Online Appendix

Chapter 5

Data

We obtained data for the 2008–2012 elections via voter file snapshots from the vendor Catalist, LLC, and for the 2014–2022 elections from L2 Political. Voter file snapshots offer an unparalleled look into the racial turnout gap. It is important to note, however, that these snapshots change every day as voters register or are removed from the rolls. As such, estimates can be slightly different if scholars are working with different snapshots. We follow the advice of Kim and Fraga (2022) and here include the dates of the snapshots used in this chapter. Unfortunately, Catalist does not report the dates of the snapshots on which their files are based (though they are dated shortly after the election), so we can provide the dates only from the L2-based snapshots.

Table A.5.1: Snapshot Dates

State	2014	2016	2018	2020	2022
AK	2015-03-13	2017-01-27	2019-02-11	2021-02-03	2023-02-18
AL	2015-04-10	2017-03-07	2019-01-27	2021-02-04	2023-01-20
AR	2015-03-24	2017-03-29	2018-09-21	2021-01-19	2023-04-20
AZ	2015-04-22	2017-04-12	2018-09-07	2021-04-27	2023-03-21
CA	2015-05-21	2017-03-25	2019-01-31	2021-02-19	2022-12-19
CO	2015-05-05	2017-02-08	2019-08-31	2020-12-23	2022-12-19

Table A.5.1: Snapshot Dates *(continued)*

State	2014	2016	2018	2020	2022
CT	2015-03-25	2017-01-20	2019-06-03	2021-03-30	2023-04-01
DC	2015-03-07	2017-02-15	2019-01-17	2021-01-30	2023-03-11
DE	2015-02-23	2017-01-17	2019-04-02	2021-03-24	2023-04-01
FL	2015-01-28	2017-01-27	2019-02-08	2021-02-04	2023-02-11
GA	2015-05-16	2017-01-27	2018-12-22	2021-02-04	2022-12-23
HI	2015-03-05	2017-03-22	2019-04-05	2021-04-01	2023-03-21
IA	2015-03-25	2017-01-31	2019-03-06	2021-03-04	2023-01-29
ID	2015-02-23	2017-03-20	2019-03-04	2021-03-16	2023-04-20
IL	2015-03-02	2017-03-18	2019-02-21	2021-03-05	2023-03-21
IN	2015-05-06	2017-04-07	2019-02-13	2021-01-15	2023-04-01
KS	2015-02-26	2017-02-16	2019-01-31	2021-03-16	2023-04-01
KY	2015-03-05	2017-03-03	2018-09-29	2021-05-11	2023-04-01
LA	2015-02-23	2017-02-14	2019-01-15	2021-01-22	2023-01-20
MA	2015-04-02	2017-04-11	2019-02-14	2021-01-19	2023-04-26
MD	2015-02-25	2017-01-20	2018-12-14	2021-02-15	2023-03-21
ME	2015-04-29	2017-04-07	2018-09-26	2021-05-28	2023-04-01
MI	2015-02-28	2017-02-21	2019-03-22	2021-01-30	2023-02-25
MN	2015-03-03	2017-03-10	2019-04-02	2021-02-14	2023-04-01
MO	2015-03-02	2017-02-08	2019-06-03	2021-02-11	2023-02-25
MS	2015-03-17	2017-03-07	2019-03-11	2021-03-23	2023-04-20
MT	2015-03-27	2017-01-25	2019-02-07	2020-12-14	2023-02-25
NC	2015-07-29	2017-01-12	2019-02-01	2021-01-28	2023-02-18
ND	2015-04-15	2017-02-09	2019-03-22	2021-03-18	2023-03-15
NE	2015-03-25	2017-01-13	2019-01-10	2021-01-20	2023-01-16
NH	2015-03-20	2018-08-15	2019-04-10	2021-03-25	2023-05-08
NJ	2015-02-25	2017-03-31	2019-04-03	2021-03-11	2023-02-04
NM	2015-03-19	2017-02-08	2019-02-22	2021-02-25	2023-04-01
NV	2015-01-30	2017-01-13	2019-01-23	2020-12-17	2023-02-04
NY	2015-03-25	2017-03-14	2019-02-27	2021-03-15	2023-03-01
OH	2015-01-08	2017-01-09	2019-01-22	2021-01-07	2023-01-16

