polling place) and can therefore identify defiers and remove them from the sample. For each RD point, we choose a level of distance to the polling place as a ‘compliance threshold’,

d_c. A natural choice for this threshold is the distance between the RD point and the distance to the polling place in the treatment precinct. A defier would then be a voter in the control precinct with a *higher* distance to polling place than this threshold. Alternatively, we could choose the distance between the RD point and the distance to polling place in the control threshold. Then, a defier would be a voter in the treatment precinct with a distance to polling place *lower* than the threshold. To avoid dropping voters on only one side of the discontinuity, we use the midpoint of the average distance of voters in the treatment and control precincts as compliance threshold. We then drop the defiers: voters that have $d_i < d_c$ in the treatment precinct and voters with $d_i > d_c$ in the control precinct. Figure 3, Panel B, shows a depiction of the compliance threshold and the regions of a precinct where voters would be dropped due to violations of the monotonicity condition. In the sample of voters within 0.5 miles of an RD point, 32% are defiers. The resulting RD sample includes only compliers and only RD points that have both treated and control units. There are 7,388 such RD points in Pennsylvania and 2,334 in Georgia.

The key identifying assumptions are that the distributions of potential confounding factors are continuous at the RD point, that the effect of the running variable on the outcome is continuous at the RD point, and that the running variable is continuously distributed at the RD point (Card et al., 2015). The necessary first-stage effect and monotonicity conditions are satisfied by construction of the running variable and the sample selection outlined above. We present evidence in support of continuity assumptions in Appendix A.2. We test for discontinuities in the sex, age, and party affiliation of individuals. We also test for discontinuities in the census block or block-group level variables: median household income, whether it is an urban block, percent White, percent Black, percent Hispanic, percent with no high school diploma, percent that walk to work, and the number of cars per household. Of the twelve placebo outcomes in two states, there is a statistically significant discontinuity in only one case. In Pennsylvania, there is a 0.6 year decrease in age at the discontinuity. The RD coefficients for all other placebo outcomes are small and statistically insignificant. Given our large sample sizes and the fact that we test for discontinuities in many outcomes, we do not believe that the small discontinuity in age invalidates the RD approach in this setting. In Appendix A.2 we also show that there is no evidence of manipulation of the running variable.

Under the identifying assumptions, the fuzzy RD identifies the average of the treatment on the treated for a single border segment, b, evaluated at a distance that lies between w_i and w_j (Card et al. 2015, Supplemental Materials, Section A2). The fuzzy RD estimand is the ratio of the change in the likelihood of voting at the RD point over the change in distance to polling place at the RD point:

$$\tau_b^{RD} = \frac{\lim_{v \to 0^+} E[Y|v=0] - \lim_{v \to 0^-} E[Y|v=0]}{\lim_{v \to 0^+} d_i - \lim_{v \to 0^-} d_i}. \tag{2}$$

To estimate average treatment effects, we follow recommendations of Imbens and Zajonc (2011) by aggregating across many RD points, rather than estimating τ_b^{RD} separately for each RD point.[25] We estimate τ^{RD} by pooling across RD points. To improve precision we drop RD points with fewer than 50 observations and we partial out RD point fixed effects from all outcome variables. We estimate τ^{RD} using local linear regression with the MSE-optimal bandwidths of Calonico, Cattaneo and Titiunik (2015).[26]

4.2 Fixed Effects Regressions

For the border fixed effects specification, we select all voters within 0.05 miles (264 feet or 80 meters) of any election precinct border.[27] We then estimate the following equation:

$$Y_i = \delta_{s(i)} + \beta d_i + X_i' \gamma + u_i, \tag{3}$$

where Y_i, d_i, and u_i are as before, X_i is a set of covariates for individual i,[28] and $\delta_{s(i)}$ are border segment fixed effects. The identifying assumption is that all unobservable factors affecting the likelihood of voting are uncorrelated with distance to polling place, within registered voters that are assigned to the same border segment. To see if there is evidence in support of this assumption, we regress individual- and census block-level covariates on distance to polling place, including border border fixed effects. Distance

[25] Keele and Titiunik (2015) propose estimating τ_b^{RD} at several RD points and averaging across these points. However, unlike the geographic RD settings they consider, here we have many small borders with relatively low sample size at each border, rather than one long border with many observations. Imbens and Zajonc (2011) provide standard error estimates for an average of geographic RD estimates, but note that the approach is complicated. They propose using a single running variable and aggregating across RD points instead, as we do here. If we think of each RD point as a separate cutoff, then we follow the 'normalize and pool' approach for estimating treatment effects of Cattaneo, Titiunik, and Vazquez-Bare (2020).

[26] We use the Stata command rdrobust.

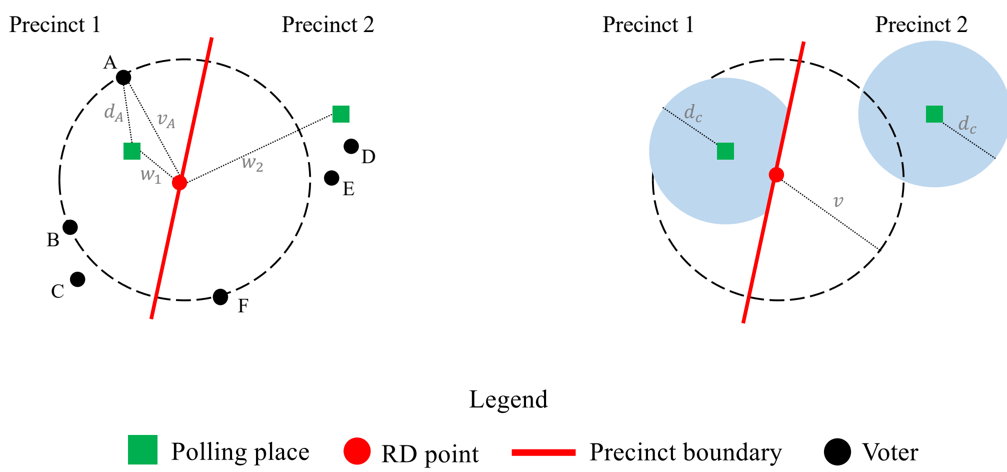
[27] If a voter is within 0.05 miles of two segments (e.g., in a corner of a precinct), then they are assigned to the closer of the two segments. For each segment, we only include voters that are assigned to one of the two polling places for the precincts separated by the border segment (e.g., we exclude voters near an endpoint of the segment that are in a third precinct).

[28] We include indicators for being a registered Democrat, registered Republican, female, and belonging to age groups 30-49, 50-64, and 65 and up. In addition to these individual-level covariates, we include census block- and block-group-level covariates: percent Black, percent Hispanic, median household income, percent without high school diploma, cars per household, percent who commute to work by walking, percent with commute time to work less than five minutes, and percent with commute time to work greater than 60 minutes.

Figure 3: Voters, Polling Places, and the Fuzzy RD Design

A. Voters and polling places

B. Compliance threshold



Note: Panel A shows an example of why treatment and control precincts should be assigned using distances between the polling places and the RD point. The average distance to polling place is higher in precinct 1, but voters closer to the RD point are relatively close to the polling place in precinct 1, relative to precinct 2. As the running variable, v_i tends to zero, d_i tends to w_i. Panel B shows the compliers and defiers for a compliance threshold d_c. The blue shaded area is the set of lat./lon. such that voters are within d_c miles to their polling place. Voters in the shaded area in Precinct 1 and in the non-shaded area in Precinct 2 are compliers for this given d_c. Voters in the non-shaded area in precinct 1 and in the shaded area in precinct 2 are defiers. As the running variable, v, goes to zero, all voters are compliers.

to polling place is statistically significantly correlated with most of the twelve placebo outcomes (Appendix A.3).[29] This raises concerns that border fixed effects do not fully control for confounding factors.[30] One way to improve would be to redefine border segments so that they are smaller, but since borders are irregularly shaped there is no guarantee that doing so would improve balance. This highlights another advantage of the RD approach, in which we avoid ad-hoc decisions about the definition of a border

[29] In Pennsylvania, distance to polling place is negatively correlated with the indicators for being female, a registered Democrat, and in an urban block, and with age and the percent that walk to work. Distance is positively correlated with an indicator for being a registered Republican and median household income. In Georgia, distance to polling place is negatively correlated with indicators for being female, a registered Democrat probability of being female, age, the probability of being registered Democrat, and in an urban block, and with the percent of the block that is Hispanic, percent with no high school degree, and percent that walk to work. Distance is positively correlated with median household income and the percent of cars per household.

[30] In Cantoni (2020), there is minimal concern about correlations between distance to polling place and covariates in the border fixed effects specification. This could be due to the fact that he studies small, straight borders in cities. The same approach may not work well for irregularly shaped borders in suburban and rural areas.

segment sample.

In a second more conservative fixed effects approach, we pair individual voters with the nearest voter on the other side of the border.[31] The estimating equation is the same as 3, but with fixed effects for each matched pair, rather than for a border segment. Thus, variation used to estimate treatment effects comes from only *across* precinct variation in distance to polling place, like with the RD approach. The identifying assumption is that, within each matched pair, distance to polling place is unrelated to confounding factors. While this seems plausible, in Table A.6 we find statistically significant correlations between distance to polling place and covariates even within matched pairs.[32] This suggests that the matching estimator may be susceptible to biases in settings where borders are irregularly shaped or where the population is relatively sparse, compared to the original densely populated urban setting of Cantoni (2020).

5 Empirical Results

We first report the RD estimates for 2018 general elections in Pennsylvania and Georgia (Table 1). The RD plots in Figure 4 show corresponding graphical evidence for both the first-stage and reduced form results.

In Pennsylvania, the average increase in distance to the polling place at the discontinuity is 0.46 miles (SE=0.03). The estimated effect of distance to polling place on at-poll voting is -2.71 (SE=1.16), meaning that for every additional mile of distance to polling place the likelihood of voting in person falls by 2.71 percentage points (p.p.), a 5% reduction. There is a much smaller and statistically insignificant increase in voting by absentee ballot (0.39 p.p., SE=0.26), such that the estimated decrease in the overall likelihood of voting is 2.35 p.p. per mile (SE=1.15).

In Georgia, the average increase in distance to the polling place at the discontinuity is 1.30 miles (SE=0.09). The estimated effect of distance to polling place on at-poll voting is -1.25 p.p. (SE=0.74), a 5% reduction. There is a similarly sized increase in absentee voting (1.20 p.p., SE=0.60), such that there is no statistically significant effect on the overall likelihood of voting (-1.13 p.p., SE=1.01). The null result for overall turnout is fairly precise. We can rule out that turnout decreases by more than 3.1 p.p. with 95% confidence.

[31] Each voter is paired with the nearest voter in the other precinct that is also within 0.05 miles of the border segment. Voters are matched with replacement. Since a single voter can belong to multiple matched pairs, we cluster standard errors by border and precinct (following Dube et al. 2010 , Cantoni 2020).

[32] The voter with the higher distance to polling place tends to be in a block with higher median household income, and is less likely to be in an urban block in both states. In Georgia, the voter with higher distance to polling place is also less likely to be Hispanic, is more likely to have a high school diploma, and is less likely to walk to work.

In both states, we see similar results if we measure the effects of an increase in Euclidean distance to polling place. We also estimate the effect of an increase in travel time to the polling place by car. Time to polling place at the discontinuity increase by 2 minutes in Pennsylvania and by 3 minutes in Georgia. The second-stage estimates again suggest small average effects: the likelihood of voting in person falls by 0.94 p.p. in Pennsylvania (SE=0.25) in Pennsylvania and by 0.31 p.p. (SE=0.25) in Georgia (Appendix C). We also find similar effects when estimating the effect of travel route distance to polling place on the likelihood of participating in the 2018 primary election or in the 2016 presidential election (available for Pennsylvania only, Appendix D).

