Online Appendix
Polling Place Location and the Costs of Voting

Gaurav Bagwe
Georgetown University

Juan Margitic
Georgetown University

Allison Stashko
University of Utah

July 5, 2022

Contents

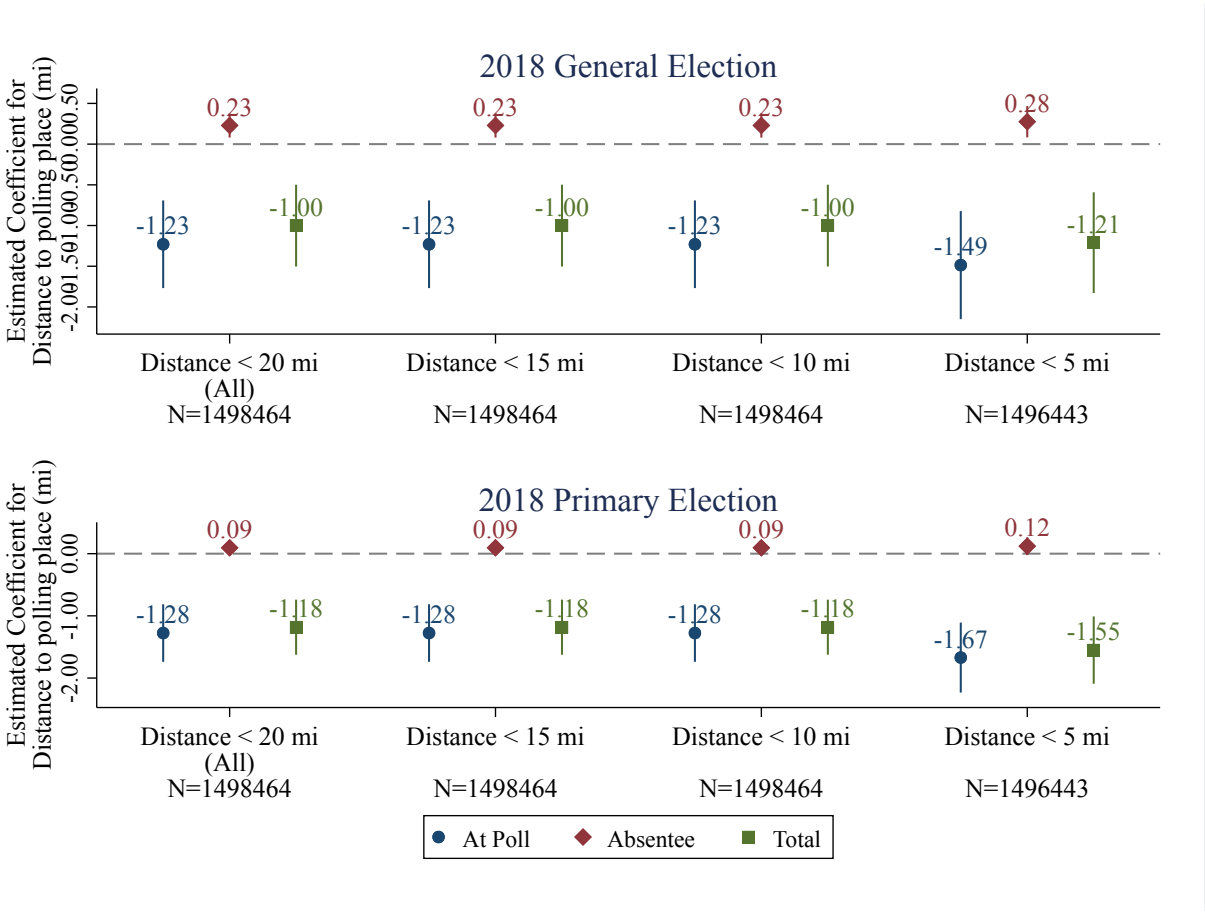
A Data Appendix

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

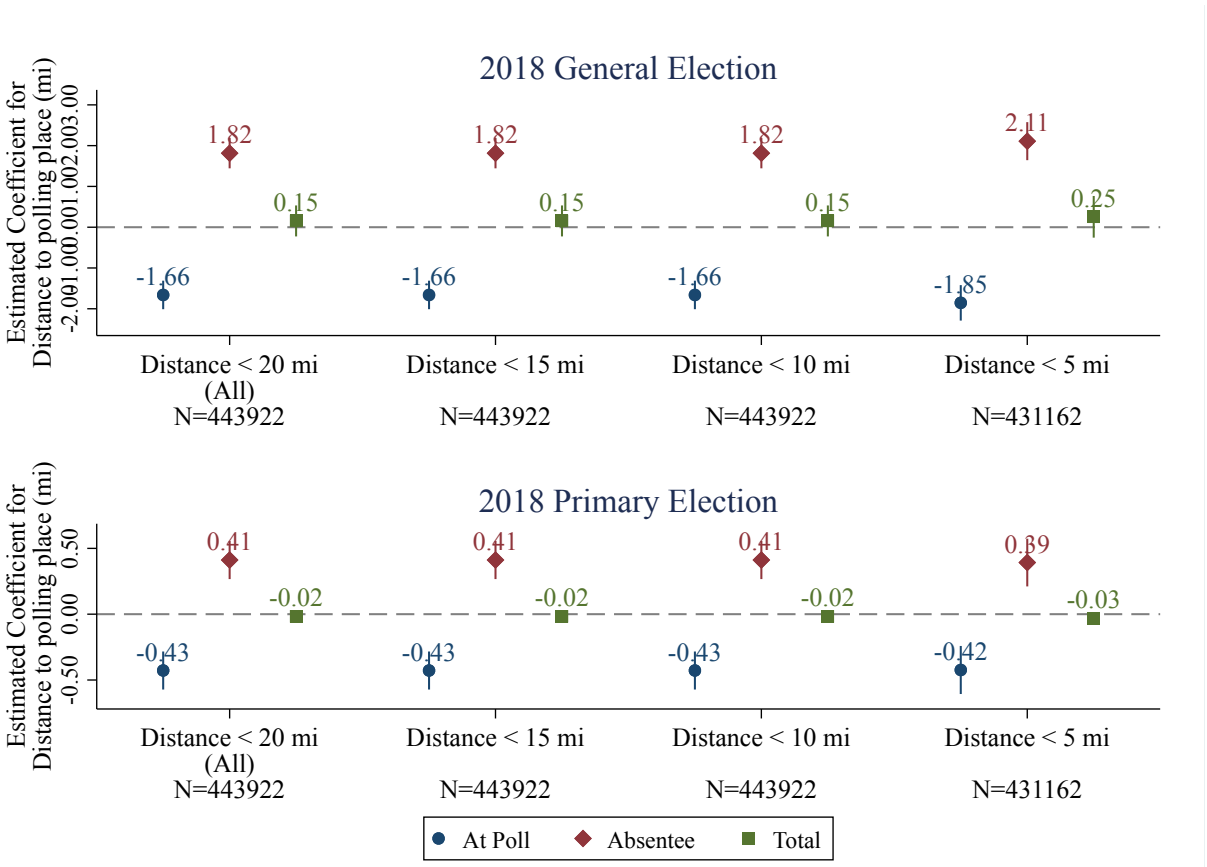
B Robustness Checks

Figure B.1: Individual-level Border Fixed Effects estimates: Vary maximum distance to polling place, Pennsylvania