Table A.5.1: Snapshot Dates *(continued)*

State	2014	2016	2018	2020	2022
OK	2015-03-26	2017-01-12	2019-03-01	2021-02-08	2023-02-25
OR	2015-04-16	2017-01-13	2019-02-24	2021-02-05	2023-03-11
PA	2015-05-01	2017-02-14	2019-09-23	2021-02-17	2023-02-04
RI	2015-03-06	2017-01-18	2019-03-15	2021-02-10	2023-02-25
SC	2015-04-09	2017-02-24	2019-03-12	2021-04-16	2023-03-11
SD	2015-03-13	2017-02-20	2019-01-14	2021-01-22	2023-02-25
TN	2015-02-23	2017-02-17	2019-01-30	2021-03-29	2023-02-04
TX	2015-04-15	2017-03-12	2019-02-24	2021-03-25	2023-03-15
UT	2015-03-06	2017-01-25	2019-03-07	2021-03-26	2023-02-18
VA	2015-04-18	2017-03-29	2019-03-12	2021-02-18	2023-04-01
VT	2015-03-20	2017-02-14	2019-03-08	2021-03-04	2023-02-25
WA	2015-05-05	2017-05-24	2019-01-08	2020-12-09	2023-01-20
WI	2015-03-03	2017-03-30	2019-02-01	2021-02-24	2023-02-18
WV	2015-03-16	2017-04-03	2019-03-22	2021-03-11	2023-04-20
WY	2015-03-30	2017-02-02	2019-04-02	2021-01-13	2023-04-20

Comparability of Catalist and L2 Data

Given that the switch from Catalist to L2 corresponds with the timing of the *Shelby County* decision, it is important to ensure that our datasets are comparable. The data from the two vendors could differ in two key ways: they could include a different set of voters (as each vendor might "prune" the registration lists differently), or, conditional on registration, the predicted race of voters could differ if the vendors geocode voters to different block groups. Here, we show that these factors are unlikely to be driving our results.

First, we assess the total number of ballots estimated by summing the individuals marked as participating in the election. These are unlikely to match official ballot counts exactly, as some voters may be removed between election day and the date on which the voter file snapshot is collected (see Kim and Fraga, 2022). Table A.5.2 shows that these

errors are unlikely to impact our analyses. We divide the number of estimated ballots from the voter file in each state-year by official ballot counts for that state-year; specifically, the number of ballots cast for the highest office, the best measure available in all states in all years.[1] This means that voter file estimates can exceed official counts due to undervotes at the top of the ballot. Our dependent variable here is the absolute value of the difference in ballot counts between the two sources divided by the official ballot counts.

Table A.5.2 shows that for both Catalist and L2, the average absolute difference between voter file and official counts is less than 5 percent (see the *Intercept*). For neither vendor were these errors different in formerly covered states (*(Formerly) Covered State*) which, in the case of the Catalist were of course still covered. While the number of estimated ballots cast according to the L2 data are closer to the "true" number of ballots cast, the L2 numbers are not closer in formerly covered than in formerly uncovered states (*(Formerly) Covered State × L2 Data*; by way of reminder, the fixed effects regression absorbs the other variables with which they are colinear).

Similarly, we find no evidence in our racial estimates relying on data from different vendors. We do not use the proprietary racial estimates provided by Catalist and L2, relying instead on self-reported race where available and the BISG approach detailed in the body of the manuscript. Nevertheless, it remains possible that differences in the vendors' geocoding result in placing voters into different block groups and, thus, impact estimated race.