Table 1: Fuzzy RD Estimates

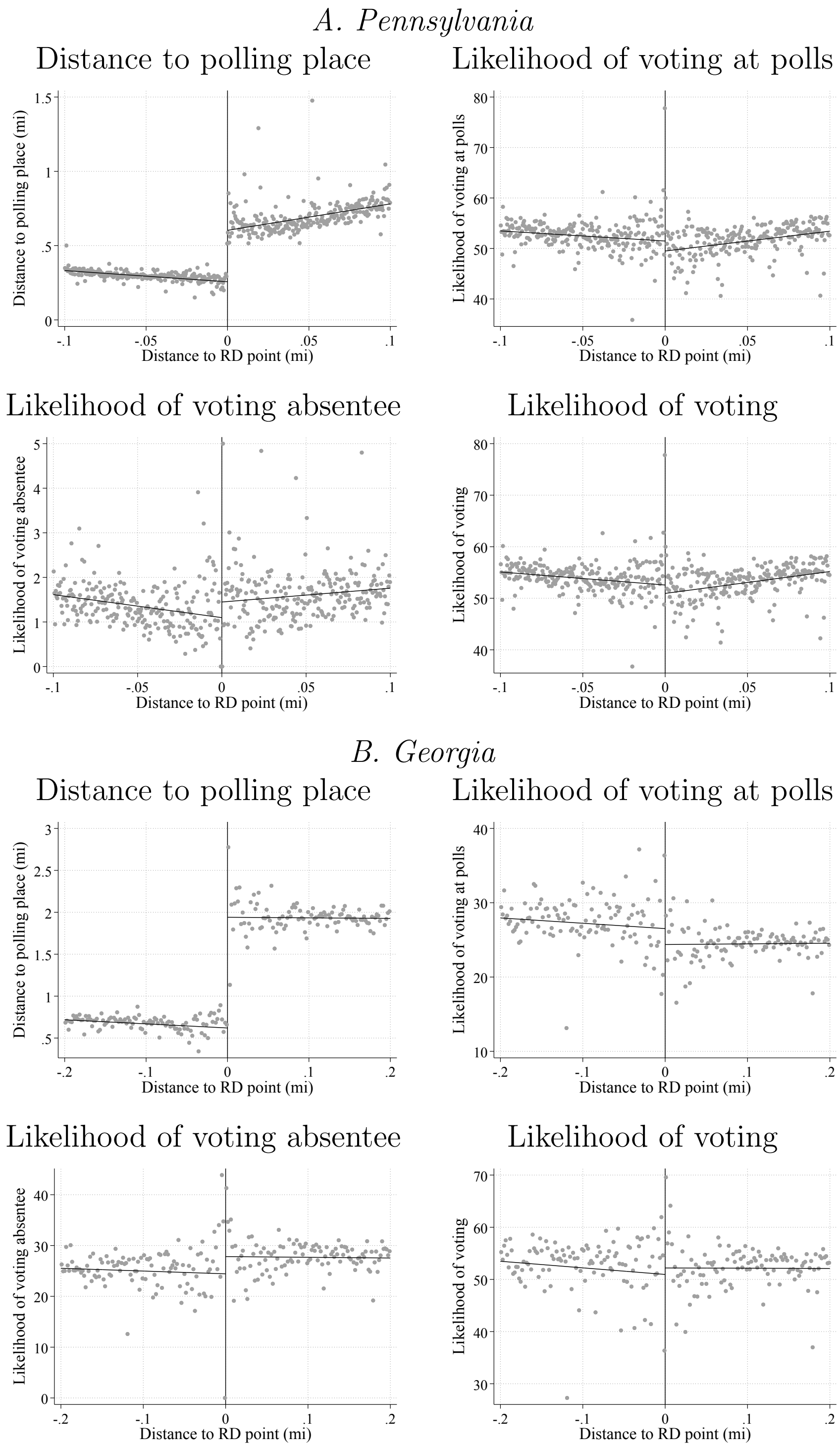
	A. Pennsylvania			
	First-stage	Second-stage		
	Distance (mi)	At Poll	Absentee	Total
RD Estimate	0.46***	-2.71**	0.39	-2.35**
	(0.03)	(1.16)	(0.26)	(1.15)
N	2,315,556	2,315,553	2,315,553	2,315,553
Effective N, Left	274,495	274,495	300,648	280,281
Effective N, Right	317,324	317,324	345,736	324,006
Bandwidth	0.09	0.09	0.10	0.09
Outome mean	0.76	56.21	2.10	58.31

	B. Georgia			
	First-stage	Second-stage		
	Distance (mi)	At Poll	Absentee	Total
RD Estimate	1.30***	-1.25*	1.19**	-1.13
	(0.09)	(0.74)	(0.59)	(1.01)
N	1,227,157	1,227,157	1,227,157	1,227,157
Effective N, Left	84,056	84,056	201,608	72,870
Effective N, Right	135,745	135,745	313,674	116,091
Bandwidth	0.16	0.16	0.26	0.15
Outome mean	1.53	26.02	27.98	54.00

Note: This table reports RD Estimates, as specified in Equation 2. Outcome variables indicate whether a registered voter voted in person (At Poll), by mail (Absentee), or by either method (Total). Outcomes are scaled so that coefficients are measured in percentage points. Distance (mi) measures the travel route distance between a voter and their assigned polling place. Bandwidths are MSE-optimal. Standard errors allow for clustering at the RD point level.

The RD estimates suggest that an increase in distance to polling place deters registered voters from voting in-person. In Pennsylvania, this translates into a drop in overall turnout. In Georgia, voters substitute to mail-in voting so that there is a null effect on overall turnout. It is important to remember that Pennsylvania and Georgia

Figure 4: Discontinuities in distance to polling place and likelihood of voting



Note: This figure shows binned scatter plots and linear trends to the left and right of the RD point discontinuity. Distance to the RD point is the Euclidean distance between a voter and RD point (mi). Distance to the polling place is travel route distance between the voter and polling place (mi). We use the Calonico, Cattaneo, and Titiunik (2015) mimicking variance evenly-spaced method to select bins with the Stata command *rdplot*.

differed in their requirements for voting by mail in 2018. Pennsylvania required voters who request absentee ballots to provide an excuse while Georgia did not. Without variation in no-excuse absentee voting policies during the study period, we can not directly test the hypothesis that the availability of no-excuse absentee voting drives the higher uptake of absentee voting in Georgia.[33] However, the estimates are consistent with a larger share of voters substituting to mail-in voting if it's convenient to do so in response to a change in the cost of voting at polls. To interpret the magnitude of the estimated effects, and to compare estimates to existing studies in different contexts, it is important to consider the possibility of nonlinear effects.

5.1 Nonlinear effects

Although the non-parametric approach allows for a non-linear relationship between distance to polling place and voting decisions, such heterogeneity makes it difficult to interpret the LATE estimates. There are at least two reasons to expect non-linearities in this setting. It could be that the costs of voting in person are increasing in distance to polling place, as with a standard convex cost function. In that case, marginal effects would be larger in magnitude for voters further away from their polling place. Alternatively, voters who live close to a polling place may be more sensitive to this cost of voting if they are less likely to use a car to get to the polling place. Small changes in distance to polling place could be more costly for voters who walk or use public transportation than for those who use cars.

To test for non-linear effects, we estimate treatment effects separately for subsamples of RD points based on quartiles of distance to polling place (Figure 5 and Tables E.1 and E.2).[34] In both states, there is a pattern of non-linear effects in which voters closest to polling places are most sensitive to the costs of traveling to the polls. In Pennsylvania, voters in the bottom quartile (who are 0.3 miles from the polling place on average) are 7.54 p.p. (SE=0.33) less likely to vote in person per additional mile of distance to polling place, whereas there is a null effect for voters in the top quartile (who are 1.5 miles from the polling place on average, the RD coefficient is 0.60 p.p., SE=0.38). We reject the hypothesis that the coefficients for the top and bottom quartile are equal with a p-value of 0.01. Comparing bottom and top quartiles of voters by distance to polling place in Georgia (who are 0.7 vs 2.6 miles from the polling place),

[33] The switch from excuse-only to no-excuse absentee voting had no immediate effect on the sensitivity of voters to distance to polling place in the case of Minnesota (Cantoni 2020). In Georgia, no-excuse absentee voting may function differently, since it was a long-standing policy by the time of the 2018 election.

[34] Each RD point identifies the marginal effect $\partial y/\partial d_i$ at a level d_i between the distance of control and treatment observations close to the RD point. We take the average distance to to polling place among control and treatment observations and use the midpoint of the two averages to assign a single level of distance to polling place to an RD point.

point estimates imply a 6.77 p.p. (SE=3.61) versus 1.64 p.p. (SE=0.59) reduction in the likelihood of voting at polls, though we fail to reject that the coefficients are equal with p-value of 0.16. The rate at which voters substitute to absentee ballots differs across states and quartiles of voting. In Pennsylvania, voters closest to polling places are more likely to substitute to absentee ballots as distance to polling place increases. In Georgia, there is no clear pattern and the estimates are noisy. However, there is no statistically significant decrease in the overall likelihood of voting at any quartile of distance to polling place. Overall, the evidence of nonlinear effects suggests that other covariates, namely mode of transportation, are important moderating factors for the effect of distance to polling place on the likelihood of voting in person. We test this hypothesis more directly in Section 5.4.

The estimates for bins of RD point by distance to polling place help to interpret the magnitude of the coefficients. We take the average of the four estimated LATEs, weighting by observations per bin. Given the presence of non-linear effects, this is a more accurate estimate of an average treatment effect than the estimates in Table 1. The point estimates in Figure 5 imply that, in Pennsylvania, a mile increase in distance to polling place causes a 3.3 p.p. reduction in the likelihood of voting at polls and a 2.7 p.p. reduction in the likelihood of voting overall. In Georgia, a mile increase in distance to polling place causes a 1.8 p.p. reduction in the likelihood of voting at polls and a 1.5 p.p. reduction in the likelihood of voting overall. The differences across states may reflect that voters are closer to polling places, on average, in Pennsylvania than in Georgia.

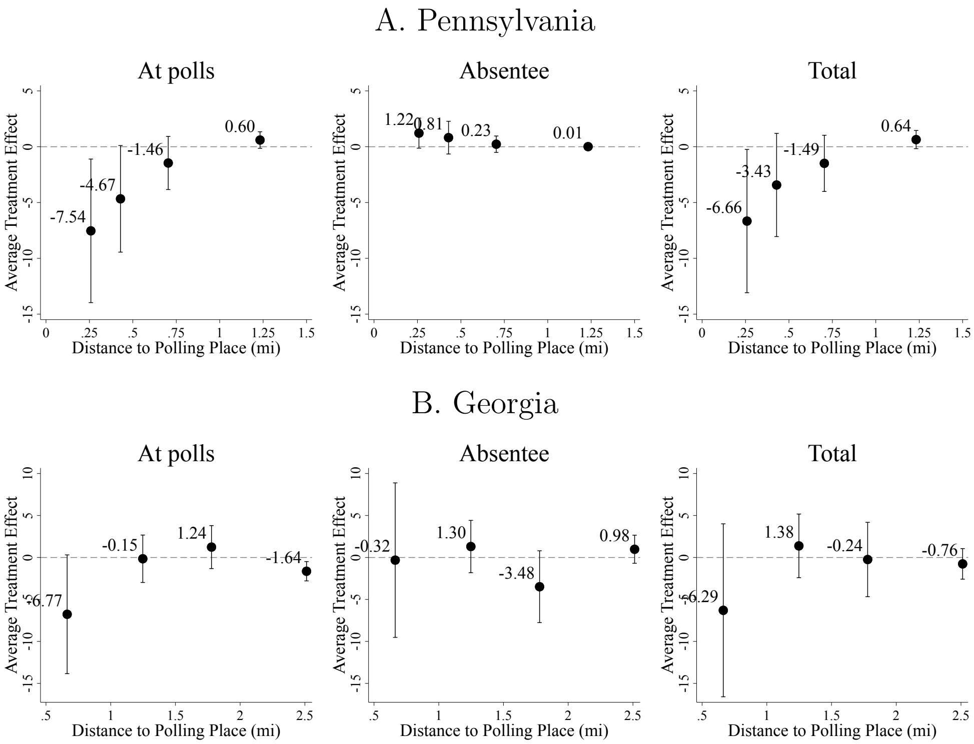
Another way to interpret coefficients is to take the average of the change in turnout due to a standard deviation change in distance to polling place. This accounts for the fact that voters who are close to polling places tend to be in small precincts. Thus, voters who are the most sensitive are also less likely to experience large increases in distance to polling place.[35] In Pennsylvania, a one standard deviation increase in distance to polling place reduces the likelihood of voting in person by 0.58 p.p. and reduces the likelihood of voting overall by 0.45 p.p.. In Georgia, a one standard deviation increase in distance to polling place reduces the likelihood of voting in person by 1.09 p.p. and reduces the likelihood of voting overall by 0.77 p.p..