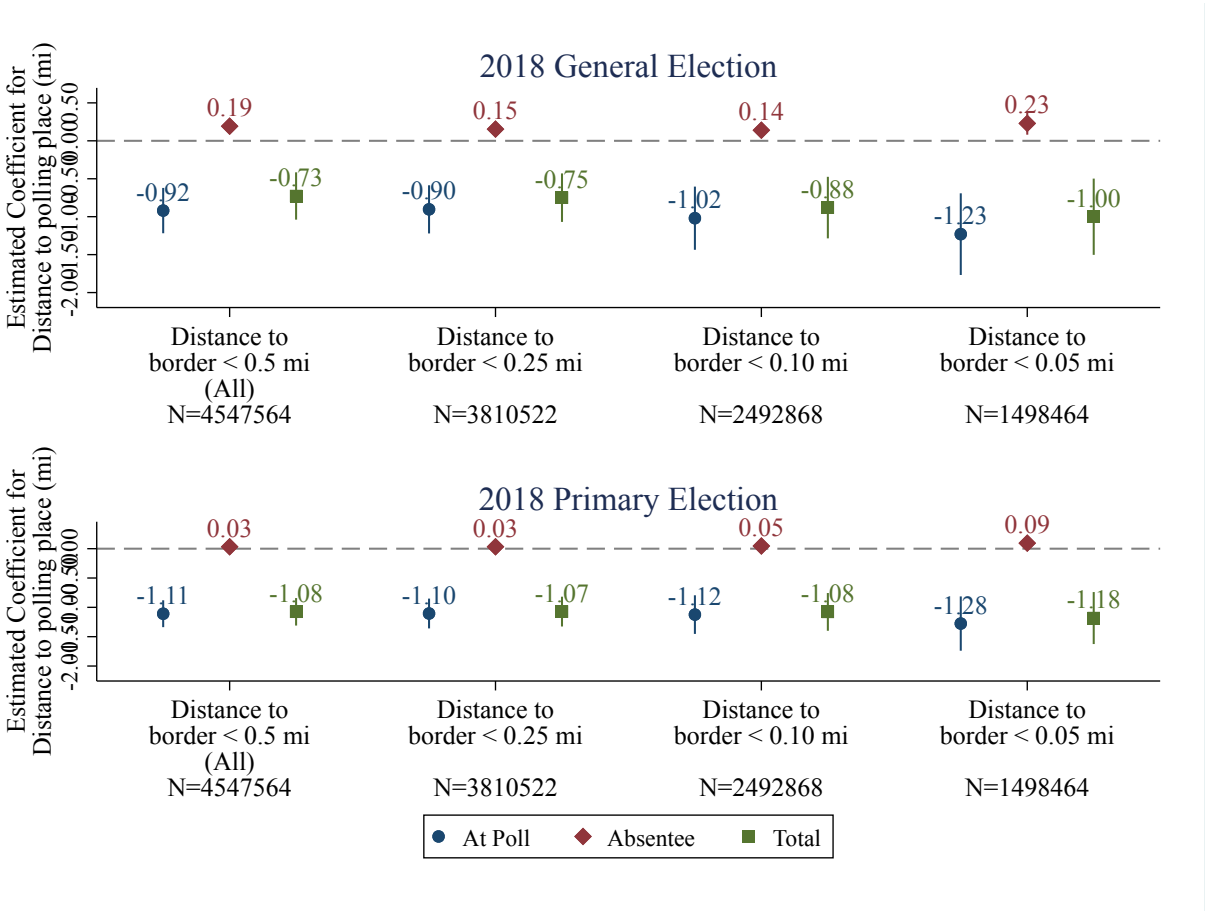
Note: The y-axis measures the coefficient on distance to polling place on the likelihood of voting at polls, by absentee ballot, or by either method. Outcome variables are re-scaled so that coefficients measure the percentage point change in the likelihood of voting. Each symbol represents a point estimate in a separate regression for the sub-sample indicated in the legend. Vertical lines indicate 95% confidence intervals. Standard errors allow for clustering at the border level. All regressions include border fixed effects, individual-level controls (registered Democrat, registered Republican, age 30-49, age 50-64, age 65 and up, female) and block-level controls (population, voting age population, percent Black, percent Hispanic, median household income, percent without a high school diploma, percent that walk to work, percent with commute time less than 5 minutes, and percent with commute greater than 60 minutes).