To test the comparability of the data, we identified about 94,000,000 individuals who were registered following both the 2012 (according to the Catalist data) and 2014 (according to L2) elections and did not move. We conducted the match using the voter's unique state identification number (not the distinct vendor's identification number), voters' house numbers, and voters' ZIP codes. This is not possible for all states: for 9 states, Catalist did not provide state ID numbers that match ID numbers in the L2 files. Nevertheless, the extremely large pool of voters for whom the comparison is valid indicates that there is no

[1] McDonald, Michael P. 2008–2022. "Voting Statistics." United States Elections Project. Accessed April 11, 2024.

Table A.5.2: Ballot Count Errors

	Catalist	L2	All Years	Differential Change
(Formerly) Covered State	-0.005	0.002		
	(0.008)	(0.005)		
L2 Data			-0.013**	
			(0.004)	
Covered State × L2 Data				0.008
				(0.009)
Intercept	0.048***	0.032***	0.046***	
	(0.005)	(0.003)	(0.004)	
State Fixed Effects				✓
Year Fixed Effects				✓
Num.Obs.	153	255	408	408
R2	0.003	0.002	0.034	0.404
R2 Adj.	-0.004	-0.002	0.032	0.305

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Standard errors clustered by state.

reason to think that the vendors' geocoding processes are driving our results.

Given that voters' racial estimates are updated each year as the racial composition of the citizen voting-age population in an assigned block group changes, we would expect a correlation coefficient between the two data sources that approaches (but is not exactly) 1. The correlation coefficient on the predicted probability of being white / nonwhite was 0.97.

Table A.5.3 formalizes this in of a two-period, two-way fixed effect model with year and voter fixed effects. The coefficient on *Formerly Covered State × L2 Data* indicates that registrants (models 1 and 2) and voters (models 3 and 4) in formerly covered states saw their likelihood of being either white or Black decrease slightly more in the formerly covered places than in the rest of the country. These marginal differences, however, are extremely small (the predicted probability of being white among those who voted both years, for instance, decreased by 0.07 percentage points more in these regions) and too small to explain the observed effects of *Shelby County* on the turnout gap.

We thus conclude that switching from Catalist to L2 in 2014 does not threaten our

Table A.5.3: Change in Racial Predictions

	Registrants		Voters	
	P(White)	P(Black)	P(White)	P(Black)
Formerly Covered State × L2 Data	-0.0008***	-0.0030***	-0.0007***	-0.0024***
	0.0000	0.0000	0.0000	0.0000
Voter Fixed Effects	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓
Num.Obs.	187220058	187220058	95748810	95748810
R2 Adj.	0.967	0.972	0.966	0.975

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Robust standard errors clustered by voter.

analyses. The different vendors do not report different ballot counts in total and, conditional on participation, our models do not return meaningfully different racial estimates for individual voters.

Estimating Voter Race

Since most states do not provide self-reported race or ethnicity in their voter files, we rely on a modified version of Bayesian Improved Surname Geocoding (BISG; see Imai and Khanna, 2016), which employs Census data on racial distributions among surnames and neighborhoods to predict voters' race.[2] BISG incorporates information about the racial distribution of each voter's surname *and* the racial distribution of their local geography (ie, block group). We make one important modification to Imai and Khanna's approach: While the canonical version of BISG uses the decennial *total* population for geographic demographics, we instead use the Citizen Voting Age Population (CVAP), using the 5-year estimates ending with each election year (see below for more information on this change).[3] This shift yields a more

[2]Following standard practice, we also estimate the race of voters who declined to denote their race in the self-identification states, or identify as "other."

[3]The ACS estimates did not begin until the 2005–2009 period; we use these as the BISG input for the 2008 election. We use the 2017–2021 numbers for the 2022 election, as they were not available yet at the time of writing. The same caveat applies for the use of CVAP as a denominator for turnout, discussed below, as well as other ACS-based demographic controls.

accurate estimation of the voting population in counties where self-identified race can be used as a comparison than either total population or total adult population. As we show below, a more recent methodological proposal (Bayesian Instrumental Regression for Disparity Estimation, or BIRDiE (McCartan et al., 2023)) similarly results in worse estimates than our approach.

