Online Appendix
Polling Place Location and the Costs of Voting

Gaurav Bagwe Juan Margitic Allison Stashko

November 22, 2023

Contents

A Data Appendix

Table A.1: Variable Definitions, Units of Observation, and Data Sources

Variable	Definition	Unit of Observation	Source
Turnout	Votes per voting age population	Block	PA Secretary of State
Distance to Polling Place	Miles from block interior centroid to polling place	Block	Computed value
Race, Ethnicity, Gender, Age	Percent of population in demographic group	Block	2010 Census
Car Ownership	Number of cars per housing units	Block-group	2006-2010 ACS
Way to Work	Percent of workers 16 and older using mode of transportation to work	Block-group	2006-2010 ACS
Time to Work	Time to work among workers 16 and older who do not work from home	Block-group	2006-2010 ACS
Median Income	Median household income for the past 12 months	Block-group	2006-2010 ACS
Home Ownership	Percent of households owning home	Block-group	2006-2010 ACS
Education	Percent of population older than 25 belonging to education group	Block-group	2006-2010 ACS

B Fuzzy RD Framework

In this section we explain why using a single-dimensional regression discontinuity requires a Fuzzy RD approach. By collapsing a two-dimensional treatment assignment vector (latitude and longitude of the voter) to one dimension (distance from voter to border), we necessarily lose some information about treatment assignment. The Fuzzy RD framework applies because we can not write treatment as a function of the running variable alone, but as a function of the running variable and other (known) variables. To see this explicitly, we write the treatment variable as a function of the running variable (v_i), the distance between a voter and the RD point (w_i, as in Figure 3), and the angle between the line segments that connect a voter with the RD point and the RD point with the polling place (γ_i):

$$d_i = \sqrt{v_i^2 + w_i^2 - 2|v_i| w_i \cos(\gamma_i)}. \tag{B.1}$$

Clearly, for the same value of v_i, we can observe different values of d_i depending on w_i and γ_i. Thus, by collapsing the RD to one dimension, we lose some information about treatment assignment, but we gain by fitting the a standard fuzzy RD framework which means we can plot the RD results and estimate the RD in a transparent, data-driven manner.

Note that, as v_i tends to zero from the left, d_i tends to w_i, where w_i is the distance between the RD point and the polling place of the control precinct. As v_i tends to zero from the right, d_i tends to w_j, where w_j is the distance between the RD point and the polling place of the treatment precinct. By definition of the running variable, we have $w_i < w_j$. Thus, there is a discontinuity in d_i at the cutoff of $v_i = 0$. Second, the jump in distance to polling place is monotonic for voters sufficiently close to the cutoff, since $v_i < 0$ and $v_j > 0$ for voters i and j whenever $w_i < w_j$.

C Alternative Measures of Distance to Polling Place

In this section we replace travel route distance to the polling place (miles) with Euclidean distance (mi) to the polling place (Table C.1) and with travel time by car (minutes) to the polling place (Table C.2). Note that sample sizes change because we determine if an observation is a complier for each definition of distance to polling place, in line with the monotonicity condition, as discussed in Section 4.1. The estimates indicate that, in Pennsylvania, the likelihood of voting in person falls by 7.4 p.p. per Euclidean mile of distance (similar to travel route distance) and 0.9 p.p. per minute of travel time by car. In Georgia, the likelihood of voting in person falls by 2.4 p.p. per Euclidean mile of distance and 0.3 p.p. per minute of travel time by car. We prefer travel route distance to Euclidean distance since it is more accurate measure of the travel-costs associated with geographic distance to polling place. A limitation of using travel time by car is that we do not know the mode of transportation of all voters. If we assume that voters within 0.5 miles of their polling place walk instead of drive, at the standard rate of 3 miles per hour, then the estimated reduction in the likelihood of voting in person per additional minute of travel time is a 0.35 p.p. in Pennsylvania and 0.38 p.p. in Georgia.

Table C.1: RD Estimates using Euclidean Distance to Polling Place

	A. Pennsylvania			
	First-stage	Second-stage		
	Euclidean Distance (mi)	At Poll	Absentee	Total
RD Estimate	0.23***	-7.44***	0.48	-7.20***
	(0.02)	(2.23)	(0.46)	(2.17)
N	3,236,019	3,236,015	3,236,015	3,236,015
Effective N, Left	495,622	495,622	662,235	541,964
Effective N, Right	229,644	229,644	303,236	248,945
Bandwidth	0.07	0.07	0.09	0.08
Outome mean	0.49	56.27	2.09	58.36

	B. Georgia			
	First-stage	Second-stage		
	Euclidean Distance (mi)	At Poll	Absentee	Total
RD Estimate	0.70***	-2.37**	1.12	-1.17
	(0.05)	(0.98)	(1.26)	(1.60)
N	1,604,550	1,604,550	1,604,550	1,604,550
Effective N, Left	251,258	251,258	200,104	212,485
Effective N, Right	188,451	188,451	149,314	160,483
Bandwidth	0.19	0.19	0.17	0.18
Outome mean	1.02	26.01	28.12	54.14

Note: 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. Euclidean distance (mi) measures the Euclidean distance between a voter and their assigned polling place. Bandwidths are MSE-optimal. Standard errors allow for clustering at the RD point level.

Table C.2: RD Estimates using Travel Time by Car to Polling Place

	A. Pennsylvania			
	First-stage	Second-stage		
	Travel Time (min)	At Poll	Absentee	Total
RD Estimate	1.77***	-0.94***	0.07	-0.86***
	(0.11)	(0.25)	(0.06)	(0.25)
N	2,775,932	2,775,929	2,775,929	2,775,929
Effective N, Left	580,933	580,932	539,294	571,945
Effective N, Right	330,130	330,130	306,930	325,187
Bandwidth	0.11	0.11	0.10	0.10
Outome mean	2.35	56.36	2.11	58.47

	B. Georgia			
	First-stage	Second-stage		
	Travel Time (min)	At Poll	Absentee	Total
RD Estimate	2.80***	-0.31	0.56**	0.20
	(0.16)	(0.28)	(0.28)	(0.35)
N	1,460,672	1,460,672	1,460,672	1,460,672
Effective N, Left	202,362	202,362	252,058	249,463
Effective N, Right	167,634	167,634	208,463	206,046
Bandwidth	0.19	0.19	0.22	0.22
Outome mean	3.83	26.07	27.94	54.01

Note: 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. Travel time (min) measures the travel time between a voter and their assigned polling place, assuming registered voters drive at posted speed limits. Bandwidths are MSE-optimal. Standard errors allow for clustering at the RD point level.

Table C.3: RD Estimates using Travel Time to polling place, assuming voters walk if within 0.5 miles of the polling plae and drive otherwise

	A. Pennsylvania			
	First-stage	Second-stage		
	Travel Time (min)	At Poll	Absentee	Total
RD Estimate	5.28***	-0.35***	0.04**	-0.30***
	(0.28)	(0.12)	(0.02)	(0.12)
N	2,309,728	2,309,725	2,309,725	2,309,725
Effective N, Left	303,339	303,339	582,597	297,321
Effective N, Right	166,588	166,588	315,795	162,515
Bandwidth	0.07	0.07	0.11	0.07
Outome mean	7.02	55.44	1.99	57.43

	B. Georgia			
	First-stage	Second-stage		
	Travel Time (min)	At Poll	Absentee	Total
RD Estimate	6.08***	-0.38***	0.44***	0.00
	(0.49)	(0.12)	(0.16)	(0.20)
N	979,696	979,696	979,696	979,696
Effective N, Left	181,314	181,314	142,488	111,258
Effective N, Right	104,187	104,187	82,929	66,174
Bandwidth	0.21	0.21	0.19	0.17
Outome mean	6.59	25.97	27.91	53.88

Note: 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. Travel time (min) measures the travel time between a voter and their assigned polling place, assuming that voters walk at a rate of 3 miles per hour if they are within 0.5 miles of the polling place, and drive at posted speed limits otherwise. Bandwidths are MSE-optimal. Standard errors allow for clustering at the RD point level.