5.2 Fixed Effect Regression Results

In Table 2 we report estimated coefficients for travel route distance to polling place in specifications that include precinct fixed effects (Panel A), border segment fixed

[35] We report standard deviation in distance to polling place by bin in Tables E.1 and E.2. A standard deviation of distance to polling place is 0.17 and 0.41 miles for the bottom quartiles and 1.37 and 1.53 miles for the top quartiles in PA and GA, respectively.

Figure 5: Nonlinear effects: Average treatment effects by distance to polling place and by time to polling place



Note: This figure shows binned RD estimates. The RD sample is divided into quartiles based on the average distance to the polling place associated with an RD point. The distance to polling place associated with an RD point is the midpoint between the mean distance to polling place in the control precinct and the mean distance to polling place in the treatment precinct. The y-axis is measured in percentage points of likelihood of voting per minute of travel time to polling place. The x-axis shows the average travel time of each sub-sample. Vertical lines indicate 95% confidence intervals. Standard errors allow for clustering at the RD point level.

effects (Panel B), or matched pair fixed effects (Panel C). Coefficients are stable across the three specifications in both Pennsylvania and Georgia. In Pennsylvania, the fixed effects specifications yield precisely estimated null effects for all voting outcomes. With border fixed effects (Panel B), the estimated effect of distance to polling place on the likelihood of voting at polls is -0.24 (SE=0.16), meaning we can rule out that the likelihood of voting decreases by more than 0.55 percentage points. There is an estimated 0.14 p.p. increase in the likelihood of voting by mail (SE=0.04) and a 0.09 p.p. decrease in the overall likelihood of voting (SE=0.15). In Georgia, the border FE estimates are -1.12 p.p. (SE=0.12) for voting in person, 1.31 p.p. (SE=0.13) for voting by mail, and 0.13 p.p. (SE=0.13) for the overall likelihood of voting.

To summarize, estimates from fixed effects specifications are similar to those from the RD estimation for Georgia, but not for Pennsylvania. To reconcile these findings, it could be that the border FE and matching FE specifications are misspecified because they omit confounding factors or because they impose a linear relationship between distance to polling place and voting. We argue that the latter explanation is more likely in this setting. Although it is possible that confounding factors are correlated with distance to polling place within a precinct or border segment sample, it is unlikely to be the case for matched pairs of voters. Given that point estimates change very little when we estimate effects using variation within a precinct (Table 2, Panel A), within a set of voters near the same border (Panel B), or within a pair of voters on opposite sides of a border (Panel C), we expect that omitted variable bias does not explain the differences between estimates in Table 1 and Table 2.

Instead, the presence of non-linear effects can explain the differences between the estimated LATEs in Table 1 and ATEs in Table 2. The LATEs estimated by the RD approach represent the weighted average of the marginal effect of distance to polling place on the likelihood of voting, with higher weights for observations closest to an RD point. The ATEs estimated by the FE specifications can also be interpreted as weighted average treatment effects, using the weighting results of Angrist and Pischke (2009). Observations receive relatively higher weights if they belong to a border segment or matched pair with large variation in distance to polling place. Given the non-linear effects in Figure 5, one under-estimates the average marginal effect when up-weighting borders with large differences in distance to polling place. Indeed, even if we estimate effects for sub-samples of voters by distance to polling place, we estimate smaller effects relative to the binned RD estimates (Tables E.3 and E.4).[36]

Several robustness checks confirm that nonlinear effects, and other sources of treatment effect heterogeneity, can reconcile our findings with those in the existing literature. To start, heterogeneous effects can explain why our estimates are an order of magnitude smaller than those in Cantoni (2020), even though the border FE methodology is essentially the same.[37] In Appendix G, we select Census Blocks in Pennsylvania and Georgia that are similar, based on observables, to the Census Blocks in the Minneapolis and Boston areas (using replication data for Cantoni, 2020). Point estimates are indistinguishable between the matched sample, which contains more urban areas, and the original sample of Cantoni (Table G.1). In general, voters tend to be more sensitive to distance to polling place in large cities with short distances to polling place (see Appendix H for estimates specific to Philadelphia, Pittsburgh, and Atlanta). Given the large differences in estimated distance to polling place within a state, we further

[36]This highlights an advantage of the RD approach. The LATE is well-defined for bins of RD points, since the distances to polling place at the RD point determines the range at which we estimate the marginal treatment effect. By comparison, it's not obvious how to bin border segments in the fixed effects approach.

[37]The key difference is that Cantoni estimates turnout rates for parcels of land or Census Blocks, whereas we focus on individual likelihood of voting. We also estimate effects at the Census Block level in Appendix F.

investigate heterogeneous effects in Section 5.4.

Table 2: Fixed Effects Estimates

	A. Precinct Fixed Effects					
	Pennsylvania			Georgia		
	At Poll	Absentee	Total	At Poll	Absentee	Total
Distance to Polling Place (mi)	-0.20	0.28***	0.08	-1.09***	1.37***	0.28**
	(0.23)	(0.06)	(0.23)	(0.11)	(0.12)	(0.13)
N	1,507,854	1,507,854	1,507,854	390,572	390,572	390,572
Outcome mean	51.90	1.64	53.54	24.73	27.70	52.44
R^2	0.10	0.03	0.11	0.04	0.17	0.20
	B. Border Fixed Effects					
	Pennsylvania			Georgia		
	At Poll	Absentee	Total	At Poll	Absentee	Total
Distance to Polling Place (mi)	-0.24	0.14***	-0.09	-1.18***	1.31***	0.13
	(0.16)	(0.04)	(0.15)	(0.12)	(0.13)	(0.13)
N	1,507,854	1,507,854	1,507,854	390,572	390,572	390,572
Outcome mean	51.90	1.64	53.54	24.73	27.70	52.44
R^2	0.10	0.03	0.11	0.05	0.18	0.20
	C. Matched Pair Fixed Effects					
	Pennsylvania			Georgia		
	At Poll	Absentee	Total	At Poll	Absentee	Total
Distance to Polling Place (mi)	0.41	0.12	0.53	-1.26***	1.32***	0.06
	(0.65)	(0.08)	(0.64)	(0.28)	(0.33)	(0.31)
N	1,868,442	1,868,442	1,868,442	654,424	654,424	654,424
Outcome mean	50.37	1.64	52.01	24.92	28.46	53.39
R^2	0.55	0.51	0.55	0.52	0.58	0.60

Note: This table reports estimates from linear fixed effects regressions. Regressions in Panel A include precinct fixed effects, regressions in Panel B include border fixed effects, and regressions in Panel C include matched-pair fixed effects. Outcome variables indicate whether a registered voter voted in person (At Poll), by mail (Absentee), or by either method (Total). Outcomes are scaled so that coefficients are measured in percentage points. Distance (mi) measures the travel route distance between a voter and their assigned polling place. Standard errors allow for clustering at the precinct level in Panel A, at the border segment level in Panel B, and at the border segment and precinct level in panel C.

5.3 Voter registration costs and search costs

In this section we address additional costs of voting that are related to polling place locations. First, a larger distance to polling place may reduce the likelihood that a voter registers to vote, if the voter is aware of the polling place location in advance. In Georgia, eligible voters are automatically registered to vote if they provide their information to the Department of Driver Services (e.g., to get a driver's license, unless one chooses to opt out). In 2018, Pennsylvania had no such automatic registration system.

So far, we have estimated the effect of distance to polling place on voting decisions, conditional on an individual already being registered to vote. We test for voter registration effects using a block-level fixed effects regression.[38] At the block level, we have information on the voting age population, a proxy for voting eligibility, from the 2010 Census. We regress the percent of the block's voting age population that is registered to vote on the distance from the block centroid to the polling place. We include border segment fixed effects, and only include blocks that contain voters within 0.05 miles of the border. We find no evidence of a significant change in voter registration rates due to increased distance to polling place (Table F.1). The point estimates are small and statistically insignificant: in both states a mile increase in distance to polling place causes a 0.4 percentage point reduction in registration rates. We also find point estimates similar to those in Table 2 when we regress turnout (number of votes per voting age population) on distance to polling place (Table F.1).

Second, although we estimate a change in the likelihood of voting due to a change in distance to polling place, in practice such changes may create additional short-term costs of voting because a voter has to locate the new polling place. Such search costs are the focus of Clinton et al. (2020) and Yoder (2018). We construct a panel of voters and polling place locations for Pennsylvania (from 2016 to 2018). Using a differences-in-differences approach, we find that if a polling place moves, there is no statistically significant effect on the likelihood of voting. But, if a voter moved to a new election precinct, then there is an estimated 3.0 p.p. reduction in the likelihood of voting at polls in the 2018 general election (Appendix I).

5.4 Heterogeneous effects

In this section, we investigate differences in the responses to distance to polling place by demographic, economic, and political factors. To test for heterogeneous effects, we estimate RD coefficients for sub-samples.[39] We test if coefficients are equal across sub-samples using standard t-tests, with p-values reported in Tables in Appendix J. These results are suggestive: we do not have exogenous variation in the covariates of

[38] Cantoni (2020) similarly aggregates to the block-level, noting that the benefit of avoiding selection bias comes at the cost of less precise measurement of distance.

[39] We could otherwise estimate interaction effects, but that approach over-rejects the null-hypothesis of no heterogeneous effects when the linear model is misspecified (Hsu and Shen, 2019).

interest, but we can observe associations between covariates and estimated treatment effects. We first consider heterogeneous effects for covariates that we observe at the individual-level from voter registration files: age, sex, political party affiliation, and race and ethnicity (available for Georgia only). Using data from census block groups, we test for heterogeneous effects by educational attainment, income, and mode of transportation to work. Since we do not observe these characteristics at the individual level, we estimate effects for sub-samples of RD points with above vs. below median values of the continuous covariate.

Age and Sex. One might expect that younger voters are more sensitive to costs of voting, since they may not have formed the habit of voting (Fujiwara et al. 2016; Plutzer 2002). On the other hand, the oldest voters may be more likely to request an absentee ballot if distance to polling place is larger due to health and accessibility concerns. In Figure 6, we see suggestive evidence that older voters are more sensitive to distance to polling place than younger voters in Georgia and in Pennsylvania, though the differences between those aged 65 and up and other age groups are only statistically significant in Georgia. We find no differences in sensitivity to distance to polling place between male and female registered voters (Table J.1).