Our estimation process assigns a probability that each voter is white, Black, Asian, Latino, or "other;" the cross-category probabilities for each voter sum to 1. To estimate the number of ballots cast by, say, Black voters in a given county, we have two options. The first would be to use discrete assignment, calling a voter Black if that is the most probable category assigned to them by our procedure, even if that probability is just a bit more than half—which wouldn't indicate a great deal of certainty. We would then sum all of the voters estimated to be Black in a county, and then the white voters, etc.

We do not use this approach. Rather, to calculate how many Black voters there are within a given county, we sum the predicted probabilities of being Black for each voter living there. In other words, if we had 10 voters who each had a 55% likelihood of being white and a 45% likelihood of being Black, we would estimate that—in aggregate—we had 5.5 white voters and 4.5 Black ones. Discrete assignment, meanwhile, would call each of these 10 voters "white" and therefore overstate both white turnout and the turnout gap. This is the approach taken by some prior work, which relies on Catalist-defined discrete racial estimates (e.g., Billings et al., 2024; Komisarchik and White, 2025). While our chosen path results in non-whole numbers of estimated ballot counts in each county (which is why we stress it here), incorporating the uncertainty in each voter's race provides better overall estimates. To be clear then, we calculate county-level turnout rates by aggregating the individual records to the county level and dividing these estimated ballot counts by the county's CVAP for each racial group from the 5-year ACS estimates ending with the relevant election year.

In all cases, we follow Komisarchik and White (2025) by dropping counties that had small populations of interest (that is, where the nonwhite population was less than 100

for the white–nonwhite models, and where the Black population was less than 100 for the white–Black models) in any year. While the nonwhite share of the population in some small counties is relatively large (King County, Texas, for instance had a total population of 189 in the 2023 ACS, and is 26% Latino), racial estimates from BISG are considerably noisier in these smaller counties. We also note that any errors in BISG should be incorporated into the baseline estimates for each county; the TWFE design is robust to BISG errors unless these errors systematically changed only in the formerly covered counties, and only in the same year as *Shelby County*. This is highly unlikely.

Alternative Source for Self-Identification in 2014, 2016

The "processed" voter files from L2 following the 2014 and 2016 elections do not include self-reported race from the six states with self-reported race in their raw files. The processed files do, however, include voters' unique state voter identification numbers, which we can use to merge the processed files with the raw files. This is how we obtain voters' self-reported race in these two elections. The snapshots do not align perfectly in terms of timing, but we are nonetheless able to match more than the overwhelming majority of all participants to an entry in the raw voter file, where the voter's race (or lack thereof, if they decline to denote it) can be obtained directly. Table A.5.4 shows the share of participants in each state in each year that were successfully matched to the raw voter file. BISG is used to estimate the race of the remaining voters (as it is for those who report "other" for their race, or decline to provide their race). Louisiana stopped using unique statewide IDs in its voter file in 2016, reducing the match rate in that state that year.

Alternative Racial Predictions

We argue that using a BISG algorithm in which the underlying population distribution is drawn from the Citizen Voting Age Population (CVAP) is reasonable because CVAP provides a better estimate of the demographics of *potential voters* than we would obtain

Table A.5.4: Match Rates Between L2 and Raw Voter Files

	Match Rates	
State	2014	2016
AL	100.00%	99.24%
FL	100.00%	100.00%
GA	99.96%	100.00%
LA	98.43%	88.01%
NC	100.00%	100.00%
SC	98.95%	100.00%

by using total population. In areas with many noncitizens of color, or where children are disproportionately nonwhite, using total population overestimates the nonwhite share of the electorate and biases the turnout gap downward.