D Additional Election Outcomes

In this section we report RD estimates for the 2018 primary elections in Pennsylvania and Georgia, as well as for the 2016 Presidential election in Pennsylvania. We report estimates of average effects for primary elections in both states and for the 2016 general election in Pennsylvania in Table D.1. We report RD estimates for bins of RD points based on distance to polling place (non-linear estimates) in Tables D.2, D.3, and D.4.

Table D.1: Fuzzy RD Estimates: Additional Election Outcomes

	First-stage	Second-stage		
	Distance (mi)	At Poll	Absentee	Total
A. Pennsylvania, 2018 Primary				
RD Estimate	0.45***	-2.97***	-0.18	-3.03***
	(0.03)	(0.85)	(0.17)	(0.85)
N	2,315,556	2,313,437	2,313,437	2,313,437
Effective N, Left	302,570	302,342	220,261	312,562
Effective N, Right	346,713	346,396	257,840	357,087
Bandwidth	0.10	0.10	0.08	0.10
Outome mean	0.76	18.87	0.40	19.27
B. Georgia, 2018 Primary				
	First-stage	Second-stage		
	Distance (mi)	At Poll	Absentee	Total
RD Estimate	1.30***	-0.53	0.18	-0.36
	(0.09)	(0.56)	(0.30)	(0.69)
N	1,227,157	1,227,157	1,227,157	1,227,157
Effective N, Left	86,961	86,961	77,459	74,981
Effective N, Right	140,429	140,429	125,013	120,457
Bandwidth	0.17	0.17	0.16	0.15
Outome mean	1.53	12.38	4.20	16.58
C. Pennsylvania, 2016 General				
	First-stage	Second-stage		
	Distance (mi)	At Poll	Absentee	Total
RD Estimate	0.46***	-2.14*	0.30	-1.96*
	(0.03)	(1.10)	(0.27)	(1.08)
N	2,315,556	2,315,556	2,315,556	2,315,556
Effective N, Left	249,181	249,181	282,551	256,324
Effective N, Right	291,753	291,753	326,666	298,746
Bandwidth	0.09	0.09	0.09	0.09
Outome mean	0.76	67.08	2.69	69.77

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.

Table D.2: Nonlinear RD estimates: 2018 Primary Elections, Pennsylvania

	First Quantile (mean d_i=0.28, st. dev.=0.17)			Second Quantile (mean d_i=0.45, st. dev.=0.26)		
	At Poll	Absentee	Total	At Poll	Absentee	Total
RD Estimate	-8.58***	-0.43	-9.14***	-3.05*	-0.21	-2.62
	(2.85)	(0.29)	(2.89)	(1.82)	(0.44)	(1.79)
N	578,222	578,222	578,222	578,168	578,168	578,168
Outome mean	17.74	0.31	18.06	18.40	0.40	18.80
Bandwidth	0.08	0.10	0.08	0.10	0.05	0.10
	Third Quantile (mean d_i=0.76, st. dev.=0.45)			Fourth Quantile (mean d_i=1.53, st. dev.=1.37)		
	At Poll	Absentee	Total	At Poll	Absentee	Total
RD Estimate	-2.10**	-0.12	-2.57***	0.48	0.08	0.53
	(0.90)	(0.13)	(0.96)	(0.33)	(0.08)	(0.35)
N	578,141	578,141	578,141	578,793	578,793	578,793
Outome mean	19.96	0.44	20.40	19.35	0.45	19.80
Bandwidth	0.15	0.20	0.14	0.15	0.12	0.14

H_0 Quantile 1 = Quantile 2: At polls, $p-$value=0.10; Absentee, $p-$value=0.68; Total, $p-$value=0.05.
H_0 Quantile 1 = Quantile 3: At polls, $p-$value=0.03; Absentee, $p-$value=0.34; Total, $p-$value=0.03.
H_0 Quantile 1 = Quantile 4: At polls, $p-$value=0.002; Absentee, $p-$value=0.10; Total, $p-$value=0.001.

Note: For each voting history variable, we observe whether or not a registered voter votes, by method of voting. We multiply these indicator variables by 100 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.

Table D.3: Nonlinear RD estimates: 2018 Primary Elections, Georgia

	First Quantile (mean d_i=0.72, st. dev.=0.41)			Second Quantile (mean d_i=1.20, st. dev.=0.60)		
	At Poll	Absentee	Total	At Poll	Absentee	Total
RD Estimate	-4.11	-0.42	-4.63	1.77	0.62	2.22
	(2.60)	(1.41)	(2.88)	(1.44)	(0.50)	(1.58)
N	306,677	306,677	306,677	306,590	306,590	306,590
Outome mean	13.18	3.74	16.92	12.86	4.22	17.07
Bandwidth	0.14	0.12	0.14	0.15	0.22	0.16
	Third Quantile (mean d_i=1.66, st. dev.=0.79)			Fourth Quantile (mean d_i=2.55, st. dev.=1.53)		
	At Poll	Absentee	Total	At Poll	Absentee	Total
RD Estimate	-0.60	0.13	-0.51	-0.72	0.36	-0.34
	(1.06)	(0.56)	(1.47)	(0.48)	(0.31)	(0.56)
N	306,977	306,977	306,977	306,913	306,913	306,913
Outome mean	12.01	4.22	16.23	11.46	4.63	16.09
Bandwidth	0.20	0.20	0.18	0.18	0.17	0.17

H_0 Quantile 1 = Quantile 2: At polls, $p-$value=0.05; Absentee, $p-$value=0.49; Total, $p-$value=0.04.
H_0 Quantile 1 = Quantile 3: At polls, $p-$value=0.21; Absentee, $p-$value=0.72; Total, $p-$value=0.20.
H_0 Quantile 1 = Quantile 4: At polls, $p-$value=0.20; Absentee, $p-$value=0.59; Total, $p-$value=0.14.

Note: For each voting history variable, we observe whether or not a registered voter votes, by method of voting. We multiply these indicator variables by 100 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.