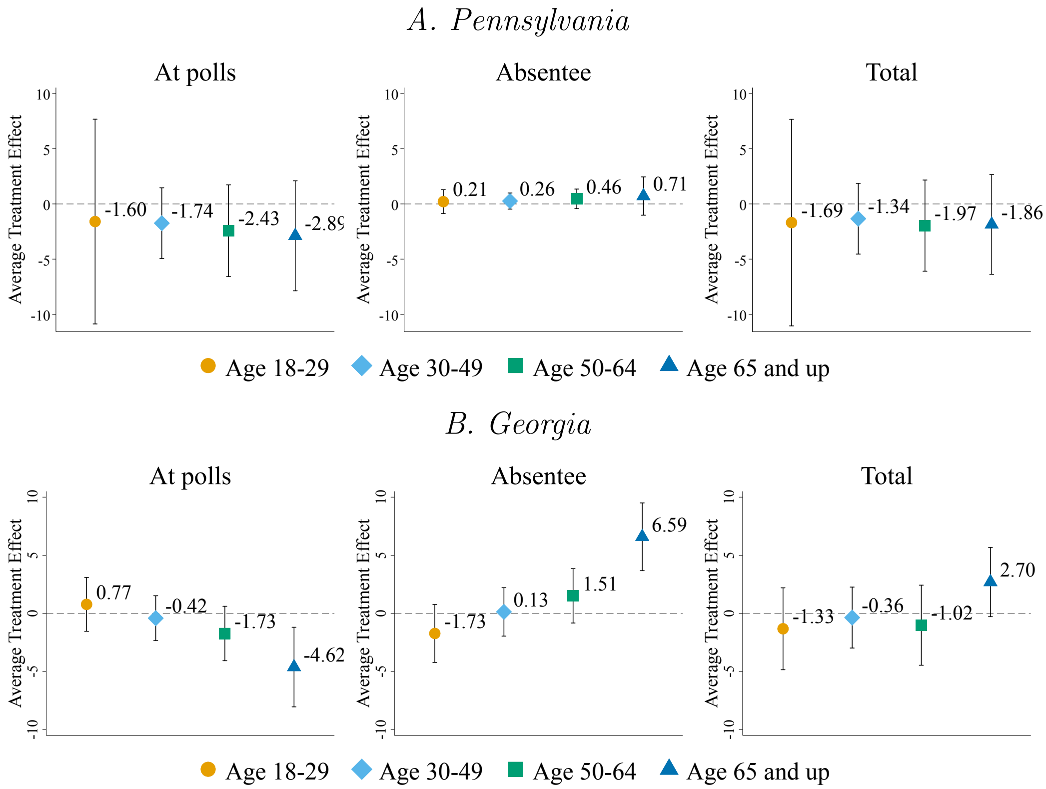
Race and Ethnicity. Race and ethnicity are especially important to consider in the context of the cost of voting due to both a persistent turnout gap between white and non-white voters (Fraga 2018; Ansolabehere et al. 2021) and long-standing concerns of voter disenfranchisement. In recent years, decisions to close or move polling places have come under increased scrutiny due to concerns over voter suppression.[40] Related to these concerns, recent studies show that Black voters are more likely to experience longer waiting times at polls (Chen et al. 2020) and are more likely to have their mail-in ballots rejected (Shino et al. 2021).

In Figure 7, we find no significant differences in how the overall likelihood of voting responds to distance to polling place between White non-Hispanic, Black non-Hispanic, and Hispanic voters. However, Black non-Hispanic voters are statistically significantly more likely to substitute to mail-in voting as distance to polling place increases than Hispanic voters.

Existing evidence of differential effects of distance to polling place by race is mixed. Cantoni (2020) finds that areas with a larger non-white population are more sensitive to distance to polling place, while Clinton et al. (2020) find that non-white voters are less likely to substitute to early voting in response to polling place changes than white voters. In studying recent polling place and precincting decisions in North Carolina, Shepherd et al. (2021) find no evidence of manipulation of polling place choices that

[40] "The Georgia Governor's Race Has Brought Voter Suppression Into Full View", *The Atlantic*. Retrieved Novermber 6, 2018. "Republican Voter Suppression Efforts Are Targeting Minorities, Journalist Says", *NPR*. Retrieved October 23, 2018.

Figure 6: Heterogeneous Effects by Age



Note: This figure shows RD estimates for sub-samples of registered voters by age. The y-axis is measured in percentage points of likelihood of voting per mile of distance to polling place. Vertical lines indicate 95% confidence intervals. Standard errors allow for clustering at the RD point level.

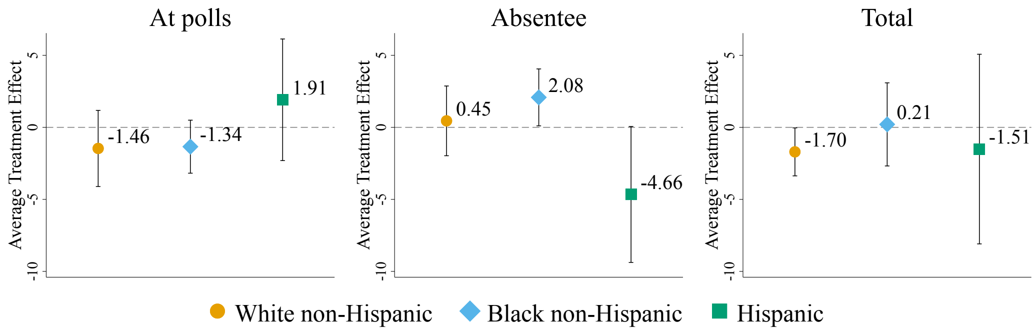
would systematically affect voters differently by race. The mixed findings suggest that context-specific factors and electoral design are important to take into account when considering whether or not polling place locations and changes will have disparate racial impacts.

Political Party Affiliation. Making use of the party affiliation from voter registration files, we estimate the effect separately for registered Democrats, registered Republicans, and for voters that do not register with either party (Figure 8). In both states, there is no significant difference across groups in the effect of distance to polling place on the overall likelihood of voting. In Georgia, an increase in distance to polling place causes a higher rate of substitution to absentee voting among Republicans than among Democrats. There is no such differential in substitution to mail-in voting in Pennsylvania. Georgia and Pennsylvania are both swing states, so these findings may not generalize to states with strong support for one political party.

Income and Education. Participation in elections in the United States has historically been higher among those with relatively high income and educational attainment

Figure 7: Heterogeneous Effects by Race and Ethnicity

B. Georgia



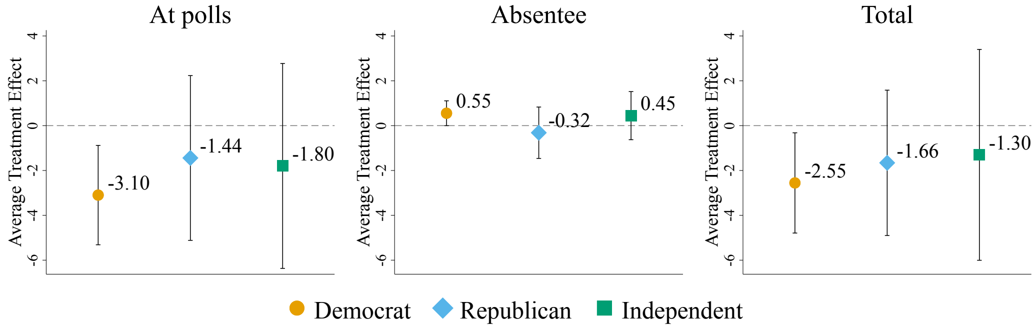
Note: This figure shows RD estimates for sub-samples of registered voters by race and ethnicity, as reported on voter registration files. This information is available at the individual-level in Georgia but not in Pennsylvania. The y-axis is measured in percentage points of likelihood of voting per mile of distance to polling place. Vertical lines indicate 95% confidence intervals. Standard errors allow for clustering at the RD point level.

(Milligan et al. 2004, Sondheimer and Green 2010, Leighley and Nagler 2014). The costs associated with traveling to the polling place could potentially lead to unequal political representation by income and education. In theory, the direction of how income might moderate the effect of distance to polling place on turnout is unclear. On one hand, high-income voters may have a higher opportunity cost of voting because of their income-earning opportunities. On the other hand, low-income voters may have a higher opportunity cost of voting due to lower flexibility to take time out of work and because they are more likely to receive hourly wages. In Figure 9, the likelihood of voting in person is more sensitive to distance to polling place in relatively low income areas in both Pennsylvania and Georgia, though differences between low and high income areas are statistically insignificant. However, in Georgia, voters in low income areas are statistically significantly less likely to take up mail-in voting, such that a mile increase in distance to polling place reduces the overall likelihood of voting by 3.63 p.p. (SE=1.63). For voters in relatively high income areas, a mile increase in distance to polling place causes a small reduction in voting at polls (0.49 p.p., SE=1.06) and a comparatively large uptake in voting by absentee ballot (2.48 p.p., SE=0.97), such that there is a null effect on the likelihood of voting overall.

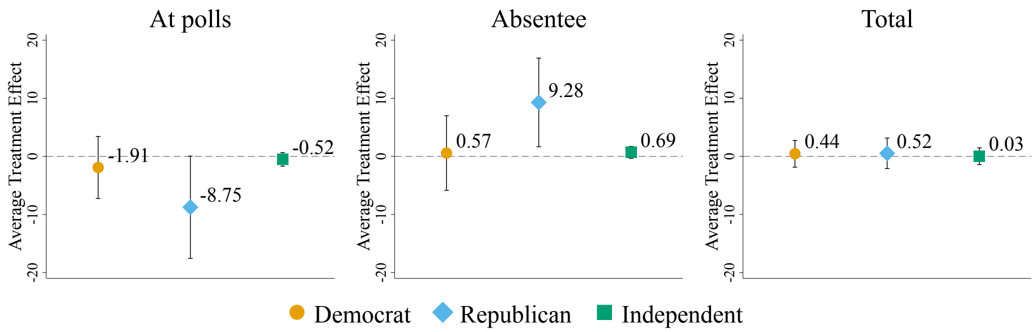
There is a similar pattern by educational attainment (Table J.6), where point estimates indicate that areas with low educational attainment are associated with larger reductions in voting at polls in response to an increase in distance to polling place (Pennsylvania and Georgia), and also less likely to substitute to mail-in voting (Pennsylvania only). The patterns by income and education suggest that there may be barriers to the take-up of voting by mail, even in states like Georgia where voters can

Figure 8: Heterogeneous Effects by Political Party

A. Pennsylvania



B. Georgia



Note: This figure shows RD estimates for sub-samples of registered voters by party affiliation, as reported on voter registration files. The y-axis is measured in percentage points of likelihood of voting per mile of distance to polling place. Vertical lines indicate 95% confidence intervals. Standard errors allow for clustering at the RD point level.

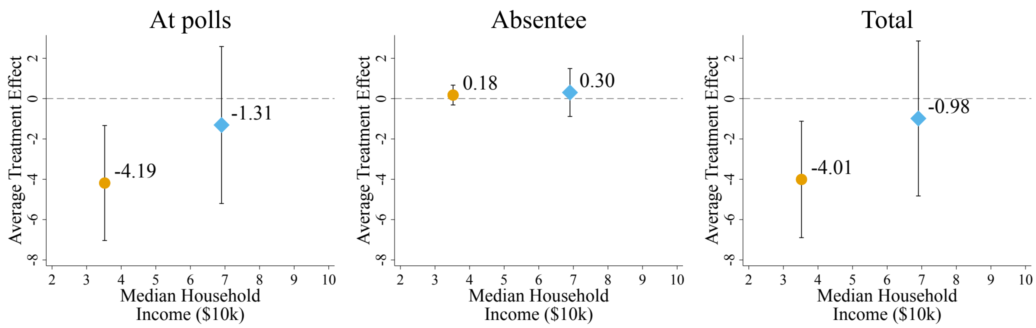
need no excuse to receive an absentee ballot.