Here, we present results supportive of that conclusion. We calculate 2020 turnout rates for each racial group in each county in the six states where self-identified race is available (Alabama, Florida, Georgia, Louisiana, North Carolina, and South Carolina). We then compare the turnout rates that BISG would estimate for each racial group using either CVAP, adult population, or total population as the underlying geographic distribution. We remove the counties where the CVAP of the group of interest is less than 100, and then calculate the absolute value of the "error"—that is, the difference between the *actual* turnout rate based on self-reported data, and the *estimated* turnout rate from BISG. Table A.5.5 presents the mean of these county-level errors (the mean absolute error, or MAE) for each racial group using each estimation strategy.

The above estimates all rely on the methodology developed by Imai and Khanna (2016). In the intervening years, however, researchers have proposed new approaches. In 2023, a team of researchers (McCartan et al., 2023) released a working paper implementing what they call Bayesian Instrumental Regression for Disparity Estimation (BIRDiE). To test whether our approach suffers meaningfully relative to this new approach when aggregated to the county level in states with self-identification on their voter files, we also replicate the MAE analysis using BIRDiE, run at the block group level using CVAP. Those estimates are

Table A.5.5: MAE: Different Target Populations for BISG. Counties Weighted Equally

Variable	Primary	Adult	Total	BIRDiE
White	5.3%	7.7%	9.0%	7.2%
Black	14.2%	14.1%	15.2%	21.8%
Latino	8.3%	16.2%	20.7%	8.9%
Asian	52.8%	22.4%	23.3%	60.7%
Nonwhite	10.8%	19.1%	22.5%	13.8%
White–Nonwhite Gap	16.1%	26.8%	31.6%	21.0%
White–Black Gap	19.4%	21.7%	24.1%	29.0%

Table A.5.6: MAE: Different Target Populations for BISG. Counties Weighted by CVAP

Variable	Primary	Adult	Total	BIRDiE
White	3.2%	7.0%	8.5%	4.2%
Black	7.3%	7.9%	9.2%	12.9%
Latino	7.5%	14.9%	18.5%	6.8%
Asian	24.1%	10.9%	11.2%	31.0%
Nonwhite	5.4%	14.0%	17.4%	7.1%
White–Nonwhite Gap	8.7%	21.0%	25.9%	11.4%
White–Black Gap	9.7%	14.1%	17.1%	17.0%

presented in the final column.

Table A.5.5 indicates that the CVAP approach is preferable to the other strategies for white and nonwhite turnout. While the error for Black turnout is higher using the CVAP approach than using adult population, it is nonetheless smaller than when race is estimated using total population. Importantly, using CVAP is clearly far superior to approaches including noncitizens for estimating Latinos' race. Although CVAP results in poorer estimates for Asian Americans than the other approaches using alternate target populations, less than 1.5% of the CVAP in this region is Asian. In no case does BIRDiE return better estimates than our primary approach.

Table A.5.6 once again shows the MAE, but this time weights counties by the relative size of the CVAP of interest in each county. When we weight by CVAP, the BISG approach using CVAP as the target population outperforms both of the other approaches for every racial group. In no case does BIRDiE return better estimates than our primary approach.

This test is likely conservative, as 2020 is the year in which the other approaches are most likely to outperform our preferred approach. Traditional BISG approaches update their underlying geographic racial probabilities only once a decade, following the decennial Census. They are therefore most likely to be "correct" in the year of the decennial Census. Our approach leveraging the 5-year ACS estimates, meanwhile, updates the geographic distribution with every election cycle. This means that in 2014 and 2016—elections far from the decennial Censuses of 2010 and 2020—our approach is likely to perform even better, as they incorporate mid-decade shifts in racial distributions.

We thus conclude that using CVAP as the underlying racial distribution for the BISG analyses is justified. However, as we note below in our discussion of the causal effect of *Shelby County v. Holder* on the racial turnout gap, our results are consistent regardless of how the BISG algorithm is used (due to the computing intensity of the BIRDiE approach, we do not estimate race using BIRDiE for all voters in all states in all years, as we do with the alternative BISG populations).