Table D.4: Nonlinear RD estimates: 2016 Presidential Election, Pennsylvania

	First Quantile (mean d_i=0.28, st. dev.=0.17)			Second Quantile (mean d_i=0.45, st. dev.=0.26)		
	At Poll	Absentee	Total	At Poll	Absentee	Total
RD Estimate	-6.32**	0.40	-5.94*	-1.97	0.35	-1.23
	(3.10)	(0.67)	(3.08)	(2.12)	(0.69)	(2.03)
N	578,699	578,699	578,699	578,652	578,652	578,652
Outome mean	63.91	1.78	65.69	65.47	2.45	67.91
Bandwidth	0.09	0.09	0.09	0.11	0.09	0.11
	Third Quantile (mean d_i=0.76, st. dev.=0.45)			Fourth Quantile (mean d_i=1.53, st. dev.=1.37)		
	At Poll	Absentee	Total	At Poll	Absentee	Total
RD Estimate	-1.04	0.01	-1.04	-0.06	0.26**	0.18
	(1.39)	(0.40)	(1.45)	(0.33)	(0.13)	(0.37)
N	578,698	578,698	578,698	579,394	579,394	579,394
Outome mean	68.98	3.04	72.02	69.96	3.49	73.44
Bandwidth	0.13	0.15	0.12	0.17	0.13	0.15

H_0 Quantile 1 = Quantile 2: At polls, $p-$value=0.20; Absentee, $p-$value=0.34; Total, $p-$value=0.18.
H_0 Quantile 1 = Quantile 3: At polls, $p-$value=0.06; Absentee, $p-$value=0.48; Total, $p-$value=0.03.
H_0 Quantile 1 = Quantile 4: At polls, $p-$value=0.02; Absentee, $p-$value=0.10; Total, $p-$value=0.01.

Note: For each voting history variable, we observe whether or not a registered voter votes, by method of voting. We multiply these indicator variables by 100 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.

E Nonlinear Effects

Table E.1: Fuzzy RD Estimates: Nonlinear effects in Pennsylvania

	First Quantile (mean d_i=0.28, st. dev.=0.17)			Second Quantile (mean d_i=0.45, st. dev.=0.26)		
	At Poll	Absentee	Total	At Poll	Absentee	Total
RD Estimate	-7.536**	1.215*	-6.662**	-4.669*	0.814	-3.427
	(3.280)	(0.686)	(3.274)	(2.434)	(0.749)	(2.358)
N	578,698	578,698	578,698	578,651	578,651	578,651
Outome mean	51.72	1.40	53.13	54.21	1.90	56.11
Bandwidth	0.09	0.07	0.09	0.11	0.09	0.11
	Third Quantile (mean d_i=0.76, st. dev.=0.45)			Fourth Quantile (mean d_i=1.53, st. dev.=1.37)		
	At Poll	Absentee	Total	At Poll	Absentee	Total
RD Estimate	-1.458	0.226	-1.490	0.598	0.008	0.640
	(1.215)	(0.375)	(1.285)	(0.379)	(0.103)	(0.415)
N	578,698	578,698	578,698	579,393	579,393	579,393
Outome mean	59.02	2.41	61.43	59.87	2.69	62.56
Bandwidth	0.17	0.15	0.16	0.18	0.16	0.16

H_0 Quantile 1 = Quantile 2: At polls, $p-$value=0.48; Absentee, $p-$value=0.69; Total, $p-$value=0.42.
H_0 Quantile 1 = Quantile 3: At polls, $p-$value=0.08; Absentee, $p-$value=0.21; Total, $p-$value=0.14.
H_0 Quantile 1 = Quantile 4: At polls, $p-$value=0.01; Absentee, $p-$value=0.08; Total, $p-$value=0.03.

Note: This table reports local linear estimates of the RD treatment effect for quantiles of distance to polling place. Distance to polling place is the travel route distance, measured in miles. Each point estimate is for a sub-sample of observations based on the midpoint between the average of distances to polling place in the treatment and control precincts. 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.

Table E.2: Fuzzy RD Estimates: Nonlinear effects in Georgia

	First Quantile (mean d_i=0.72, st. dev.=0.41)			Second Quantile (mean d_i=1.20, st. dev.=0.60)		
	At Poll	Absentee	Total	At Poll	Absentee	Total
RD Estimate	-6.766*	-0.317	-6.294	-0.153	1.301	1.381
	(3.609)	(4.694)	(5.256)	(1.440)	(1.591)	(1.935)
N	306,677	306,677	306,677	306,590	306,590	306,590
Outome mean	27.10	25.01	52.12	26.77	27.38	54.15
Bandwidth	0.14	0.10	0.10	0.17	0.21	0.17
	Third Quantile (mean d_i=1.66, st. dev.=0.79)			Fourth Quantile (mean d_i=2.55, st. dev.=1.53)		
	At Poll	Absentee	Total	At Poll	Absentee	Total
RD Estimate	1.237	-3.482	-0.245	-1.641***	0.978	-0.765
	(1.305)	(2.182)	(2.260)	(0.586)	(0.854)	(0.928)
N	306,977	306,977	306,977	306,913	306,913	306,913
Outome mean	25.32	29.10	54.42	24.91	30.40	55.32
Bandwidth	0.19	0.12	0.16	0.20	0.14	0.15

H_0 Quantile 1 = Quantile 2: At polls, $p-$value=0.09; Absentee, $p-$value=0.74; Total, $p-$value=0.17.
H_0 Quantile 1 = Quantile 3: At polls, $p-$value=0.04; Absentee, $p-$value=0.54; Total, $p-$value=0.29.
H_0 Quantile 1 = Quantile 4: At polls, $p-$value=0.16; Absentee, $p-$value=0.79; Total, $p-$value=0.30.

Note: This table reports local linear estimates of the RD treatment effect for four quantiles of distance to polling place. Each point estimate is for a sub-sample of observations based on the average of the two distances to polling place at the RD point. 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.

Table E.3: Border FE Estimates: Nonlinear effects in Pennsylvania

	First Quantile (mean d_i=, st. dev.=)			Second Quantile (mean d_i=, st. dev.=)		
	At Poll	Absentee	Total	At Poll	Absentee	Total
Distance to Polling Place (mi)	-0.91	0.70**	-0.21	-1.45	0.41	-1.03
	(1.43)	(0.28)	(1.43)	(2.12)	(0.48)	(2.08)
N	629,577	629,577	629,577	399,295	399,295	399,295
Outome mean	50.83	1.31	52.14	50.71	1.50	52.21
Bandwidth						

H_0 Quantile 1 = Quantile 2: At polls, $p-$value=0.83; Absentee, $p-$value=0.60; Total, $p-$value=0.74.
H_0 Quantile 1 = Quantile 3: At polls, $p-$value=0.86; Absentee, $p-$value=0.07; Total, $p-$value=0.79.
H_0 Quantile 1 = Quantile 4: At polls, $p-$value=0.59; Absentee, $p-$value=0.04; Total, $p-$value=0.89.

Note: This table reports border fixed effects estimates for quartiles of distance to polling place. The border FE sample is divided into quartiles based on the average distance to the polling place within a border segment. Distance is measured as travel route distance in miles. 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. Each regression includes border segment fixed effects, individual-level covariates (age, sex, registered democrat, registered republican), and block-level covariates (percent White, non-Hispanic, percent Black, non-Hispanic, percent Hispanic, median household income, percent with no high school degree, and percent that walk to work, cars per household). Standard errors allow for clustering at the border segment level.