Transportation. Finally, the cost of distance to the polling place will vary by mode of transportation. We have already seen that voters who live closer to their polling place are more sensitive to distance to polling place in Section 5.1. Voters who live closer to their polling place are also less likely to drive to the polls. It could be that nonlinear effects in Figure 5 represent associations between distance to polling place and time to polling place. We test this hypothesis by constructing an alternative measure of time to polling place, which takes mode of transportation into account. for nonlinear effects by time to travel to the polling place, where we assume that voters who are within 0.5 miles of the polling place walk along the travel route at 3 miles per hour, and that all other voters drive along the travel route according to the speed limit.[41] Figure 10 shows that there is no evidence of non-linear effects of time to travel to the polling place on voting outcomes, if we assume that those closest to

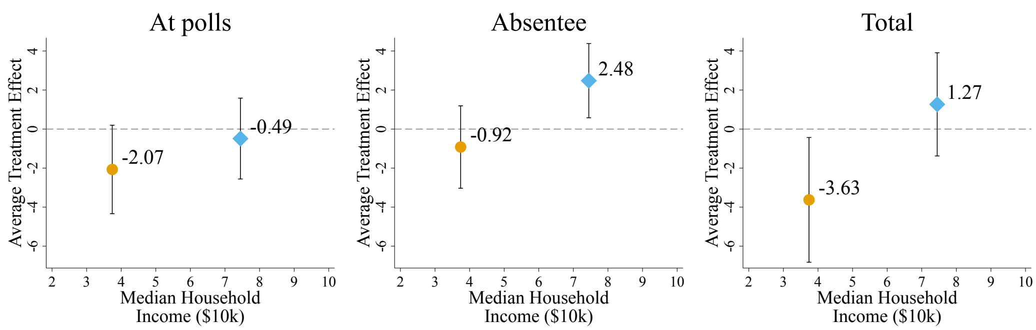
[41] Three miles per hour is the average walking speed used by Google Maps.

Figure 9: Heterogeneous Effects by Income

A. Pennsylvania



B. Georgia

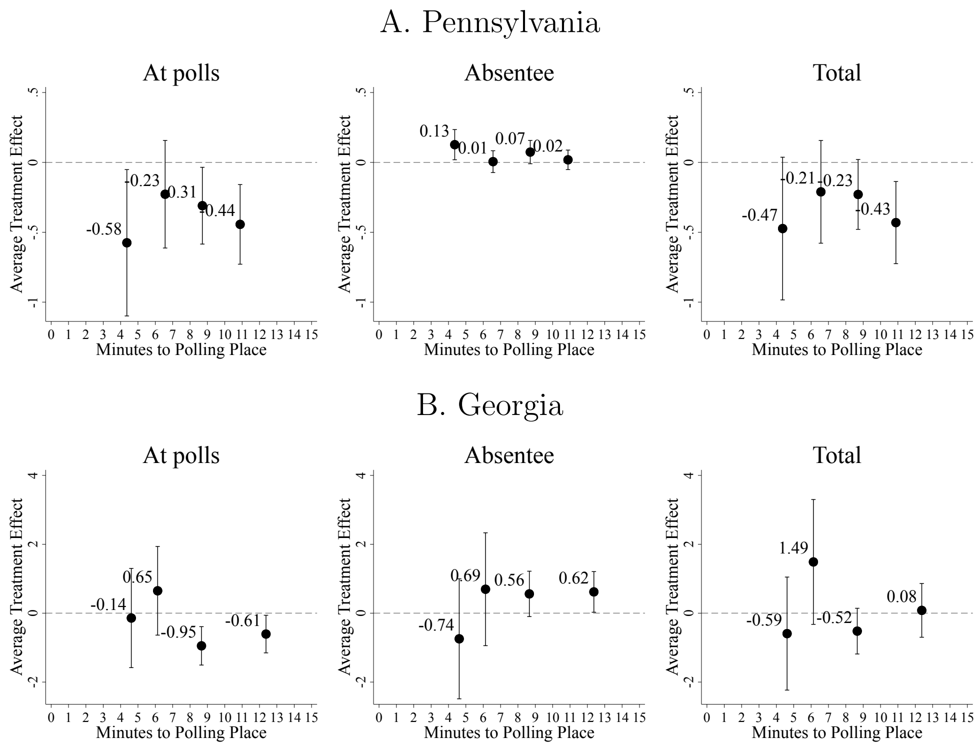


Note: This figure shows RD estimates for sub-samples RD points with above and below median income. The income-level of an RD point is measured as the average of median household income among registered voters. Median household income is measured in 10,000 US dollars. The y-axis is measured in percentage points of likelihood of voting per mile of distance to polling place. The x-axis shows the average median household income of each sub-sample. Vertical lines indicate 95% confidence intervals. Standard errors allow for clustering at the RD point level.

the polling place walk instead of drive. Although there is no way to know the actual mode of transportation and travel time experienced by voters, we interpret this as suggestive evidence that the non-linear effects in Figure 5 reflect differences in mode of transportation and travel time costs. To further test this hypothesis, we estimate effects for sub-samples of RD points in areas where there is an above and below median rate of commuters who travel to work by walking. In Table J.7, we see that voters in areas with above median rates of commuting by walking are more sensitive to distance to polling place in Pennsylvania. The effect of an additional mile to polling place reduces the likelihood of voting in person by 4.6 p.p. (SE=1.5) in areas with high rates of commuting by foot, versus only 0.7 p.p. (SE=2.11) in areas with low rates of commuting by foot. In Georgia, there are no significant differences, which may be due to the lower rates of people who commute by walking overall. There are no significant differences in substitution to voting by mail by mode of transportation in either state.

This is consistent with the idea that transportation affects the cost of voting in person, but not the costs of voting by mail.

Figure 10: Average treatment effects by travel time to polling place, assuming that those within 0.5 miles walk to the polls and all others drive



Note: This figure shows binned RD estimates. The RD sample is divided into quartiles based on the average travel time to polling place associated with an RD point, where we assume that voters within 0.5 miles of the polling place walk instead of drive. The travel time to polling place associated with an RD point is the midpoint between the mean travel time to polling place in the control precinct and the mean travel time to polling place in the treatment precinct. The y-axis is measured in percentage points of likelihood of voting per minute of travel time to polling place. The x-axis shows the average travel time of each sub-sample. Vertical lines indicate 95% confidence intervals. Standard errors allow for clustering at the RD point level.

6 Turnout-maximizing Polling Place Locations

In the previous empirical sections, we estimate the effect of a one mile increase in distance to polling place on the likelihood of voting in person, by mail, and overall. From these estimates alone, it is difficult to understand how polling place location choices influence turnout in an election. For instance, any time a polling place is moved, the distance to polling place increases for some voters and decreases for others. To understand the importance of polling place locations for turnout in aggregate, we must choose a counterfactual allocation of polling places. With the goal of understanding how important polling place locations are for voter participation, we use turnout-maximizing polling place locations as the relevant counterfactual.[42] This means we abstract from the relative costs and benefits of voters participating in person or by mail. Since the effect of distance to polling place on the overall likelihood of voting is null in Georgia and negative in Pennsylvania, we do this exercise for Pennsylvania only.

We explore three scenarios for counterfactual polling place allocations. First, we find the turnout-maximizing latitude and longitude in each precinct. Second, we take existing infrastructure into account and identify the turnout-maximizing location among existing public buildings in a precinct. Third, we consider doubling the number of polling locations precincts, and find the two turnout-maximizing public buildings per precinct.

For these counterfactual exercises, we first estimate parameters of a model of vote choice. We estimate a logit model using only the sample of voters within 0.05 miles of a precinct border. We include a cubic polynomial of distance to polling place, border fixed effects, and covariates. The estimated marginal effects are consistent with the nonlinear RD estimates in Figure 5 (see Online Appendix K for details).

Next, we specify a planner's problem to maximize turnout by choosing the latitude and longitude of a polling place for a given precinct. We numerically solve for the optimal polling place location for each precinct using information about the existing allocation of voters and the estimates from the logit model. Note that, due to non-linear effects, the distance-minimizing polling place generally differs from the turnout-maximizing polling place. From this exercise, we can compare the distances between existing and optimal polling places and the votes gained by implementing the turnout-maximizing polling places.

The gains to voter participation from optimizing polling places may be over-stated from the first exercise, since many precincts have only a handful of buildings that are suitable candidate polling place locations. We construct a new dataset of polling place candidates by combining the locations of all public schools, libraries, court houses, police stations, and places of worship.[43] We also include the existing polling place as a candidate. The number of polling place candidates per precinct ranges from 1 to 18, with a median value of 2. For 30% of precincts, the only known candidate

[42] This objective is also more closely aligned with the stated objective of local election officials who choose where to put polling places. Although there are concerns that local election officials may have partisan aims, Ferrer et al. (2023) find no effect of the partisan make-up of a county election commission on the county's turnout or vote shares.

[43] These data are available from the Homeland Infrastructure Foundation-Level Data (HIFLD) of the Geospatial Management Office of the U.S. Department of Homeland Security, available at: https://hifld-geoplatform.opendata.arcgis.com/.

building is the existing polling place. We compute the simulated level of turnout for each candidate public building and select the turnout-maximizing building. For the third exercise, we find the turnout-maximizing pair of public buildings. Precincts with only 1 public building are excluded from this exercise.

6.1 Results

Figure 11 shows the locations of existing and optimal polling places, both unconstrained and limited to public buildings, for all precincts in Pennsylvania, as well as for a comparative urban and rural area within the same county. This rough look at the raw data and our computed optimal polling places shows that the counterfactual locations are not too far off from existing polling places. Although the computation of optimal polling locations does not constrain the location to be within existing precinct boundaries, we find that the optimal polling place is located within the same voting precinct as the existing polling place 62.2% of the time.[44] This suggests that there may be gains to re-precincting, that is to re-draw precincts so that they include the optimal polling place.[45]

The average distance between the current polling place and the optimal polling location is 0.82 miles, while the average distance to the optimal public building is only 0.3 miles.[46] This reflects the fact that there are some precincts in which the optimal polling location differs from the existing polling location, but where the optimal public building coincides with the existing polling location.

The distribution of gains in turnout from implementing turnout-maximizing polling places are shown in Figure 12. The predicted gains to using optimal polling place locations are small, on average. The mean increase in votes is 12 votes per precinct, equivalent to a 1.25 percentage point increase in turnout (1.58%).[47] The relatively small magnitude of the change in turnout is perhaps not surprising given that most polling places are located near the optimal polling place and that voter sensitivity to a

[44] We drop 3 precincts (0.03% of all precincts) from this analysis. Two precincts were excluded where the distance between optimal public and existing polling place was greater than 35 miles, due to the high likelihood of geocoding errors for some voters in the precinct. In addition, we exclude one precinct that has a single registered voter.

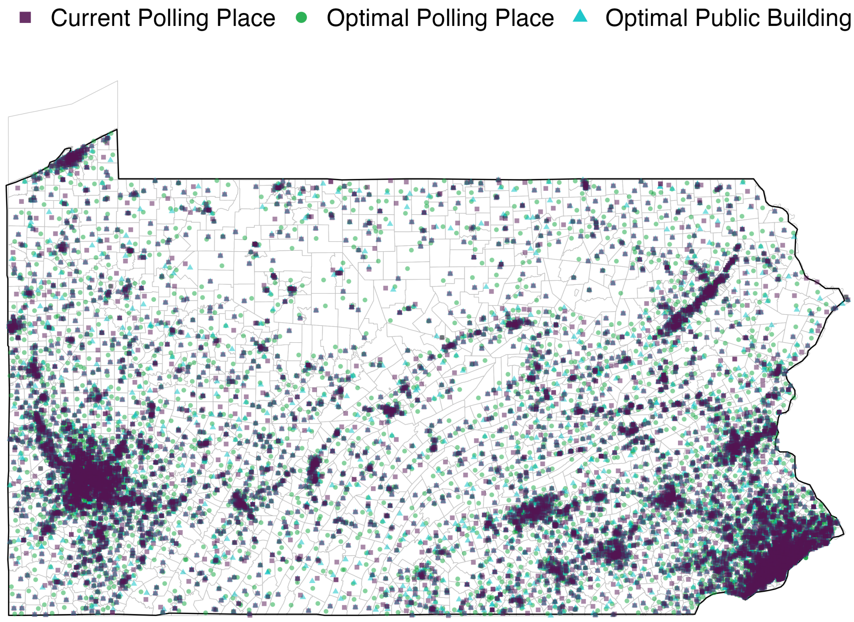
[45] Cantoni (2020) considers the optimal re-precincting problem, given existing locations of polling places. There, the mean distance to polling place is 10% smaller (0.04 miles) under the efficient re-precincting. Given that precinct boundaries change relatively infrequently, typically as part of a decennial redistricting process, we think that the optimal polling location problem is also policy-relevant. Cantoni also considers a distance-minimization problem, whereas we consider a turnout-maximization problem.