Additional Exposition of Results

Potential Threats to Difference-in-Differences Assumptions

To be sure, there were major changes in national politics over our period of study that influenced the racial turnout gap: namely, the presidencies of Barack Obama and Donald Trump. While Obama's candidacy probably disproportionately increased Black turnout—and, thus, decreased the racial turnout gap—the opposite may have been true of Trump. This is why accounting for differences between the covered and uncovered parts of the country is so important to the validity of our study: we must assume that, absent *Shelby County*, the turnout gap in covered and uncovered counties would have continued to evolve in parallel. Net of the covariate adjustments demonstrated in Table A.8.1, however, we can be relatively confident that the counties would have moved in parallel. There is no reason to think that Obama disproportionately mobilized Black voters in covered counties relative to the

uncovered counties that are upweighted or matched in our analyses. We do not attribute the full growth of the racial turnout gap in formerly covered counties to *Shelby County*, but rather the growth above and beyond that observed in very similar parts of the country that were not covered by Section 5. Any national trends in American politics should have impacted the turnout gap equally in (formerly) covered and non-covered counties in this region similarly, net of covariates.

Regression Table for Primary Results

Table A.5.7: Estimated Effect of *Shelby County* on Turnout Gaps

Time Period	White–Nonwhite	White–Black
ATT(-1)	-0.004	-0.016*
	(0.004)	(0.005)
ATT(-2)	0.003	0.004
	(0.005)	(0.005)
ATT(0)	0.006	0.013
	(0.005)	(0.005)
ATT(1)	0.005	0.023*
	(0.006)	(0.007)
ATT(2)	0.015*	0.018*
	(0.005)	(0.006)
ATT(3)	0.016	0.028*
	(0.008)	(0.007)
ATT(4)	0.043*	0.044*
	(0.006)	(0.007)
Aggregate ATT	0.017*	0.025*
	(0.005)	(0.005)
County Fixed Effects	✓	✓
Year Fixed Effects	✓	✓

Note:
* $p < 0.05$
Standard errors clustered by county.

Methodological Approach: Pre-Processing and Conditional Parallel Trends

We test the causal effect of *Shelby County v. Holder* on the racial turnout gap using difference-in-differences, comparing jurisdictions where *Shelby County* could have altered election practices (either only at the state level, or at the state and local level) to those where it could not.

As discussed in the chapter, the counties formerly covered by Section 5 of the VRA were very different from the rest of the country. There, we account for these differences by pre-processing the data using genetic matching (Sekhon, 2011). Another approach, which has become more common with the new generation of TWFE / DiD estimators, is to instead condition the parallel trends assumption on covariates. We use the estimator developed by Callaway and Sant'Anna (2021). Alternately, entropy balancing (Hainmueller, 2012) assigns each potential "control" unit—here, counties that were not covered under Section 5 at the time of *Shelby County*—a unique weight such that, in aggregate, control units look similar to treated ones along a broad set of covariates.

We show here that both alternate approaches return very similar results to those reported in Chapter 5 using genetic matching. In both the conditional-on-covariates and entropy balancing approach we use the same covariates as those used in the genetic matching procedure: logged population, the share of the county that is non-Hispanic white and Black, median income, median age, and the proportion of the county's voters that possess a bachelor's degree or higher (all from the 5-year ACS estimates ending in the election year). Democratic vote share is not used in the conditional-on-covariates approach, as it is potentially downstream of treatment. While the entropy balancing approach follows the genetic matching approach used in the chapter of using only 2012 characteristics (the final pre-treatment year), the conditional-on-covariates approach updates these characteristics each year with the 5-year ACS estimates.