Table E.4: Border FE Estimates: Nonlinear effects in Georgia

	First Quantile (mean d_i=0.53, st. dev.=0.22)			Second Quantile (mean d_i=1.06, st. dev.=0.11)		
	At Poll	Absentee	Total	At Poll	Absentee	Total
Distance to Polling Place (mi)	-3.91**	6.11***	2.19	-3.03	2.89	-0.15
	(1.67)	(1.47)	(2.48)	(2.79)	(2.60)	(2.87)
N	97,233	97,233	97,233	58,837	58,837	58,837
Outome mean	26.55	23.81	50.36	24.59	26.03	50.61
Bandwidth						
	Third Quantile (mean d_i=1.50, st. dev.=0.15)			Fourth Quantile (mean d_i=3.40, st. dev.=1.69)		
	At Poll	Absentee	Total	At Poll	Absentee	Total
Distance to Polling Place (mi)	-2.53	0.77	-1.76	-1.05***	1.13***	0.07
	(1.93)	(1.90)	(2.24)	(0.16)	(0.17)	(0.16)
N	64,918	64,918	64,918	169,584	169,584	169,584
Outome mean	23.86	26.90	50.75	24.07	30.83	54.90
Bandwidth						

H_0 Quantile 1 = Quantile 2: At polls, $p-$value=0.79; Absentee, $p-$value=0.28; Total, $p-$value=0.54.
H_0 Quantile 1 = Quantile 3: At polls, $p-$value=0.59; Absentee, $p-$value=0.03; Total, $p-$value=0.24.
H_0 Quantile 1 = Quantile 4: At polls, $p-$value=0.09; Absentee, $p-$value=0.00; Total, $p-$value=0.39.

Note: This table reports border fixed effects estimates for quartiles of distance to polling place. The border FE sample is divided into quartiles based on the average distance to the polling place within a border segment. Distance is measured as travel route distance in miles. 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. Each regression includes border segment fixed effects, individual-level covariates (age, sex, registered democrat, registered republican), and block-level covariates (percent White, non-Hispanic, percent Black, non-Hispanic, percent Hispanic, median household income, percent with no high school degree, and percent that walk to work, cars per household). Standard errors allow for clustering at the border segment level.

F Block-level Border Fixed Effects Regressions

In this section we estimate the effect of distance to polling place on voter registration rates and turnout rates in each state, using block-level analysis. In each census block we observe the total voting age population (VAP), which is a proxy for the population eligible to vote. Percent registered is the number of voters we observe in the voter registration file, divided by the VAP of the block. We measure at-poll, absentee, and total turnout as the number of votes cast divided by the VAP. Because VAP comes from 2010 Census data and voter registration files are from 2018, the turnout rates and registration rates are imperfectly measured. The measures are especially noisy for small blocks. Among blocks with fewer than 50 people, the average number of registered voters is 137% and 114% of VAP in Pennsylvania and Georgia, respectively. The percent registered voters in blocks with population above 50 is 84% and 96% on average, which is much more reasonable. To reduce noise in our outcome measures, we only include blocks with population of 50 or higher.

Table F.1: Block-level Border Fixed Effects Regressions

	A. Pennsylvania			
	Percent Registered	At Poll	Absentee	Total
Distance to Polling Place (mi)	-0.42	-0.74	0.06	-0.68
	(0.77)	(0.47)	(0.05)	(0.48)
N	29,014	29,014	29,014	29,014
Outcome mean	85.42	46.47	1.56	48.04
R^2	0.47	0.61	0.49	0.62
	B. Georgia			
	Percent Registered	At Poll	Absentee	Total
Distance to Polling Place (mi)	-0.41	-1.55***	1.57***	0.03
	(0.64)	(0.21)	(0.29)	(0.39)
N	10,074	10,074	10,074	10,074
Outcome mean	99.66	26.12	29.79	55.91
R^2	0.41	0.56	0.57	0.54

Note: A unit of observation is a Census Block. Distance to polling place measured as travel-route distance in miles. At-polls is the number of votes cast in a polling place per registered voters in the block, times 100. Likewise for turnout by absentee ballots and in total. The coefficient for distance to polling place thus represents the change in percentage points of likelihood of voting per mile All regressions include border segment fixed effects and the following controls: population, voting age population, percent registered Democrat, percent registered Republican, percent age 30-49, percent age 50-64, percent age 65 and up, percent female, percent Black, percent Hispanic, median household income, percent without a high school diploma, percent that walk to work, and cars per household. Standard errors clustered at the border level are reported in parentheses.

G Border FE Estimates for a Sample Matched to Cantoni (2020)

In this section we use the replication data from Cantoni (2020)[1] to reconcile the differences in estimates in his study versus in our Table F.1. We construct a sample of blocks in Georgia and Pennsylvania that are more similar to the urban blocks in the sample of Cantoni (2020), which includes the Boston Massachusetts (MA) area and Minneapolis, Minnesota (MN). Note that Cantoni does not provide individual-level estimates, but reports block-level and parcel-level estimates. A parcel is a unit of land within a city, smaller than a Census block, but there are no state-wide parcel data for Pennsylvania and Georgia. Thus, we compare block-level estimates only. As we do not have the geospatial data used to construct the Cantoni Replication data, we cannot use our RD estimation framework there.

To find blocks that are comparable to Boston and Minneapolis, we pool block-level data from PA, GA, MN, and MA, and estimate a propensity score for the likelihood of being in the MA or MN sample. We use a logit specification and the covariates used by Cantoni (population, income, race, car ownership, and education). Compared to the statewide PA and GA samples, the census blocks in these areas are higher in population, income, and education. Then, we construct a matched sample by selecting the blocks with the highest propensity score. To improve precision, we select border segments at a time. That is, we compute the average propensity score of a border segment and include all observations that belong to border segments with the highest average propensity score. In one matched sample, we choose a sample size roughly equivalent to that of Cantoni (2020). In a second matched sample, we choose a sample size four times as large as that of Cantoni (2020), again to improve precision. For comparability, we use turnout rates in general elections of midterm years for the outcome variables (2018 in Georgia and Pennsylvania, 2014 in Massachusetts and Minneapolis).

Table G.1 reports the border fixed effects estimates for full state samples (columns 1 and 2), for the matched samples (columns 3 and 4) and for the urban areas in Massachusetts and Minnesota (column 5). If we use only the top 2% of blocks by propensity score, the estimated coefficient for distance to polling place is -8.68 (SE=5.44), roughly 8 times larger than the point estimate for all of Georgia, and in contrast to the null estimate for all of Pennsylvania. The point estimate in the matched sample is, however, comparable to the point estimate

[1] Cantoni, Enrico. 2020. "Replication package for: A Precinct Too Far: Turnout and Voting Costs." American Economic Association [publisher]. Accessed at `https://www.aeaweb.org/journals/dataset?id=10.1257/app.20180306` on 2023-10-23.

from the sample of urban blocks in Massachusetts and Minnesota ($\beta = -5.44$, $SE = 2.48$, as in Table 4, Panel C of Cantoni (2020)).

Table G.1: Border FE Estimates for Sub-samples that are observationally similar to sample from urban areas in Cantoni (2020)

	State Samples		Matched Sample (PA and GA)		Cantoni (2020)
	(1) PA	(2) GA	(3) Top 10%	(4) Top 2%	(5) MA and MN
Distance to polling place (mi)	0.767	-1.149	1.941	-4.789	-5.435**
	(1.351)	(1.024)	(2.127)	(5.915)	(2.484)
N	73883	24706	8454	1694	1694
y variable mean	56.887	69.010	77.054	82.875	38.229
R^2	0.298	0.421	0.402	0.316	0.595

Note: The Urban Areas sample is provided by Cantoni (2020) and include data from the Boston, Massachusetts area (MA) and Minneapolis, Minnesota (MN). The Matched Samples include observations near border segments that have the highest average propensity score (top 10% of observations and top 2% of observations) for the likelihood of being in the MA and MN samples. The dependent variable is turnout in the 2018 midterm election for Georgia (GA) and Pennsylvania (PA), and the dependent variable is turnout in the 2014 midterm election for MA and MN. All regressions include border fixed effects and the following additional controls: population, median household income, percent non-white, percent with no car, and percent with no high school diploma. Standard errors clustered at the border level are reported in parentheses.