[46] The corresponding median distances are 0.34 miles and 0 miles, respectively. A median of 0 miles implies that, for the majority of precincts, there is no benefit in switching to an alternative public building from their original polling place.

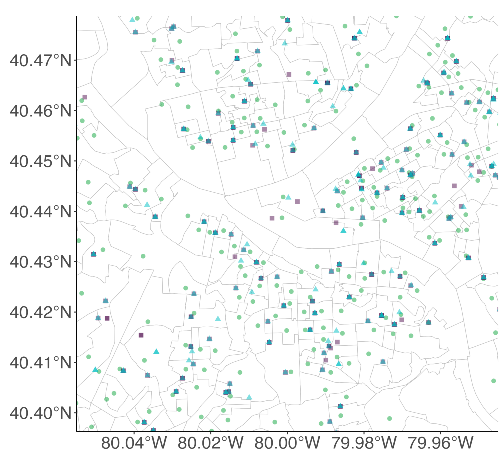
[47] We use the estimated turnout for both the existing and optimal polling places in order to compute percent changes. The average estimated turnout is 79% of voting-age population.

Figure 11: Optimal and Existing polling place locations in Pennsylvania

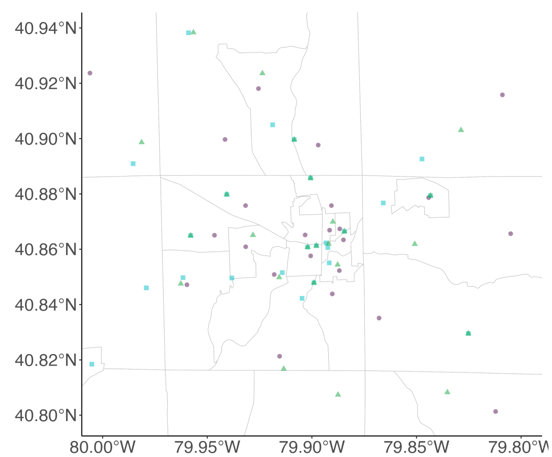
(a) All of Pennsylvania



(b) Urban area in Allegheny County, PA



(c) Rural area in Allegheny County, PA



Note: This figure shows the locations of optimal and existing polling places. Panel A shows a map of the state of Pennsylvania, with current polling locations indicated in red and optimal polling places indicated in blue. The gray lines show precinct borders.

change in distance to polling place is also relatively small. The gains to turnout from using the optimal public building are smaller, as expected. The mean increase in votes is 3.91 per precinct, equivalent to a 0.34 p.p. increase in turnout (0.43%). In 57% of precincts, the current polling place is the optimal choice among known candidate buildings. If, however, the number of polling places per precinct doubles, the mean increase in votes is 32.43 per precinct, equivalent to a 2.82 p.p. increase in turnout (3.57%).

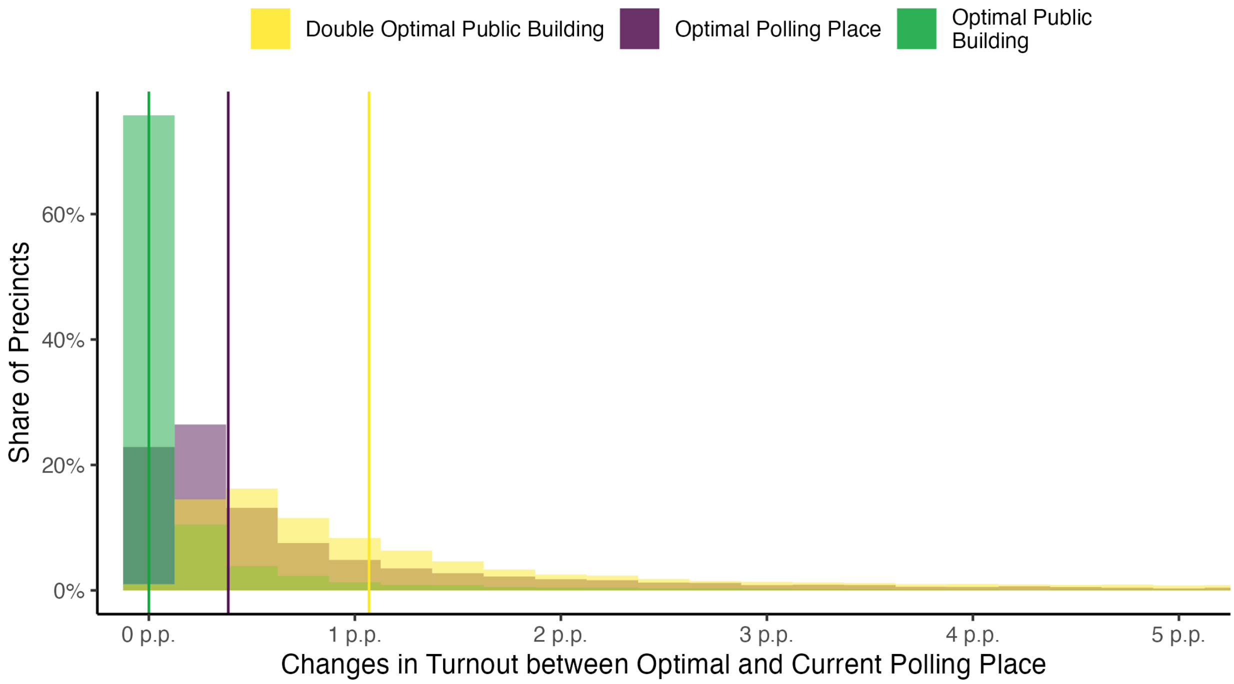
According to these simulations, if Pennsylvania were to implement all of the optimal polling places, then turnout would increase by 1.33 p.p., which is 1.68% or 110,179 votes. If they were to use all optimal public buildings, then turnout would increase by 0.43 p.p., equivalent to 0.55% or 35,907 votes. Thus, there are modest gains to turnout from implementing the turnout-maximizing polling places statewide, holding the number of polling places fixed. By doubling the number of polling places, the total increase in turnout would be 2.6 p.p., equivalent to a 3.25% increase or 213,045 more votes.

For perspective, it helps to compare these counterfactual exercises to other voter mobilization tactics and election policies. In a meta-analysis of 147 field experiments in which eligible voters receive mail encouraging them to vote, the average effect on turnout per mailer is 0.16 p.p. (Green et al. 2013). Kaplan and Yuan 2020 estimate that an additional day of early voting increases turnout by 0.22 percentage points. In swing states like Pennsylvania, these small gains in turnout could be meaningful given the extremely close elections in recent years.

The gains from doubling the number of polling places are on par with more successful voter mobilization tactics. A large-scale field experiment in the U.S., mailings that add elements of social pressure (e.g., the voting history of the resident as well as of their neighbors) increase turnout by 2.2 p.p. (Gerber et al. 2017). The average effect of canvassing across 71 studies is 2.54 p.p. (Green et al. 2013). Finally, Enos and Fowler (2018) estimate that large scale presidential campaigns, including all advertising and canvassing, can increase turnout by 7-8 p.p. in the most heavily targeted states.

While doubling the number of polling places could be costly, we can also use counterfactual simulations to find a small number of precincts with large gains to re-optimizing existing polling place locations. The distributions of gains to turnout in Figure 12 are heavily skewed. Among the 92 precincts in the 99th percentile of gains from implementing the optimal public building, the average increase in turnout would be 13.1 p.p. (16.58%). Compared to other voter mobilization efforts, relocating or increasing the number of polling places could be a cost-effective way to improve voter participation.

Figure 12: Distribution of differences in turnout between turnout-maximizing polling places and existing polling place



Note: This figure presents the distribution of turnout differences between the current polling place and the optimal polling location at the precinct level. Median values are marked by vertical lines. To enhance readability, the histograms omit the top 0.1% of observations. The greatest distance between the current and optimal polling locations is 4.99 miles, whereas the maximum distance between the optimal public and current polling locations is 13.35 miles.

7 Conclusion

We study the causal effect of distance to polling place on voter participation and voting method, at polls or by mail, in two large swing states. On average, there is a small negative effect of distance to polling place on the likelihood that a registered voter goes to the poll to vote. A standard deviation increase in distance to polling place reduces the likelihood of voting in person by 0.6 and 0.5 percentage points in Pennsylvania and Georgia, respectively. In Georgia, voters substitute to voting by mail, such that there is no net effect of distance to polling place on the likelihood of voting. In Pennsylvania, substitution to mail-in voting is limited and there is a negative net effect of distance to polling place on the likelihood of voting. An important difference between the states in the 2018 elections is that voters needed no excuse to vote by mail in Georgia and did need an excuse to vote by mail in Pennsylvania.

In our preferred specification, we adapt a geographic RD methodology for the case of a continuous treatment variable, where the change in treatment is not necessarily monotonic at the discontinuity. Our application is useful for future studies of voting costs, which might similarly exploit election precinct boundaries in other settings, and for research involving geographic discontinuities more broadly.

We find that effects are non-linear, with voters closest to polling places being most sensitive to an increase in costs of voting, likely due to differences in mode of transportation. Low income areas and areas with lower educational attainment are associated with a relatively smaller take-up of mail-in voting as distance to polling place increases, which corresponds to larger declines in overall turnout. Together, the non-linear effects and other treatment effect heterogeneity suggest that any analysis of costs of voting should be context specific.

The results highlight some important lessons for studies of electoral design and voter participation in the future. First, it is important to use large datasets to study costs of voting in large elections, where power is needed to detect very small effects. The ability to estimate small effects with precision is especially important in settings where we expect close margins of victory. In the 2020 presidential election, the margin of victory for Joe Biden was less than one percentage point in both Georgia and Pennsylvania. Second, it is important to take context into account when determining the effect of any election design change. Cantoni and Pons (2022) show that a voter's context and location influences whether or not they vote. This paper adds that voters' sensitivity to costs of voting is also highly context-dependent.

The findings in this paper can help election commissions that face costly trade-offs in choosing how many polling places to open and where to place them. Our counterfactual exercise shows that the causal estimates from our paper are not merely of academic interest but are policy-relevant if local officials want a practical way of increasing turnout. In Pennsylvania, using the set of turnout-maximizing public buildings for polling places could increase turnout by 0.55% in aggregate, with some precincts experiencing substantial gains to turnout of up to 16.6%. In future work, one might take into account the differential costs of voting in person and by mail and of opening and closing polling places, in order to determine the cost efficiency of implementing turnout-maximizing polling place locations. With a more nuanced understanding of when voters choose to vote in person, vote by mail, or abstain, election commissions have an opportunity to reduce costs of voting or to make costs of voting more equal across the population.

References

Alipour, J.-V. and V. Lindlacher (2023). No surprises, please: Voting costs and electoral turnout. *CESifo Working Paper, No. 9759*.

Amos, B., D. A. Smith, and C. Ste. Claire (2017). Reprecincting and Voting Behavior. *Political Behavior 39*(1), 133–156.

Angrist, J. D. and J.-S. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton university press.