Figure B.2: Individual-level Border Fixed Effects estimates: Vary maximum distance to polling place, Georgia



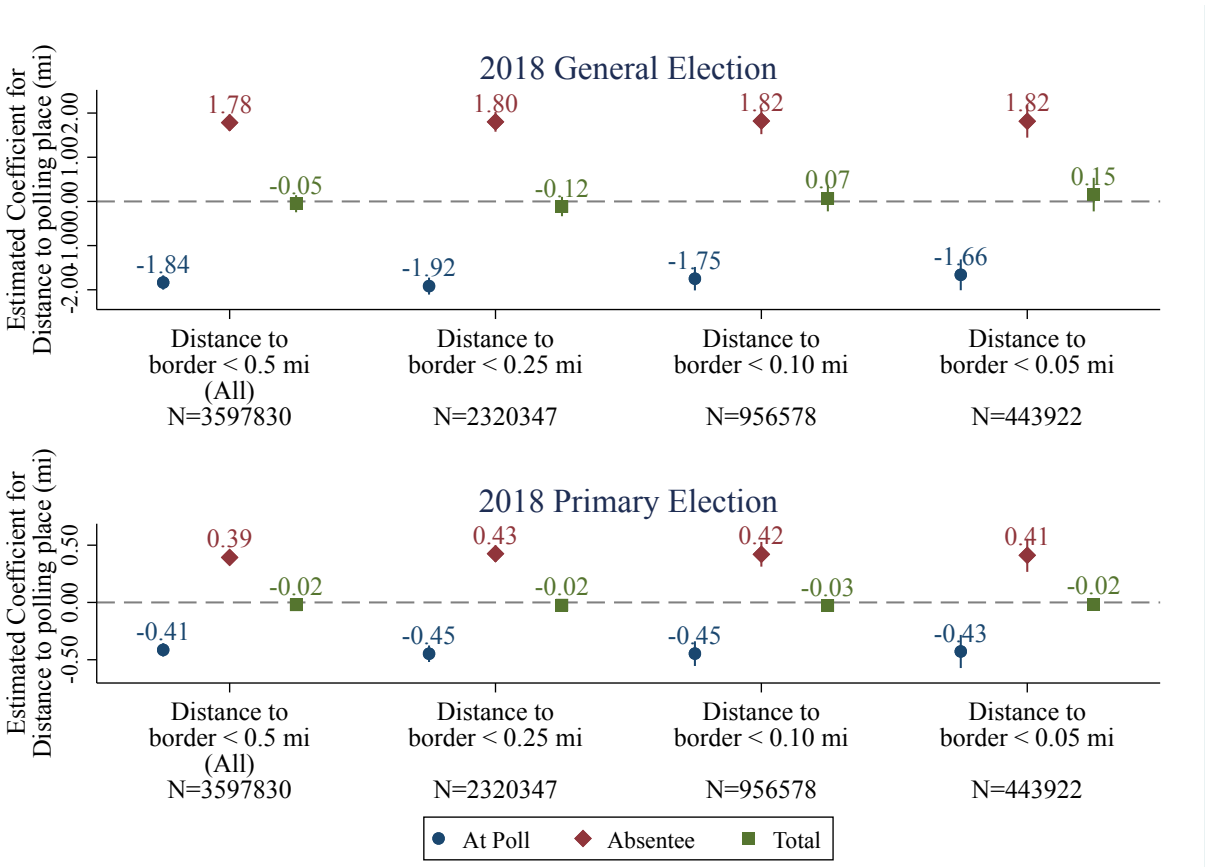
Note: The y-axis measures the coefficient on distance to polling place on the likelihood of voting at polls, by absentee ballot, or by either method. Outcome variables are re-scaled so that coefficients measure the percentage point change in the likelihood of voting. Each symbol represents a point estimate in a separate regression for the sub-sample indicated in the legend. Vertical lines indicate 95% confidence intervals. Standard errors allow for clustering at the border level. All regressions include border fixed effects, individual-level controls (registered Democrat, registered Republican, age 30-49, age 50-64, age 65 and up, female) and block-level controls (population, voting age population, percent Black, percent Hispanic, median household income, percent without a high school diploma, percent that walk to work, percent with commute time less than 5 minutes, and percent with commute greater than 60 minutes).