H RD Estimates for Three Large Cities

Table H.1: RD Estimates for Rural and Urban Areas

	A. Philadelphia, Pennsylvania			
	First-stage	Second-stage		
	Distance (mi)	At Poll	Absentee	Total
RD Estimate	0.40***	-5.05***	0.59	-4.52**
	(0.05)	(1.91)	(0.37)	(1.89)
N	380,650	380,650	380,650	380,650
Effective N, Left	105,655	105,101	82,814	105,655
Effective N, Right	135,394	134,691	105,670	135,394
Bandwidth	0.09	0.09	0.07	0.09
Outome mean	0.42	52.98	1.08	54.06

	A. Pittsburgh, Pennsylvania			
	First-stage	Second-stage		
	Distance (mi)	At Poll	Absentee	Total
RD Estimate	0.59***	-2.11	0.35	-2.02
	(0.08)	(2.03)	(0.56)	(1.74)
N	352,448	352,448	352,448	352,448
Effective N, Left	84,416	63,921	44,669	84,416
Effective N, Right	80,379	61,711	43,699	80,379
Bandwidth	0.14	0.11	0.09	0.14
Outome mean	0.80	54.22	2.47	56.69

	C. Atlanta, Georgia			
	First-stage	Second-stage		
	Distance (mi)	At Poll	Absentee	Total
RD Estimate	1.39***	-0.76	-0.72	-0.76
	(0.14)	(1.12)	(1.68)	(1.47)
N	565,669	565,669	565,669	565,669
Effective N, Left	29,830	39,350	22,009	29,830
Effective N, Right	40,587	57,290	31,066	40,587
Bandwidth	0.13	0.15	0.11	0.13
Outome mean	1.37	27.47	28.98	56.45

Note: This table reports RD estimates for the three largest cities in the sample. The Philadelphia sample includes urban blocks in Philadelphia County, the Pittsburgh sample includes urban blocks in Allegheny county, and the Atlanta sample includes urban blocks in Fulton and DeKalb counties. Bandwidths are MSE-optimal. Standard errors allow for clustering at the RD point level.

I Difference in Differences Estimates

In this section we exploit individual-level variation in distance to polling place across time. We create a panel of voters using voting registration and voter history files from the 2016 and 2018 elections in Pennsylvania. Previous instances of voter registration files (from 2016) were unavailable for Georgia. We estimate two specifications. The first specification estimates the effect of a change in distance to the polling place using within-voter variation:

$$vote_{it} = \beta distance_{it} + \delta_i + \gamma_{ct} + \epsilon_{it}, \tag{I.1}$$

where δ_i are individual fixed effects and γ_{ct} are county-year fixed effects. In a second specification, we disentangle the effects of a change in voter residence versus a change in polling place location:

$$\begin{aligned} vote_{it} =& \beta PL\ moved_{it} + \zeta PL\ moved_{it} \times distance_{it} \\ &+ \mu Voter\ Moved_{it} + \eta Voter\ moved_{it} \times distance_{it} \\ &+ \iota PL\ moved_{it} \times Voter\ moved_{it} + \psi distance_{it} \times PL\ moved_{it} \times Voter\ moved_{it} \\ &+ \delta_i + \gamma_{ct} + \epsilon_{it}, \end{aligned} \tag{I.2}$$

where $PL\ moved_{it}$ is an indicator that equals 1 if voter i in election t is assigned to a polling location different from the one assigned in election $t-1$ and $Voter\ moved_{it}$ is an indicator variable that takes value 1 if voter i in election t has a different home address than during election $t-1$. Note that there can only be a change in distance if either the voter or the polling place moves, so we do not identify a coefficient for $distance_{it}$ alone. This specification allows us to identify the effect of distance to polling place separately from the effect of a change in polling place.

In Table I.1 we report estimates for Equation I.1. We estimate precise null effects of a change in distance to polling place on the likelihood of voting in total, at polls, and by absentee ballot. These results are consistent with Clinton et al. (2020) and Yoder (2018). When we separately consider voters who move versus polling places that move, we find a small negative effect of distance to polling place for those who experience a change in polling place but remain in their location and no statistically significant effect for those who moved (Table I.2). The point estimates indicate that if a voter moves, holding distance to polling place constant, then they are more likely to vote, whereas a voter whose polling location place moved, holding distance constant is less likely to vote. The effect of changes in the polling place on turnout is the focus of Clinton et al. (2020) and Yoder (2018). While the size of our difference-in-differences estimates is smaller than those reported in these papers, they are similar in order of magnitude.

Table I.1: The effect of distance to polling place on turnout: Difference in Differences Estimation in Pennsylvania

	General Election		
	At Poll	Absentee	Total
Distance (miles)	0.0554	0.0086	0.0640
	(0.0863)	(0.0165)	(0.0830)
N	14504036	14504036	14504036
y variable mean	63.92	2.66	66.58
R^2	0.755	0.657	0.757

Note: Distance to polling place measured in miles. The dependent variables are indicators for whether or not a registered voter has voted at the polling place, through absentee ballot, or through either voting method. All regressions include Individual Fixed Effects and County by Year FE. Standard errors clustered at the precinct level are reported in parentheses.

Table I.2: Difference in Differences Estimates from Pennsylvania:The effect of polling place changes, voter location changes, and distance to polling place on turnout:

	General Election		
	At Poll	Absentee	Total
PL Moved	0.3856*	-0.0744*	0.3112
	(0.2027)	(0.0409)	(0.1957)
Voter Moved	-2.9912**	0.6797***	-2.3115*
	(1.2586)	(0.1243)	(1.3149)
PL Moved × Dist. (mi)	-0.1283	-0.0068	-0.1351
	(0.0892)	(0.0192)	(0.0855)
Voter Moved × Dist. (mi)	-0.2751	-0.0029	-0.2780
	(0.7409)	(0.1740)	(0.7536)
Voter Moved × PL Moved	-0.2019	-0.0792	-0.2811
	(1.2975)	(0.1371)	(1.3520)
Voter Moved × PL Moved	-0.2272	0.0163	-0.2109
× Dist. (mi)	(0.7186)	(0.1798)	(0.7311)
N	14504036	14504036	14504036
y variable mean	63.92	2.66	66.58
R^2	0.755	0.657	0.757

Note: All regressions include individual voter fixed effects and county-year fixed effects. The dependent variables are indicators for whether or not a registered voter has voted at the polling place, through absentee ballot, or through either voting method. For readability, we multiply the dependent variables by 100 so that the coefficients can be interpreted as percentage point changes in the likelihood of voting. Standard errors clustered at the precinct level are reported in parentheses.