As we show in Table A.5.8, both entropy balancing and genetic matching both result in a set of (weighted) control counties that look very similar to the formerly-covered counties

Table A.5.8: Balance Table for Entropy Balancing and Genetic Matching

Variable	Covered Counties	Full Set of Uncovered Counties	Entropy Balanced Uncovered Counties	Genetically Matched Uncovered Counties
Share White	66.0%	85.5%*†	66.0%	69.4%*
Share Black	17.1%	3.6%*†	17.1%†	15.8%
Obama 2012 Vote Share	40.3%	38.4%*	40.3%	40.6%
Population	129,185	84,334*	153,690	115,883
Median Income	$44,628	$46,620*†	$44,628	$44,342
Median Age	39.3	40.5*†	39.3	39.1
Share with Bachelor's Degree or Higher	19.1%	19.9%*	19.1%	18.9%

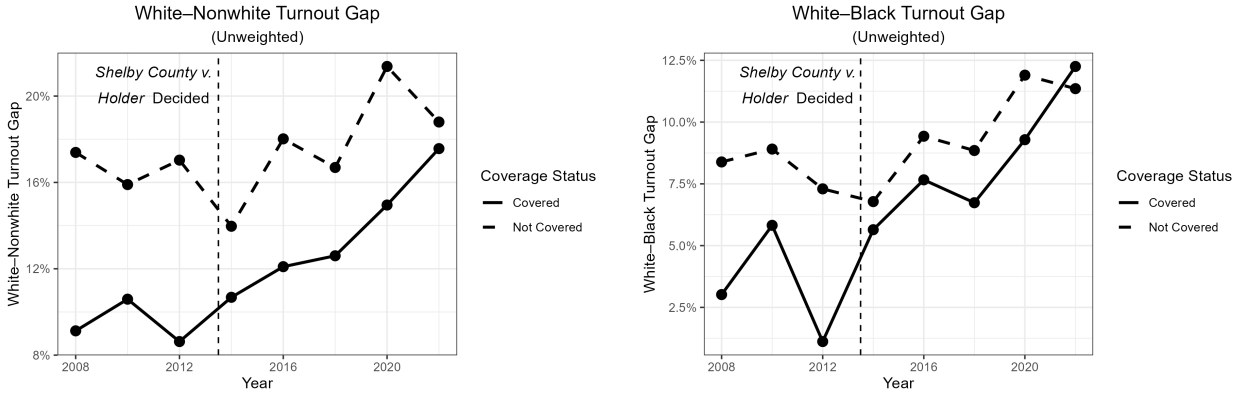
Notes:
Balance table shown for white–nonwhite gap models.
Logged population is used for weighting and matching, but raw population presented here for legibility.
* Mean different from covered counties (t-test, $p < 0.05$).
† Distribution different from covered counties (Kolmogorov–Smirnov test, $p < 0.05$).

both in terms of their means (using t-tests) and distributions (using Kolmogorov–Smirnov tests).

Next we show the results using both the conditional-on-covariates and entropy balancing approaches. We begin by plotting the non-pre-processed time series of the white–nonwhite and white–Black turnout gaps in Figure A.5.1. A number of trends are clear. First, we can see that turnout gaps in formerly covered ("treated") counties were substantially lower in the pre-*Shelby County* 2008–2012 period than in the "control" counties not formerly subject to preclearance. Chief Justice Roberts noted this very trend in his *Shelby County* opinion; he believed it suggested that racial disenfranchisement in 2013 was less acute than it was in 1965, which justified ending preclearance. Yet, Figure A.5.1 also shows that once the treated counties were no longer subject to preclearance, these differences gradually disappeared: The white–nonwhite turnout gap reached parity in 2022, and the white–Black turnout gap was actually *higher* in formerly covered counties by the end of the period.