Figure B.3: Individual-level Border Fixed Effects estimates: Vary maximum distance to border segment, Pennsylvania



Note: The y-axis measures the coefficient on distance to polling place on the likelihood of voting at polls, by absentee ballot, or by either method. Outcome variables are re-scaled so that coefficients measure the percentage point change in the likelihood of voting. Each symbol represents a point estimate in a separate regression for the sub-sample indicated in the legend. Vertical lines indicate 95% confidence intervals. Standard errors allow for clustering at the border level. All regressions include border fixed effects, individual-level controls (registered Democrat, registered Republican, age 30-49, age 50-64, age 65 and up, female) and block-level controls (population, voting age population, percent Black, percent Hispanic, median household income, percent without a high school diploma, percent that walk to work, percent with commute time less than 5 minutes, and percent with commute greater than 60 minutes).

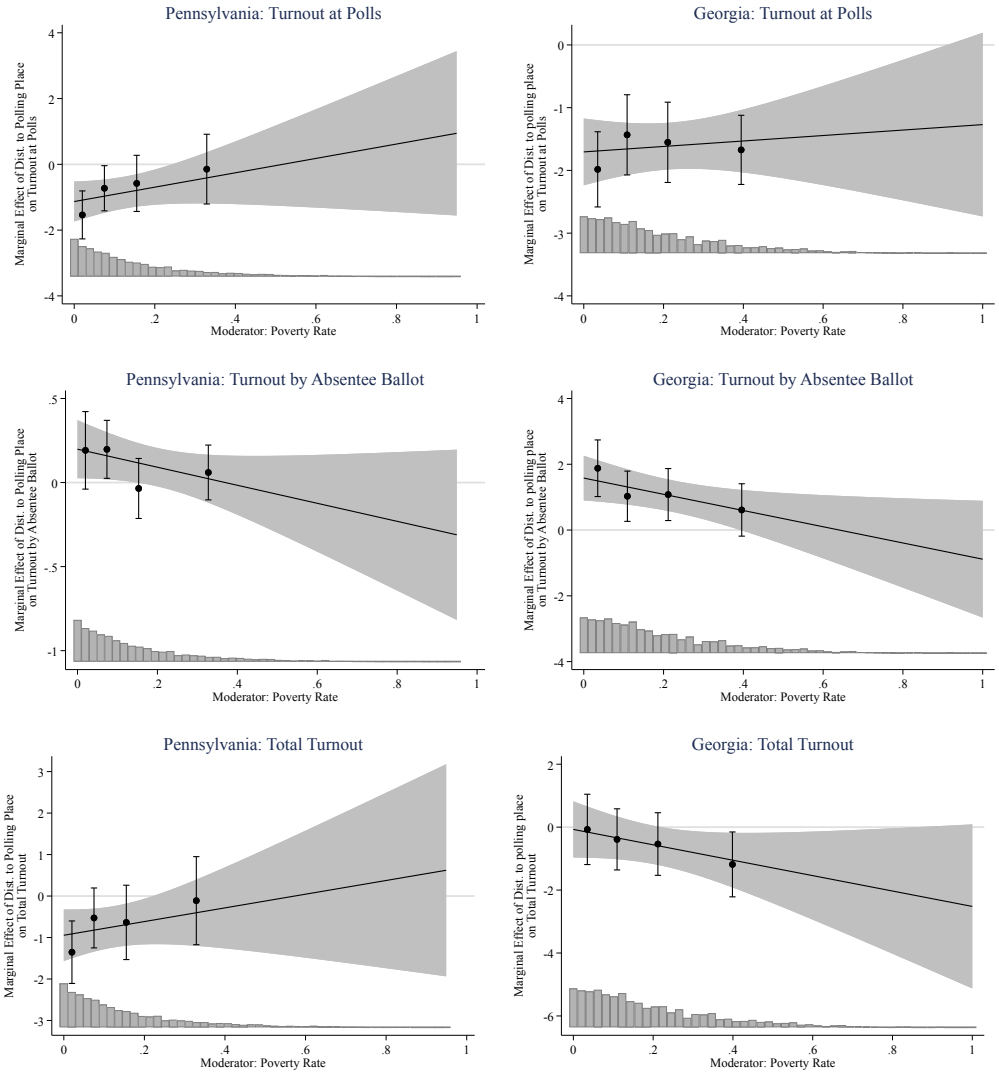
Figure B.4: Individual-level Border Fixed Effects estimates: Vary maximum distance to border segment, Georgia