J Heterogeneous Effects

Table J.1: Heterogeneous Effects: Sex

A. Pennsylvania

	Female			Male		
	At Poll	Absentee	Total	At Poll	Absentee	Total
RD Estimate	-2.016*	0.233	-1.865*	-3.686**	0.555*	-3.145*
	(1.120)	(0.302)	(1.118)	(1.795)	(0.332)	(1.799)
N	862,191	862,191	862,191	1,018,933	1,018,933	1,018,933
Outome mean	58.81	2.18	60.99	51.27	2.01	53.28
Bandwidth	0.17	0.16	0.17	0.09	0.10	0.09

H_0 Female=Male: At polls, $p-$value=0.43; Absentee, $p-$value=0.47; Total, $p-$value=0.55.

B. Georgia

	Female			Male		
	At Poll	Absentee	Total	At Poll	Absentee	Total
RD Estimate	-0.543	1.228*	-0.219	-2.052**	0.956	-1.886
	(0.958)	(0.727)	(0.942)	(0.938)	(0.841)	(1.395)
N	663,210	663,210	663,210	561,522	561,522	561,522
Outome mean	26.92	29.98	56.90	24.98	25.63	50.61
Bandwidth	0.15	0.23	0.17	0.16	0.20	0.14

H_0 Female=Male: At polls, $p-$value=0.26; Absentee, $p-$value=0.81; Total, $p-$value=0.32.

Note: This table reports local linear estimates of the RD treatment effect for sub-samples of voters. Outcomes are residualized, after removing RD point fixed effects. We test if coefficients are equal for pairwise comparisons of sub-samples and report p-values above for each outcome variable. Bandwidths are MSE-optimal. Standard errors allow for clustering at the RD point level.

Table J.2: Heterogeneous Effects: Age

A. Pennsylvania

	Age 18-29			Age 30-49		
	At Poll	Absentee	Total	At Poll	Absentee	Total
RD Estimate	-0.543	0.111	-0.462	-0.873	0.396	-0.399
	(5.036)	(0.423)	(5.026)	(1.495)	(0.316)	(1.507)
N	353,948	353,948	353,948	771,522	771,522	771,522
Outome mean	33.93	2.62	36.55	49.71	0.98	50.69
Bandwidth	0.12	0.11	0.12	0.13	0.10	0.14
	Age 50-64			Age 65 and up		
	At Poll	Absentee	Total	At Poll	Absentee	Total
RD Estimate	-0.442	0.309	-0.008	-2.061	1.057	-0.779
	(1.969)	(0.327)	(1.978)	(2.309)	(0.866)	(2.138)
N	609,235	609,235	609,235	578,407	578,407	578,407
Outome mean	65.95	1.61	67.56	68.39	3.80	72.19
Bandwidth	0.08	0.14	0.08	0.11	0.09	0.11

H_0 Age 65 and up=Age 18-29: At polls, $p-$value=0.78; Absentee, $p-$value=0.33; Total, $p-$value=0.95.
H_0 Age 65 and up=Age 30-49: At polls, $p-$value=0.67; Absentee, $p-$value=0.47; Total, $p-$value=0.88.
H_0 Age 65 and up=Age 50-64: At polls, $p-$value=0.59; Absentee, $p-$value=0.42; Total, $p-$value=0.79.

B. Georgia

	Age 18-29			Age 30-49		
	At Poll	Absentee	Total	At Poll	Absentee	Total
RD Estimate	0.774	-1.729	-1.325	-0.416	0.127	-0.361
	(1.183)	(1.272)	(1.799)	(0.986)	(1.063)	(1.338)
N	285,003	285,003	285,003	446,284	446,284	446,284
Outome mean	20.77	15.17	35.94	28.73	22.56	51.28
Bandwidth	0.19	0.16	0.16	0.17	0.16	0.16
	Age 50-64			Age 65 and up		
	At Poll	Absentee	Total	At Poll	Absentee	Total
RD Estimate	-1.732	1.508	-1.019	-4.619***	6.588***	2.696*
	(1.194)	(1.194)	(1.758)	(1.748)	(1.487)	(1.521)
N	289,890	289,890	289,890	204,059	204,059	204,059
Outome mean	28.90	36.12	65.02	23.60	46.39	69.99
Bandwidth	0.16	0.20	0.15	0.20	0.25	0.16

H_0 Age 65 and up=Age 18-29: At polls, $p-$value=0.01; Absentee, $p-$value<0.01; Total, $p-$value=0.09.
H_0 Age 65 and up=Age 30-49: At polls, $p-$value=0.04; Absentee, $p-$value<0.01; Total, $p-$value=0.13.
H_0 Age 65 and up=Age 50-64: At polls, $p-$value=0.17; Absentee, $p-$value=0.01; Total, $p-$value=0.11.

Note: This table reports local linear estimates of the RD treatment effect for sub-samples of voters. Outcomes are residualized, after removing RD point fixed effects. We test if coefficients are equal for pairwise comparisons of sub-samples and report p-values above for each outcome variable. We only report p-values for comparisons to the sub-sample of voters aged 65 and up. Coefficients for the other age-groups are not statistically significantly different from each other (p-values are greater than 0.05 for all other pairwise comparisons). Bandwidths are MSE-optimal. Standard errors allow for clustering at the RD point level.

Table J.3: Heterogeneous Effects: Party

A. Pennsylvania

	Democrats			Republicans			Independent		
	At Poll	Absentee	Total	At Poll	Absentee	Total	At Poll	Absentee	Total
RD Estimate	-3.097***	0.553*	-2.555**	-1.442	-0.316	-1.659	-1.798	0.446	-1.303
	(1.131)	(0.283)	(1.141)	(1.875)	(0.585)	(1.654)	(2.332)	(0.547)	(2.400)
N	1,272,655	1,272,655	1,272,655	717,840	717,840	717,840	325,058	325,058	325,058
Outome mean	57.52	2.16	59.68	61.54	2.36	63.91	39.29	1.31	40.60
Bandwidth	0.12	0.12	0.12	0.11	0.10	0.14	0.12	0.09	0.12

H_0 Democrat=Republican: At polls, $p-$value=0.44; Absentee, $p-$value=0.18; Total, $p-$value=0.65.
H_0 Democrat=Independent: At polls, $p-$value=0.62; Absentee, $p-$value=0.86; Total, $p-$value=0.64.
H_0 Independent=Republican: At polls, $p-$value=0.91; Absentee, $p-$value=0.34; Total, $p-$value=0.90.

B. Georgia

	Democrats			Republicans			Independent		
	At Poll	Absentee	Total	At Poll	Absentee	Total	At Poll	Absentee	Total
RD Estimate	-1.915	0.573	0.444	-8.753*	9.281**	0.519	-0.522	0.693	0.027
	(2.722)	(3.275)	(1.175)	(4.489)	(3.893)	(1.339)	(0.598)	(0.519)	(0.738)
N	126,968	126,968	126,968	72,832	72,832	72,832	1,027,357	1,027,357	1,027,357
Outome mean	31.89	64.94	96.83	40.98	55.88	96.87	24.24	21.43	45.67
Bandwidth	0.16	0.12	0.14	0.13	0.15	0.19	0.21	0.28	0.21

H_0 Democrat=Republican: At polls, $p-$value=0.19; Absentee, $p-$value=0.09; Total, $p-$value=0.97.
H_0 Democrat=Independent: At polls, $p-$value=0.62; Absentee, $p-$value=0.97; Total, $p-$value=0.76.
H_0 Independent=Republican: At polls, $p-$value=0.07; Absentee, $p-$value=0.03; Total, $p-$value=0.75.