Ansolabehere, S., B. L. Fraga, and B. F. Schaffner (2021). The CPS Voting and Registration Supplement Overstates Minority Turnout. *Journal of Politics*, Forthcoming.

Avery, J. M. (2015). Does Who Votes Matter? Income Bias in Voter Turnout and Economic Inequality in the American States from 1980 to 2010. *Political Behavior 37*(4), 955–976.

Bentele, K. G. and E. E. O'Brien (2013). Jim crow 2.0? why states consider and adopt restrictive voter access policies. *Perspectives on Politics 11*(4), 1088–1116.

Braconnier, C., J. Y. Dormagen, and V. Pons (2017). Voter Registration Costs and Disenfranchisement: Experimental Evidence from France. *American Political Science Review 111*(3), 584–604.

Brady, H. E. and J. E. McNulty (2011). Turning out to Vote: The Costs of Finding and Getting to the Polling Place. *American Political Science Review 105*(1), 115–134.

Burden, B. C., D. T. Canon, K. R. Mayer, and D. P. Moynihan (2017). The complicated partisan effects of state election laws. *Political Research Quarterly 70*(3), 564–576.

Cantoni, E. (2020). A Precinct Too Far: Turnout and Voting Costs. *American Economic Journal: Applied Economics 12*(1), 61–85.

Cantoni, E. and V. Pons (2022). Does context outweigh individual characteristics in driving voting behavior? evidence from relocations within the united states. *American Economic Review 112*(4), 1226–72.

Card, D., D. S. Lee, Z. Pei, and A. Weber (2015). Inference on causal effects in a generalized regression kink design. *Econometrica 83*(6), 2453–2483.

Cattaneo, M. D., M. Jansson, and X. Ma (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association 115*(531), 1449–1455.

Chen, M. K., K. Haggag, D. G. Pope, and R. Rohla (2020). Racial Disparities in Voting Wait Times: Evidence from Smartphone Data. *Review of Economics and Statistics*, Forthcoming.

Clinton, J. D., N. Eubank, A. Fresh, and M. E. Shepherd (2020). Polling Place Changes and Political Participation: Evidence from North Carolina Presidential Elections, 2008-2016 *. *Political Science Research and Methods First View*, 1–18.

Dell, M. (2010). The persistent effects of peru's mining mita. *Econometrica 78*(6), 1863–1903.

Dyck, J. J. and J. G. Gimpel (2005). Distance, turnout, and the convenience of voting. *Social Science Quarterly 86*(3), 531–548.

Enos, R. D. and A. Fowler (2018). Aggregate effects of large-scale campaigns on voter turnout. *Political Science Research and Methods 6*(4), 733–751.

Ferrer, J., I. Geyn, and D. M. Thompson (2023). How partisan is local election administration? *American Political Science Review*, 1–16.

Fraga, B. L. (2018). *The Turnout Gap.* Cambridge University Press.

Fujiwara, T., K. Meng, and T. Vogl (2016). Habit formation in voting: Evidence from rainy elections. *American Economic Journal: Applied Economics 8*(4), 160–188.

Gerber, A. S., G. A. Huber, A. H. Fang, and A. Gooch (2017). The generalizability of social pressure effects on turnout across high-salience electoral contexts: Field experimental evidence from 1.96 million citizens in 17 states. *American Politics Research 45*(4), 533–559.

Gomez, B. T., T. G. Hansford, and G. A. Krause (2007). The republicans should pray for rain: Weather, turnout, and voting in U.S. presidential elections. *Journal of Politics 69*(3), 649–663.

Green, D. P., M. C. McGrath, and P. M. Aronow (2013). Field experiments and the study of voter turnout. *Journal of Elections, Public Opinion and Parties 23*(1), 27–48.

Hasen, R. L. (2012). *The Voting Wars.* Yale University Press.

Haspel, M. and H. Gibbs Knotts (2005). Location, location, location: Precinct placement and the costs of voting. *Journal of Politics 67*(2), 560–573.

Highton, B. (2017). Voter identification laws and turnout in the united states. *Annual Review of Political Science 20*(1), 149–167.

Imbens, G. and T. Zajonc (2023). Regression discontinuity design with multiple forcing variables. *Working Paper*.

Kaplan, E. and H. Yuan (2020). Early voting laws, voter turnout, and Partisan vote composition: Evidence from Ohio. *American Economic Journal: Applied Economics 12*(1), 32–60.

Keele, L. J. and R. Titiunik (2015). Geographic boundaries as regression discontinuities. *Political Analysis 23*(1), 127–155.

Leighley, J. E. and J. Nagler (2014). *Who votes now?* Princeton University Press.

Lockhart, M., S. J. Hill, J. Merolla, M. Romero, and T. Kousser (2020). America's electorate is increasingly polarized along partisan lines about voting by mail during the covid-19 crisis. *Proceedings of the National Academy of Sciences 117*(40), 24640–24642.

Martin, P. S. (2003). Voting's rewards: Voter turnout, attentive publics, and congressional allocation of federal money. *American Journal of Political Science 47*(1), 110–127.

Meredith, M. and Z. Endter (2016). Aging into absentee voting: Evidence from texas. *Working Paper*.

Meredith, M. and N. Malhotra (2011). Convenience voting can affect election outcomes. *Election Law Journal 10*(3), 227–253.

Milligan, K., E. Moretti, and P. Oreopoulos (2004). Does education improve citizenship? evidence from the united states and the united kingdom. *Journal of Public Economics 88*(9), 1667 – 1695.

Plutzer, E. (2002). Becoming a Habitual Voter: Inertia, Resources, and Growth in Young Adulthood. *The American Political Science Review 96*(1), 41–56.

Shepherd, M. E., A. Fresh, N. Eubank, and J. D. Clinton (2021). The Politics of Locating Polling Places: Race and Partisanship in North Carolina Election Administration, 2008–2016. *Election Law Journal: Rules, Politics, and Policy 20*(2), 155–177.

Shino, E., M. Suttmann-Lea, and D. A. Smith (2021). Determinants of rejected mail ballots in georgia's 2018 general election. *Political Research Quarterly 75*(1), 231–243.

Sondheimer, R. M. and D. P. Green (2010). Using experiments to estimate the effects of education on voter turnout. *American Journal of Political Science 54*(1), 174–189.

Thompson, D. M., J. A. Wu, J. Yoder, and A. B. Hall (2020). Universal vote-by-mail has no impact on partisan turnout or vote share. *Proceedings of the National Academy of Sciences of the United States of America 117*(25), 14052–14056.

Tomkins, S., K. Yao, J. Gaebler, T. Konitzer, D. Rothschild, M. Meredith, and S. Goel (2023). Blocks as geographic discontinuities: The effect of polling-place assignment on voting. *Political Analysis 31*(2), 165–180.

Wolfinger, N. and W. Raymond (2008). Family Structure and Voter Turnout. *Social Forces 86*(5), 1863–1903.

Yoder, J. (2018). How Polling Place Changes Reduce Turnout: Evidence from Administrative Data in North Carolina. *Working Paper*.

A Appendix

A.1 Summary Statistics

Table A.1: Summary Statistics

	Pennsylvania				Georgia			
	All		RD sample		All		RD sample	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Voting History								
Primary election, at polls	18.83	39.10	18.87	39.12	12.26	32.80	12.38	32.93
Primary election, absentee	0.39	6.22	0.40	6.32	4.15	19.94	4.20	20.06
Primary election, total	19.22	39.41	19.27	39.44	16.41	37.04	16.58	37.19
General election, at polls	56.28	49.60	56.21	49.61	26.06	43.90	26.02	43.88
General election, absentee	2.08	14.27	2.10	14.35	28.10	44.95	27.98	44.89
General election, total	58.36	49.30	58.31	49.30	54.17	49.83	54.00	49.84
Distance								
Travel route distance to polling place (mi)	0.74	0.79	0.76	0.88	1.62	1.17	1.53	1.15
Euclidean distance to polling place (mi)	0.47	0.52	0.50	0.60	1.01	0.79	1.00	0.78
Travel time to polling place (min)	2.39	1.79	2.37	1.97	3.87	2.07	3.71	2.13
Demographics								
Democrat	0.56	0.50	0.55	0.50	0.10	0.30	0.10	0.30
Republican	0.30	0.46	0.31	0.46	0.06	0.24	0.06	0.24
Independent	0.14	0.34	0.14	0.35	0.00	0.05	0.00	0.05
Percent Female	0.52	0.07	0.52	0.07	0.52	0.07	0.52	0.07
Percent urban	0.98	0.14	0.98	0.13	0.96	0.19	0.97	0.17
Percent Black	0.17	0.28	0.15	0.27	0.38	0.35	0.38	0.35
Percent White	0.75	0.31	0.77	0.30	0.52	0.34	0.51	0.34
Percent Hispanic	0.06	0.14	0.06	0.14	0.08	0.12	0.08	0.12
Median hh Income (10k USD)	5.54	2.99	5.55	2.94	6.06	3.18	5.97	3.19
Percent without a HS degree	0.12	0.11	0.12	0.11	0.39	0.13	0.39	0.14
Percent that walk to work	0.05	0.10	0.05	0.10	0.02	0.05	0.02	0.05
Cars per household	1.46	0.48	1.49	0.46	1.70	0.42	1.68	0.42
N	4,223,610		2,315,502		2,024,221		1,227,157	

Note: For each voting history variable, we observe whether or not a registered voter votes, by method of voting. We report these variables in percentage points to make regression coefficients easier to interpret. Demographic variables are measured at the block, or block-group level and assigned to each individual voter that resides in the geographic area. The RD sample is the Regression Discontinuity sample, described in Section 4.1.

A.2 Manipulation and Balance Tests for RD Specification

In this section we estimate RD coefficients for covariates, which serve as placebo outcomes (Tables A.2 and A.3). We find no strong evidence against the identifying assumption that confounding factors are continuous at the RD point. In each state we estimate the RD coefficient for twelve placebo outcomes. There is only a statistically significant change at the discontinuity for Age in Pennsylvania.

We use local polynomial density estimation and inference with the Stata package rddensity to evaluate the distribution of the running variable near the discontinuity. Figure A.1 shows histograms of distributions of the running variable, distance to the RD point in miles, for each state. We show the distribution for 3 times the optimal bandwidth (the default) and within the optimal bandwidth used in Table 1. The distributions are highly non-linear. This is because an RD point is a midpoint of a precinct border, so it is not generally located at a residential address. As a result, point estimates fall outside of bias-corrected Confidence Intervals proposed by Cattaneo et al. (2020). This is true even if we use a fourth-order polynomial to estimate the density function and if we use a bandwidth as small as one-fourth of the MSE-optimal bandwidth. We therefore report the conventional T-statistic for the hypothesis that density functions are equal at the discontinuity, rather than bias corrected estimates. Note that the bias-corrected version of the test is designed to correct for *over*-rejection of the null hypothesis under the conventional test. Given the non-linearities in the distributions, we prefer to use a narrower bandwidth around the discontinuity. Using a bandwidth of 0.03 and 0.07 in Pennsylvania and Georgia, we fail to reject the null hypothesis that the density of distance to the polling place is continuous at the cutoff, with p-values of 0.07 in Pennsylvania and 0.47 in Georgia (Table A.4).