Note: The y-axis measures the coefficient on distance to polling place on the likelihood of voting at polls, by absentee ballot, or by either method. Outcome variables are re-scaled so that coefficients measure the percentage point change in the likelihood of voting. Each symbol represents a point estimate in a separate regression for the sub-sample indicated in the legend. Vertical lines indicate 95% confidence intervals. Standard errors allow for clustering at the border level. All regressions include border fixed effects, individual-level controls (registered Democrat, registered Republican, age 30-49, age 50-64, age 65 and up, female) and block-level controls (population, voting age population, percent Black, percent Hispanic, median household income, percent without a high school diploma, percent that walk to work, percent with commute time less than 5 minutes, and percent with commute greater than 60 minutes).

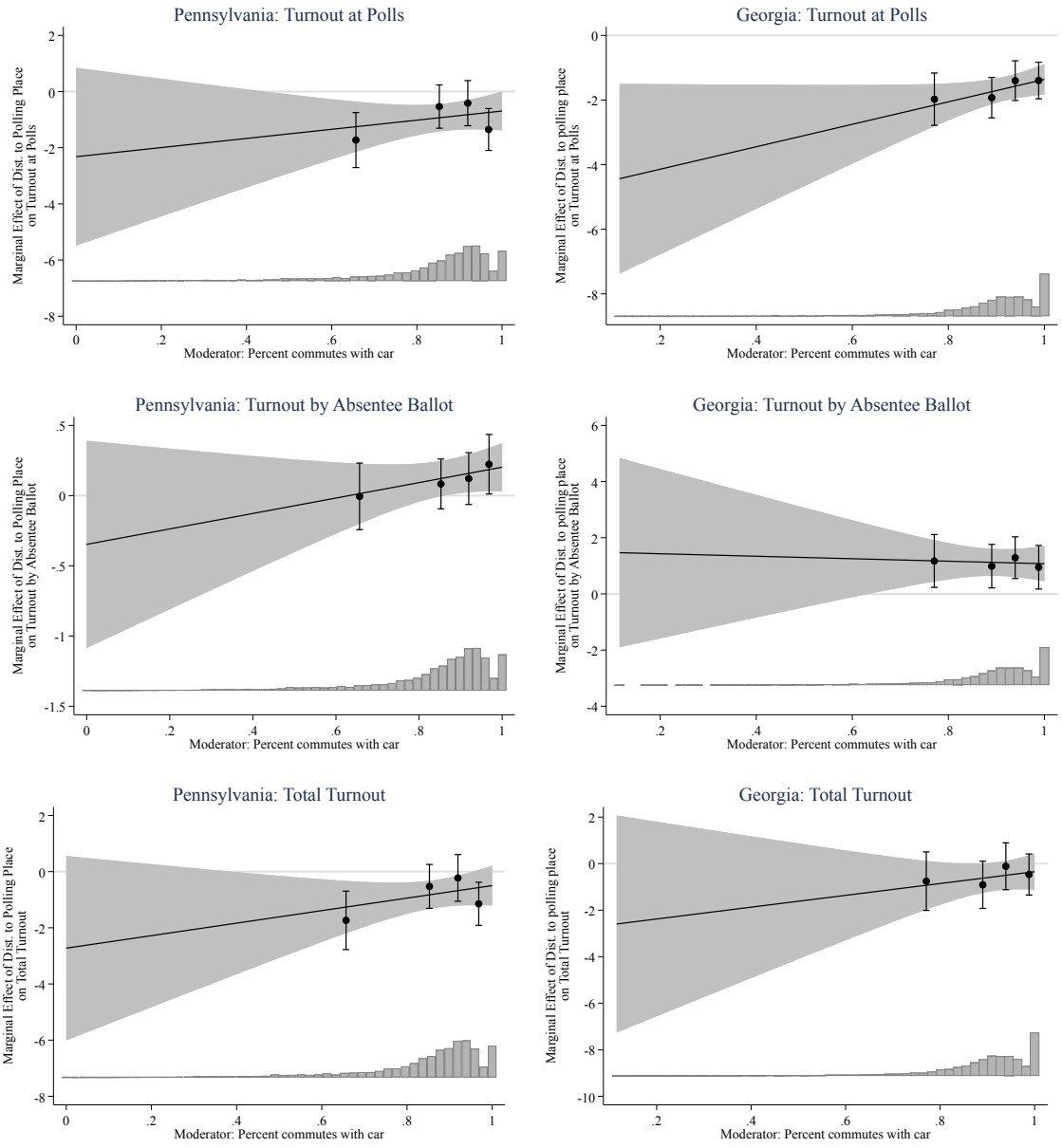
C Heterogeneous Effects

Figure C.1: Heterogeneous Effects – Poverty Rate