Note: This table reports local linear estimates of the RD treatment effect for sub-samples of voters. Outcomes are residualized, after removing RD point fixed effects. We test if coefficients are equal for pairwise comparisons of sub-samples and report p-values above for each outcome variable. Bandwidths are MSE-optimal. Standard errors allow for clustering at the RD point level.

Table J.4: Heterogeneous Effects: Race and Ethnicity

A. Georgia

	White, non-Hispanic			Black, non-Hispanic			Hispanic		
	At Poll	Absentee	Total	At Poll	Absentee	Total	At Poll	Absentee	Total
RD Estimate	-1.464	0.450	-1.702**	-1.343	2.080**	0.205	1.912	-4.659*	-1.507
	(1.348)	(1.235)	(0.850)	(0.939)	(1.008)	(1.471)	(2.154)	(2.407)	(3.358)
N	537,091	537,091	537,091	460,417	460,417	460,417	42,157	42,157	42,157
Outome mean	30.18	30.66	60.83	23.22	28.80	52.02	26.37	17.73	44.10
Bandwidth	0.14	0.16	0.26	0.21	0.23	0.18	0.21	0.17	0.16

H_0 White=Black: At polls, $p-$value=0.94; Absentee, $p-$value=0.31; Total, $p-$value=0.26.
H_0 White=Hispanic: At polls, $p-$value=0.18; Absentee, $p-$value=0.06; Total, $p-$value=0.96.
H_0 Black=Hispanic: At polls, $p-$value=0.17; Absentee, $p-$value=0.01; Total, $p-$value=0.65.

Note: This table reports local linear estimates of the RD treatment effect for sub-samples of voters. Outcomes are residualized, after removing RD point fixed effects. We test if coefficients are equal for pairwise comparisons of sub-samples and report p-values above for each outcome variable. Bandwidths are MSE-optimal. Standard errors allow for clustering at the RD point level.

Table J.5: Heterogeneous Effects: Median Household Income

A. Pennsylvania

	Below Median HH Income			Above Median HH Income		
	At Poll	Absentee	Total	At Poll	Absentee	Total
RD Estimate	-4.19***	0.18	-4.01***	-1.31	0.30	-0.98
	(1.45)	(0.25)	(1.47)	(1.99)	(0.61)	(1.96)
N	1,157,660	1,157,660	1,157,660	1,157,813	1,157,813	1,157,813
Outome mean	49.31	1.42	50.73	63.11	2.78	65.89
Bandwidth	0.10	0.11	0.10	0.06	0.06	0.06

H_0 Below = Above : At polls, $p-$value=0.24; Absentee, $p-$value=0.85; Total, $p-$value=0.22.

B. Georgia

	Below Median HH Income			Above Median HH Income		
	At Poll	Absentee	Total	At Poll	Absentee	Total
RD Estimate	-2.07*	-0.92	-3.63**	-0.49	2.48**	1.27
	(1.16)	(1.08)	(1.63)	(1.06)	(0.97)	(1.35)
N	612,633	612,633	612,633	614,524	614,524	614,524
Outome mean	23.30	24.12	47.42	28.74	31.82	60.56
Bandwidth	0.14	0.17	0.12	0.16	0.23	0.15

H_0 Below = Above : At polls, $p-$value=0.31; Absentee, $p-$value=0.02; Total, $p-$value=0.02.

Note: This table reports local linear estimates of the RD treatment effect for sub-samples of voters. Outcomes are residualized, after removing RD point fixed effects. We test if coefficients are equal for pairwise comparisons of sub-samples and report p-values above for each outcome variable. Bandwidths are MSE-optimal. Standard errors allow for clustering at the RD point level.

Table J.6: Heterogeneous Effects: Percent with No High School Diploma

A. Pennsylvania

	Below Median % with no HS Diploma			Above Median % with no HS Diploma		
	At Poll	Absentee	Total	At Poll	Absentee	Total
RD Estimate	-1.28	0.97**	-0.15	-3.07**	-0.04	-3.24**
	(1.51)	(0.41)	(1.48)	(1.55)	(0.28)	(1.54)
N	1,157,745	1,157,745	1,157,745	1,157,808	1,157,808	1,157,808
Outome mean	61.53	2.79	64.32	50.88	1.42	52.30
Bandwidth	0.12	0.14	0.13	0.09	0.10	0.09

H_0 Below = Above : At polls, $p-$value=0.41; Absentee, $p-$value=0.04; Total, $p-$value=0.15.

B. Georgia

	Below Median % with no HS Diploma			Above Median % with no HS Diploma		
	At Poll	Absentee	Total	At Poll	Absentee	Total
RD Estimate	-0.05	-0.67	-0.36	-2.04**	1.17	-1.10
	(1.19)	(1.51)	(1.41)	(0.84)	(0.99)	(1.17)
N	612,677	612,677	612,677	614,480	614,480	614,480
Outome mean	28.30	30.56	58.85	23.76	25.40	49.16
Bandwidth	0.14	0.13	0.15	0.22	0.20	0.20

H_0 Below = Above : At polls, $p-$value=0.17; Absentee, $p-$value=0.31; Total, $p-$value=0.69.

Note: This table reports local linear estimates of the RD treatment effect for sub-samples of voters. Outcomes are residualized, after removing RD point fixed effects. We test if coefficients are equal for pairwise comparisons of sub-samples and report p-values above for each outcome variable. Bandwidths are MSE-optimal. Standard errors allow for clustering at the RD point level.

Table J.7: Heterogeneous Effects: Percent that Walk to Work

A. Pennsylvania

	Below Median Commute by Walking '			Above Median Commute by Walking		
	At Poll	Absentee	Total	At Poll	Absentee	Total
RD Estimate	0.71	0.92*	1.64	-4.55***	0.04	-4.50***
	(2.11)	(0.47)	(2.04)	(1.52)	(0.33)	(1.54)
N	1,157,773	1,157,773	1,157,773	1,157,780	1,157,780	1,157,780
Outome mean	60.28	2.33	62.61	52.14	1.88	54.01
Bandwidth	0.06	0.09	0.06	0.10	0.09	0.10

H_0 Above = Below : At polls, $p-$value=0.04; Absentee, $p-$value=0.13; Total, $p-$value=0.02.

B. Georgia

	Below Median Commute by Walking '			Above Median Commute by Walking		
	At Poll	Absentee	Total	At Poll	Absentee	Total
RD Estimate	-1.299	0.903	-1.275	-1.119	1.373	-0.323
	(1.022)	(0.934)	(1.320)	(0.994)	(0.935)	(1.297)
N	613,378	613,378	613,378	613,779	613,779	613,779
Outome mean	26.42	29.09	55.51	25.63	26.86	52.49
Bandwidth	0.16	0.22	0.17	0.17	0.23	0.17

H_0 Above = Below : At polls, $p-$value=0.90; Absentee, $p-$value=0.72; Total, $p-$value=0.61.

Note: This table reports local linear estimates of the RD treatment effect for sub-samples of voters. Outcomes are residualized, after removing RD point fixed effects. We test if coefficients are equal for pairwise comparisons of sub-samples and report p-values above for each outcome variable. Bandwidths are MSE-optimal. Standard errors allow for clustering at the RD point level.