Table A.2: RD Estimates for placebo outcomes: Pennsylvania

	Female	Age	Democrat	Republican	Median HH Income	Urban
RD Estimate	0.0022	-0.6362**	0.0013	-0.0003	0.0095	-0.0006
	(0.0041)	(0.2999)	(0.0038)	(0.0031)	(0.0321)	(0.0006)
N	1,881,131	2,313,123	2,315,562	2,315,562	2,310,685	2,277,884
Outcome mean	0.46	50.16	0.55	0.31	5.55	0.98
Bandwidth	0.10	0.10	0.10	0.10	0.10	0.10

	Percent White	Percent Black	Percent Hispanic	Percent with No HS Degree	Percent that Walk to Work	Cars per Household
RD Estimate	-0.0034	-0.0026	0.0016	0.0002	-0.0006	0.0067
	(0.0039)	(0.0031)	(0.0026)	(0.0021)	(0.0020)	(0.0054)
N	2,277,884	2,277,884	2,277,884	2,312,987	2,315,305	2,011,158
Outcome mean	0.77	0.15	0.06	0.12	0.05	1.49
Bandwidth	0.10	0.10	0.10	0.10	0.10	0.10

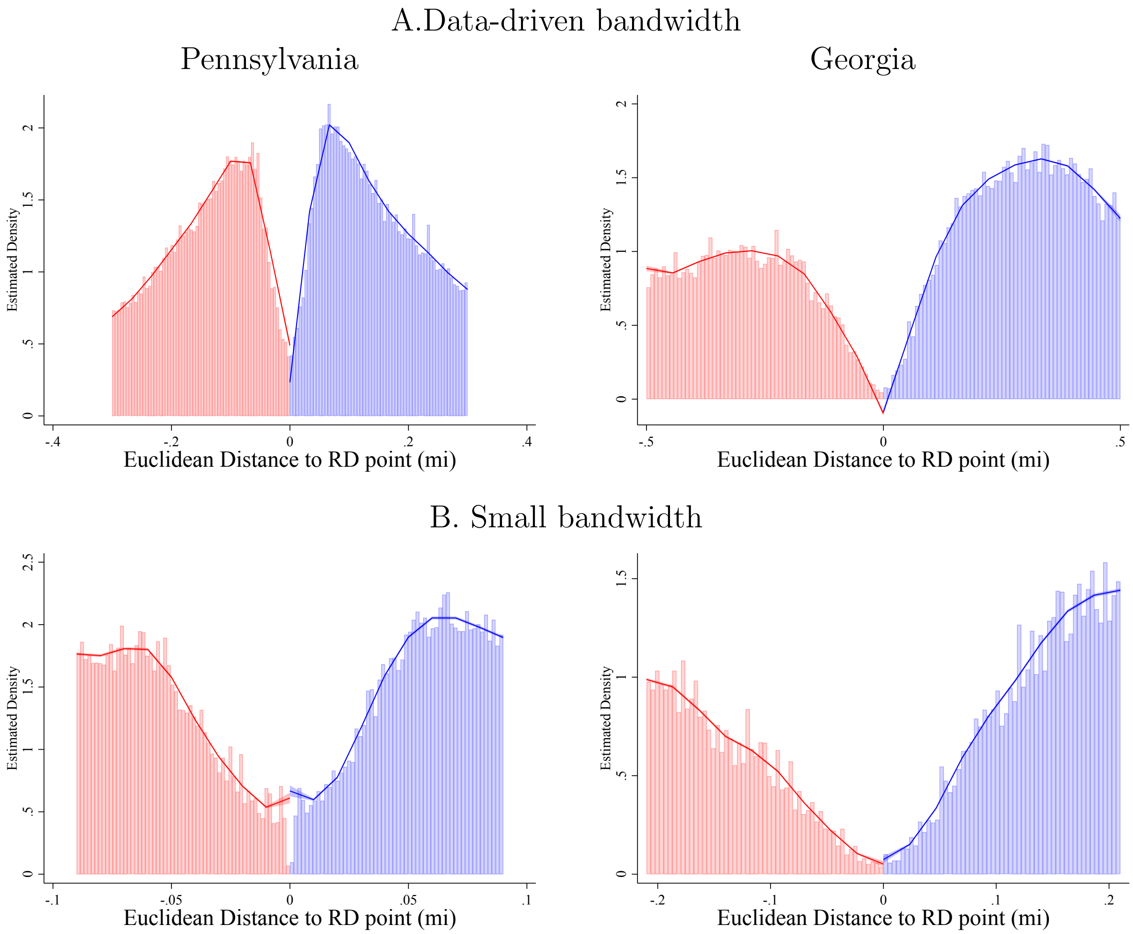
Note: This table reports the RD estimates for placebo outcomes. As in the main RD analysis, we use residualized outcomes, after removing RD point fixed effects, to improve precision. Female, Age, Democrat, and Republican are observed at the individual-level from voter registration files. Median household (HH) income, urban, race/ethnicity, percent with no high school (HS) degree, percent that walk to work, and cars per household are observed at the Census block or block-group level. Bandwidths are MSE-optimal. Standard errors allow for clustering at the RD point level.

Table A.3: RD Estimates for placebo outcomes: Georgia

	Female	Age	Democrat	Republican	Median HH Income	Urban
RD Estimate	-0.0080	-1.3687	-0.0017	-0.0017	-0.0495	-0.0048
	(0.0073)	(0.8344)	(0.0058)	(0.0039)	(0.1293)	(0.0030)
N	1,224,732	1,226,654	1,227,157	1,227,157	1,226,070	1,198,637
Outcome mean	0.54	45.49	0.10	0.06	5.97	0.97
Bandwidth	0.20	0.20	0.20	0.20	0.20	0.20
	Percent White	**Percent Black**	**Percent Hispanic**	**Percent with No HS Degree**	**Percent that Walk to Work**	**Cars per Household**
RD Estimate	-0.0199	0.0187	-0.0024	0.0003	-0.0007	0.0050
	(0.0142)	(0.0131)	(0.0052)	(0.0056)	(0.0023)	(0.0192)
N	1,198,637	1,198,637	1,198,637	1,227,157	1,227,157	1,081,886
Outcome mean	0.51	0.38	0.08	0.39	0.02	1.68
Bandwidth	0.20	0.20	0.20	0.20	0.20	0.20

Note: This table reports the RD estimates for placebo outcomes. As in the main RD analysis, we use the residualized outcomes, after removing RD point fixed effects, to improve precision. Female, Age, Democrat, and Republican are observed at the individual-level from voter registration files. Median household (HH) income, urban, race/ethnicity, percent with no high school (HS) degree, percent that walk to work, and cars per household are observed at the Census block or block-group level. Bandwidths are MSE-optimal. Standard errors allow for clustering at the RD point level.

Figure A.1: Manipulation tests: Estimated of density of running variables around the RD point



Note: These figures show histograms and estimated kernel densities of the running variable, which is the Euclidean distance from the voter to the RD point of the nearest border segment (mi). The values are negative for the control precinct (shaded in red) and positive for the treatment precinct (shaded in blue).

Table A.4: Manipulation Tests

	Data-driven bandwidth		Small bandwidth	
	Pennsylvania	Georgia	Pennsylvania	Georgia
Bandwidth	0.10	0.20	0.03	0.07
Polynomial for estimation	4	4	4	4
Effective N of obs., left	305993	126,380	43,871	58,772
Effective N of obs., right	350623	198,506	48,870	101,171
T-statistic for density test	-13.25	0.82	1.82	-0.73
P-value for density test	$< .001$	0.41	0.07	0.47

A.3 Balance for Fixed Effects Regression Specifications

Table A.5: Correlations between covariates and distance to polling place, including border fixed effects

A. Pennsylvania

	Female	Age	Democrat	Republican	Median HH Income	Urban
Distance to polling place (mi)	-0.0024**	-0.2302**	-0.0029**	0.0024*	0.0317	-0.0126***
	(0.0011)	(0.0973)	(0.0014)	(0.0014)	(0.0198)	(0.0025)
N	1,244,682	1,244,682	1,244,682	1,244,682	1,244,682	1,244,682
Outcome mean	0.47	46.63	0.64	0.22	4.59	0.97
R^2	0.02	0.13	0.14	0.16	0.89	0.78

	Percent White	Percent Black	Percent Hispanic	Percent with No HS Degree	Percent that Walk to Work	Cars per Household
Distance to polling place (mi)	-0.0009	-0.0011	-0.0003	-0.0000	-0.0007	0.0046
	(0.0017)	(0.0016)	(0.0007)	(0.0009)	(0.0006)	(0.0035)
N	1,244,682	1,244,682	1,244,682	1,244,682	1,244,682	1,244,682
Outcome mean	0.63	0.26	0.09	0.15	0.07	1.28
R^2	0.92	0.93	0.85	0.85	0.87	0.92

B. Georgia

	Female	Age	Democrat	Republican	Median HH Income	Urban
Distance to polling place (mi)	-0.0020**	-0.2396***	-0.0030***	-0.0007	0.0415**	-0.0187***
	(0.0008)	(0.0843)	(0.0008)	(0.0009)	(0.0170)	(0.0031)
N	389,902	389,902	389,902	389,902	389,902	389,902
Outcome mean	0.54	45.91	0.09	0.07	5.52	0.84
R^2	0.01	0.09	0.05	0.08	0.86	0.85

	Percent White	Percent Black	Percent Hispanic	Percent with No HS Degree	Percent that Walk to Work	Cars per Household
Distance to polling place (mi)	0.0013	0.0001	-0.0020**	-0.0016*	-0.0004	0.0136***
	(0.0022)	(0.0021)	(0.0010)	(0.0009)	(0.0003)	(0.0028)
N	389,902	389,902	389,902	389,902	389,902	389,902
Outcome mean	0.55	0.36	0.07	0.14	0.02	1.75
R^2	0.79	0.81	0.58	0.80	0.73	0.84

Note: This table reports coefficients for the travel route distance to polling place (mi). The outcome variables are the individual-level and census block-level covariates listed in each column. Each regression includes border segment fixed effects. Standard errors allow for clustering at border segment-level.

Table A.6: Correlations between covariates and distance to polling place, including matched pair fixed effects

A. Pennsylvania

	Female	Age	Democrat	Republican	Median HH Income	Urban
Distance to polling place (mi)	0.003	-0.181	-0.002	0.004	0.050*	-0.014***
	(0.006)	(0.170)	(0.006)	(0.004)	(0.029)	(0.003)
N	1,868,442	1,868,442	1,868,442	1,868,442	1,868,442	1,868,442
Outcome mean	0.47	46.32	0.64	0.21	4.58	0.98
R^2	0.09	0.18	0.21	0.23	0.83	0.74

	Percent White	Percent Black	Percent Hispanic	Percent with No HS Degree	Percent that Walk to Work	Cars per Household
Distance to polling place (mi)	-0.004	0.003	0.001	0.000	-0.001	0.006
	(0.004)	(0.004)	(0.001)	(0.001)	(0.001)	(0.005)
N	1,868,442	1,868,442	1,868,442	1,868,442	1,868,442	1,868,442
Outcome mean	0.63	0.26	0.09	0.15	0.07	1.27
R^2	0.88	0.89	0.81	0.80	0.82	0.87

B. Georgia

	Female	Age	Democrat	Republican	Median HH Income	Urban
Distance to polling place (mi)	-0.001	-0.095	-0.001	-0.002	0.020	-0.015***
	(0.002)	(0.123)	(0.002)	(0.002)	(0.015)	(0.002)
N	654,424	654,424	654,424	654,424	654,424	654,424
Outcome mean	0.54	47.28	0.09	0.08	5.46	0.84
R^2	0.04	0.11	0.09	0.13	0.84	0.85

	Percent White	Percent Black	Percent Hispanic	Percent with No HS Degree	Percent that Walk to Work	Cars per Household
Distance to polling place (mi)	-0.000	0.001	-0.001	-0.001	-0.000	0.010***
	(0.003)	(0.003)	(0.001)	(0.001)	(0.000)	(0.002)
N	654,424	654,424	654,424	654,424	654,424	654,424
Outcome mean	0.56	0.35	0.07	0.15	0.02	1.75
R^2	0.76	0.79	0.54	0.80	0.67	0.83

Note: This table reports coefficients for the travel route distance to polling place (mi). The outcome variables are the individual-level and census block-level covariates listed in each column. Each regression includes fixed effects for matched pairs of voters. Standard errors allow for clustering at precinct and segment-level.