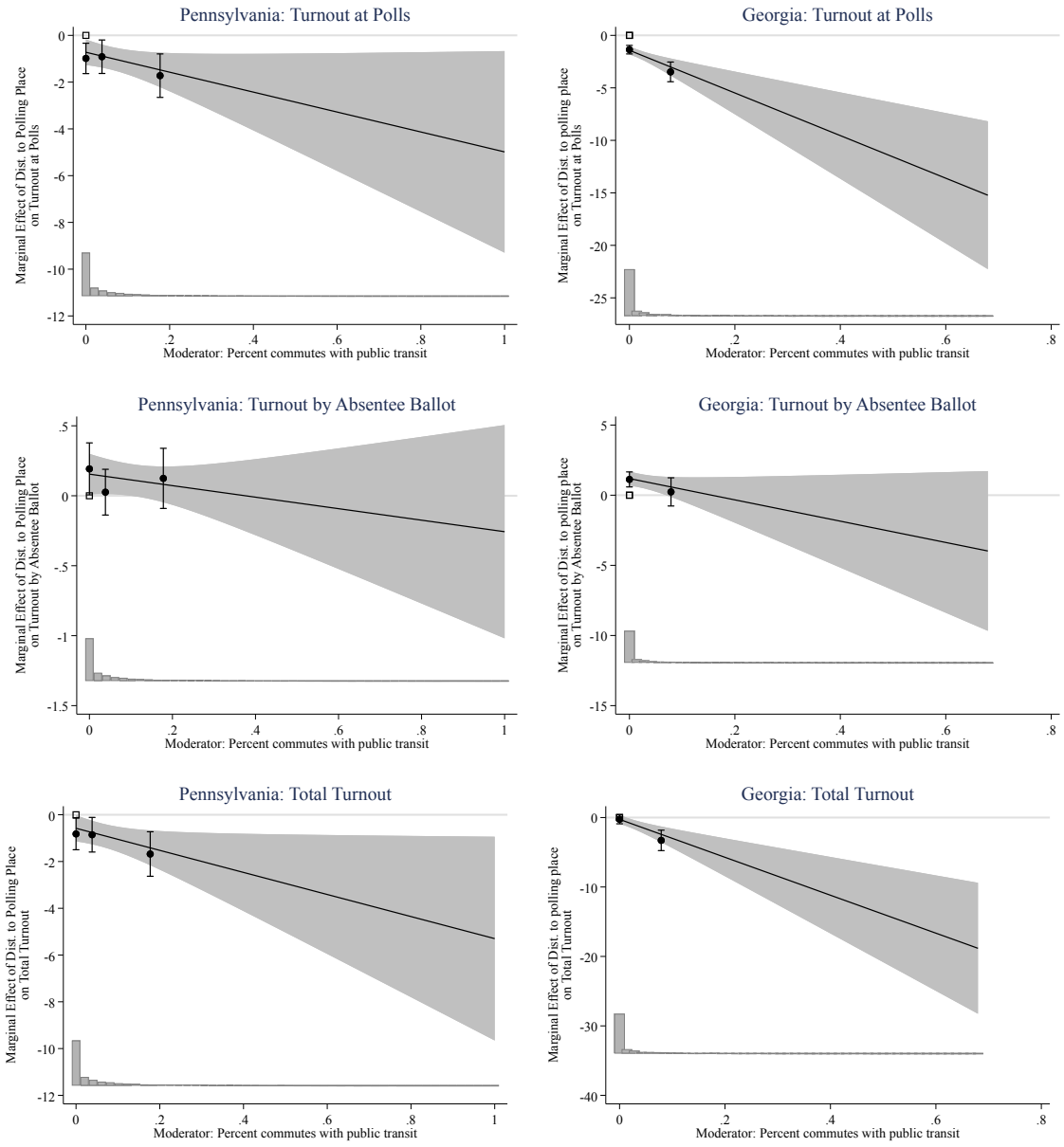
Note: This figure shows the marginal effect of distance to polling place on turnout at polls, turnout by absentee ballot, and total turnout at different levels of the moderating variable – poverty rate. The solid line plots the estimated marginal effect from the linear interaction model. The gray area shows the 95% confidence interval. The black circles and vertical lines represent the binning estimates and 95% confidence intervals. A histogram of the moderating variable is shown along the x-axis in gray. All regressions include border fixed effects and the following controls: percent registered Democrat, percent registered Republican, percent age 30-49, percent age 50-64, percent age 65 and up, percent female, population, voting age population, percent Black, percent Hispanic, median household income, percent without a high school diploma, percent that walk to work, percent with commute time less than 5 minutes, and percent with commute greater than 60 minutes.