K Turnout-Maximizing Polling Places

In this section, we solve an optimal polling place location problem. We assume that the planner's objective is to maximize aggregate turnout. Of course, in practice, officials who are responsible for selecting polling place locations might have other objectives such as representativeness of the electorate that might also factor into the planning problem. However, for the purposes of this exercise, we think that the straightforward objective of turnout-maximization provides a useful benchmark.

A planner chooses where to locate a single polling place within a precinct in order to maximize aggregate turnout. We assume that the planner knows how voters are distributed across the precinct. Each voter decides whether to vote or abstain from voting. Importantly, the location of the polling place affects voting decisions only through the cost of traveling to the polling place to vote in-person.

We model a precinct, A, as a compact two-dimensional space, $A \subset R^2$. The planner chooses coordinates for the polling place location, $(x^p, y^p) \subset A$. There are N eligible voters distributed across precinct A. A voter i is located at $(x_i, y_i) \subset A$. Let $p(x, y)$ be the mass of voters at any point $(x, y) \subset A$.

Voters decide whether to abstain ($v_i = 0$) or vote ($v_i = 1$). To keep the model tractable, we abstract away from the distinction between voting by mail or by absentee ballot. Since the planner's objective is to maximize turnout, the method of voting is a second-order concern. We also assume that voting is a function of Euclidean distance to the polling place, rather than travel route or travel time, since it is difficult to simulate counterfactual travel routes. We assume that voter i's utility can be written as follows:

$$
\begin{aligned}
u_i(v_i = 0) &= 0 + \epsilon_{0i} \\
u_i(v_i = 1) &= a_i + c(d_i) + \epsilon_{1i}
\end{aligned}
$$

where $c(.)$ is the net benefit of voting as a function of voter i's Euclidean distance to the polling place d_i, a_i denotes voter-specific net benefit of voting unrelated to distance to the polling place, and ϵ_{0i} and ϵ_{1i} are independently and identically distributed extreme value shocks. These assumptions on the functional forms of the utility and error terms translate to convenient logit choice probability functions:

$$Pr_i(v_i = 0) = \frac{1}{1 + e^{a_i + c(d_i)}} \tag{K.1}$$

$$Pr_i(v_i = 1) = \frac{e^{a_i + c(d_i)}}{1 + e^{a_i + c(d_i)}} \tag{K.2}$$

The planner's optimization problem is to pick a set of geographical coordinates for the polling place location (x^P, y^P) that solve the following:

$$\max_{\{x^p, y^p\} \subset A} \sum_{i=1}^{N} (1 - Pr_i(v_i = 0))$$

The maximand represents the aggregate precinct-level voter turnout, which is the sum of individual probabilities of voting for all N individuals in the precinct.

K.1 Computational Procedure

Below, we outline the steps of the computational procedure used to estimate the parameters of the model introduced in the previous section along with the optimal polling place locations using these parameters as inputs.

We use the border sample from Pennsylvania to estimate the parameters in equations 5 and 6 using a fixed-effects logit specification. We impose additional structure on a_i and $c(d_i)$ in order to empirically estimate $Pr_i(v_i = 1)$. Specifically, we assume that the cost function $c(d_i)$ is cubic and a_i is a linear function of all other observables that we control for in the main specifications:

$$a_i = \alpha X_i$$

$$c(d_i) = \beta_1 d_i + \beta_2 d_i^2 + \beta_3 d_i^3$$

where X_i is a vector of controls for party affiliation, age group, block-level population and block-level voting age-population, and d_i is the distance to the polling place for individual i. We report the estimates of the fixed-effects logit models in Table K.1. We use the estimated coefficients to simulate turnout for counterfactual polling place locations for the full sample.

In order to better match the predicted turnout with the level of turnout in the sample, we calibrate a precinct fixed effect, using the average turnout rate of the precinct to compute an expected log odds ratio. Note that the conditional logit estimation treats the fixed effects as nuisance parameters, and we also use only the border sample for estimation. Hence, we need to calibrate a fixed effect for all observations if we want predicted turnout to match observed turnout under the existing locations of polling places. We calibrate the fixed effect ι_p for precinct p as follows:

$$\iota_p = ln\left(\frac{\bar{t}_p}{1-\bar{t}_p}\right) - \frac{1}{N_p}\sum_i(\hat{\alpha}X_i + \hat{\beta}_1 d_i + \hat{\beta}_2 d_i^2 + \hat{\beta}_3 d_i^3), \tag{K.3}$$

where $\bar{t}_p$ is the average turnout rate in a precinct, N_p is he number of observations in a precinct, and $\hat{\alpha}_i$ and β_j are the estimated parameters. The mean predicted turnout rate for the full sample is 58.77% and the mean observed turnout rate is 59.03%.

In Table K.2, we report the average marginal effect of d_i on the probability of voting, by quartile of Euclidean distance to the polling place. The marginal effects follow a similar pattern as those implied by the nonlinear RD estimates in Table E.1 of the main text, though they are smaller in magnitude. We do not expect the magnitudes to be exactly equal, since here we use quartiles of Euclidean Distance and in Table E.1 we use quartiles of travel route distance. Further, Table E.1 uses only the RD sample and Table K.2 uses the full sample. The average marginal effects by quartile are more similar to those estimated by the border FE approach in Table E.3. This is because we are similarly estimating effects using within-border variation in distance to polling place. To the extent that the border FE approach underestimates the effect of distance to polling place on turnout, our simulated gains to turnout from using optimal polling places may also be an underestimate.

Next, we use the estimates to solve for the optimal polling place locations. To find the turnout maximizing optimal polling locations we solve a constrained optimization problem for each existing precinct using the standard Nelder-Mead algorithm. The optimal polling location is constrained to fall within an approximately 7-mile box centered around the current polling place location. We use this constraint to rule out polling locations that are optimal due to the behavior of the cubic function $c(.)$ outside of the observed range of distance to polling place. We do not constrain the optimal polling place location to be within the bounds of the precinct boundaries. We prefer to ignore this constraint because if, for a fixed set of voters, the otpimal location falls outside of the existing precinct borders, then it is a sign that the precinct borders are not optimal. In the second counterfactual exercise, we use the same model of voting to find the optimal building location. We compute predicted aggregate turnout for each candidate building in a precinct and select the building with the highest turnout rate. Finally, in the third counterfactual exercise we select the pair of public buildings in each precinct to maximize the precinct's turnout, simulating the doubling of the number of polling places per precinct.

Table K.1: Logit Estimates: Likelihood of voting

	(1)
Distance to polling place	-0.110***
	(0.029)
Distance to polling place ^{2}	0.040***
	(0.014)
Distance to polling place ^{3}	-0.004**
	(0.002)
Democrat	0.716***
	(0.006)
Republican	0.581***
	(0.007)
Age 30-49	0.344***
	(0.008)
Age 50-64	0.975***
	(0.008)
Age 65 and up	1.155***
	(0.009)
Population	0.001***
	(0.000)
Voting Age Population	-0.002***
	(0.000)
Predicted turnout rate	0.55
N	1,722,734

Table K.2: Estimated marginal effects by quartile of distance to polling place

	Euclidean distance to polling place (mi)		Estimated Marginal Effect (p.p. per mi)	
	Mean	St. Dev.	Mean	St. Dev.
1st quartile	0.13	0.06	-2.1	0.33
2nd quartile	0.35	0.7	-1.7	0.27
3rd quartile	0.76	0.18	-1.2	0.25
4th quartile	2.5	7.2	-0.24	4.78