Figure C.2: Heterogeneous Effects – Transportation by Car



Note: This figure shows the marginal effect of distance to polling place on turnout at polls, turnout by absentee ballot, and total turnout at different levels of the moderating variable – percent of workers who travel to work with a car. The solid line plots the estimated marginal effect from the linear interaction model. The gray area shows the 95% confidence interval. The black circles and vertical lines represent the binning estimates and 95% confidence intervals. A histogram of the moderating variable is shown along the x-axis in gray. All regressions include border fixed effects and the following controls: percent registered Democrat, percent registered Republican, percent age 30-49, percent age 50-64, percent age 65 and up, percent female, population, voting age population, percent Black, percent Hispanic, median household income, percent without a high school diploma, percent that walk to work, percent with commute time less than 5 minutes, and percent with commute greater than 60 minutes.

Figure C.3: Heterogeneous Effects – Transportation by Public Transit



Note: This figure shows the marginal effect of distance to polling place on turnout at polls, turnout by absentee ballot, and total turnout at different levels of the moderating variable – percent of workers who travel to work with public transportation. The solid line plots the estimated marginal effect from the linear interaction model. The gray area shows the 95% confidence interval. The black circles and vertical lines represent the binning estimates and 95% confidence intervals. A histogram of the moderating variable is shown along the x-axis in gray. All regressions include border fixed effects and the following controls: percent registered Democrat, percent registered Republican, percent age 30-49, percent age 50-64, percent age 65 and up, percent female, population, voting age population, percent Black, percent Hispanic, median household income, percent without a high school diploma, percent that walk to work, percent with commute time less than 5 minutes, and percent with commute greater than 60 minutes.

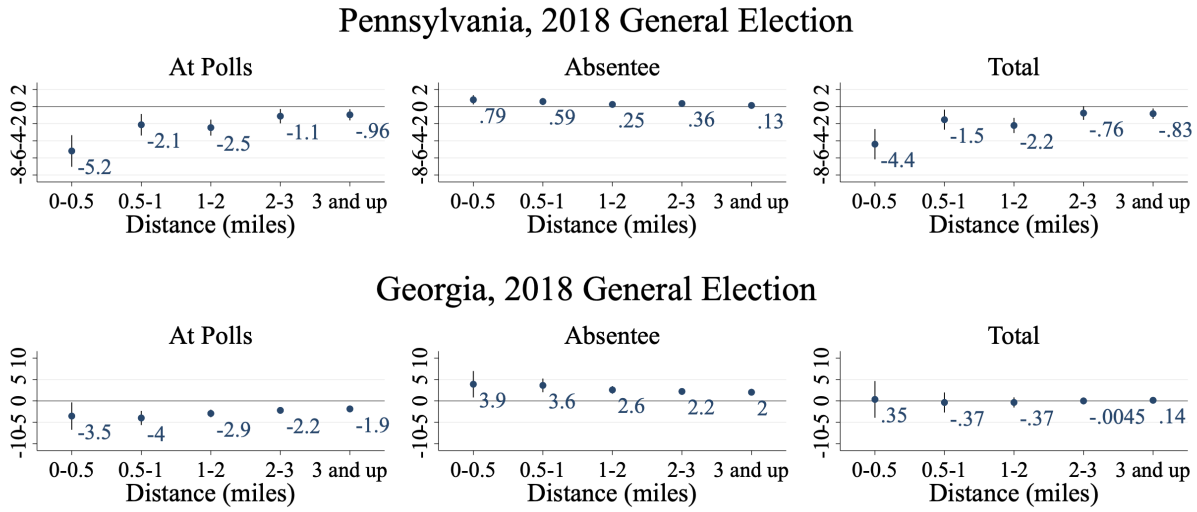
D Nonlinear Effects

We estimate the following specification to isolate the effect of distance to polling place across different intervals:

$$vote_i = \delta_{s(i)} + \gamma\ D_i + \beta\ D_i \cdot dist_i + \rho \mathcal{P}_i + \iota\ \mathcal{X}_b(i) + \epsilon_i \tag{D.1}$$

The variable D_i refers to a vector of indicator variables that take value 1 when the average distance to polling place in block i is within a particular range of miles and 0 otherwise. The distance ranges are: [0.0,0.5), [0.5,1), [1,2), [2,3), and [3,10]. We report in Figure D.1 the vector of coefficients, β for general elections. These coefficients can be interpreted as the effect of distance to polling place on the likelihood of voting within each distance category.

Figure D.1: The effect of distance to polling place on likelihood of voting in General elections: Nonlinear effects



Note: Distance to polling place is 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 border fixed effects and additional individual-level and block-level covariates: registered Democrat indicator, registered Republican indicator, population, voting age population, percent Black, percent Hispanic, median household income, percent without a high school diploma, percent that walk to work, and indicators for whether travel time to work is less than 5 minutes or greater than 60 minutes. Standard errors clustered at the border level are reported in parentheses.

E Optimal Polling Places: Logit Estimates

Table E.1: Logit Estimates: Likelihood of voting

	(1) Urban	(2) Rural
Distance (miles)	-0.033***	-0.067***
	(0.012)	(0.020)
Democrat	0.718***	0.609***
	(0.006)	(0.049)
Republican	0.550***	0.928***
	(0.008)	(0.046)
Age 30-49	0.356***	0.859***
	(0.008)	(0.048)
Age 50-64	0.976***	1.573***
	(0.008)	(0.047)
Age 65 and up	1.057***	1.750***
	(0.009)	(0.053)
Population	0.001***	0.001
	(0.000)	(0.003)
Voting Age Population	-0.002***	-0.001
	(0.000)	(0.004)
N	1624198	27476

F Optimal Polling Places: Model Details

First order conditions for x^p and y^p are:

$$\sum_{i=1}^{N} -\frac{e^{a+f(d_i)} \cdot f'(d_i) + e^{b+c(d_i)} \cdot c'(d_i)}{(1+e^{a+f(d_i)}+e^{b+c(d_i)})^2} \left(\frac{x_i - x_p}{d_i}\right) = 0$$
$$\sum_{i=1}^{N} -\frac{e^{a+f(d_i)} \cdot f'(d_i) + e^{b+c(d_i)} \cdot c'(d_i)}{(1+e^{a+f(d_i)}+e^{b+c(d_i)})^2} \left(\frac{y_i - y_p}{d_i}\right) = 0$$

These first order conditions simplify to:

$$\sum_{i=1}^{N} p_0(d_i)[p_1(d_i)f'(d_i) + p_2(d_i)c'(d_i)] \left(\frac{x_i - x_p}{d_i}\right) = 0$$
$$\sum_{i=1}^{N} p_0(d_i)[p_1(d_i)f'(d_i) + p_2(d_i)c'(d_i)] \left(\frac{y_i - y_p}{d_i}\right) = 0$$

The second order sufficient condition for x^p is:

$$\sum_{i=1}^{N} \Big(p_0'(d_i)[p_1(d_i)f'(d_i) + p_2(d_i)c'(d_i)]$$
$$+p_0(d_i)[p_1'(d_i)f'(d_i) + p_1(d_i)f''(d_i) + p_2'(d_i)c'(d_i) + p_2(d_i)c''(d_i)]\Big)\Big(\frac{x - x_p}{d_i}\Big)^2$$
$$-p_0(d_i)[p_1(d_i)f'(d_i) + p_2(d_i)c'(d_i)]\Big(\frac{(x-x_p)^2 - d_i^2}{d_i^3}\Big) < 0$$

where,

$$p_0'(d_i) = p_0(d_i)[p_1(d_i)f'(d_i) + p_2(d_i)c'(d_i)] \left(\frac{x_i - x_p}{d_i}\right) \tag{F.1}$$