This said, Figure A.5.1 does not inspire much confidence in the parallel trends assumption. Particularly with respect to the white–nonwhite turnout gap, counties in the two

Figure A.5.1: Turnout Gap Time Series, All Counties Weighted Equally

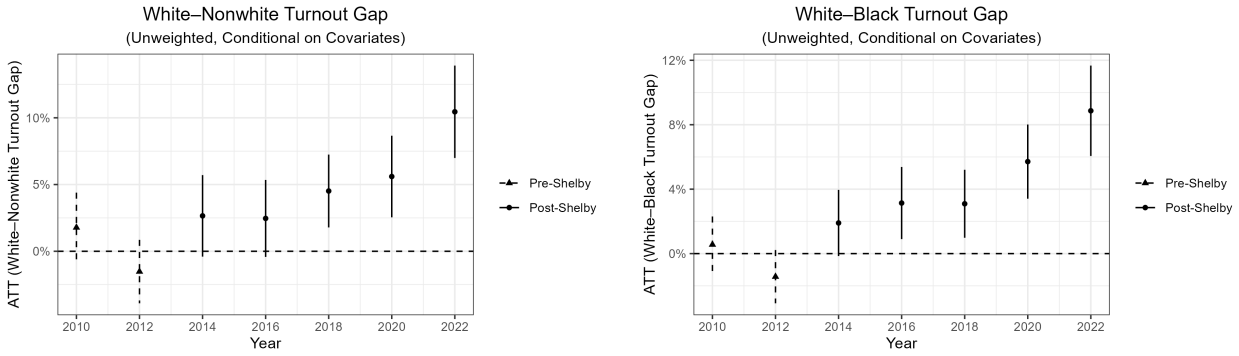


(a) White–Nonwhite Turnout Gap

(b) White–Black Turnout Gap

groups were not moving in tandem prior to 2013. This necessitates the approach we describe above: conditioning our difference-in-differences approach on covariates (in addition to year and county fixed effects). Figure A.5.2 plots the coefficients and error bars for regression models in which we condition on covariates.

Figure A.5.2: Conditional on Covariates



(a) White–Nonwhite Turnout Gap

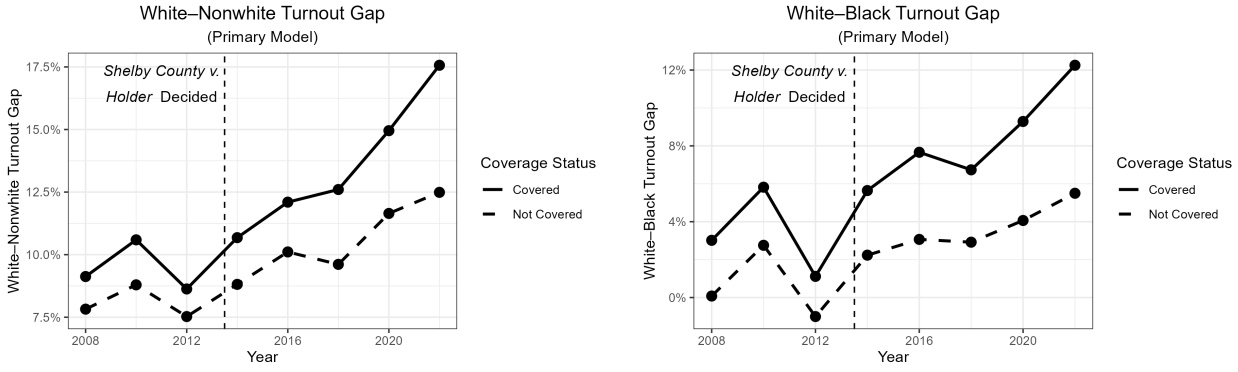
(b) White–Black Turnout Gap

Both coefficient plots indicate that, conditional on covariates, covered and uncovered counties were moving in parallel in the pre-*Shelby County* era. While this cannot prove that the parallel trends assumption is valid in the post-treatment period, it strengthens its plausibility. The figures make clear that the turnout gaps expanded considerably in

treated counties relative to control counties in the aftermath of *Shelby County*. The average treatment effect on the treated counties (ATT) is statistically significant ($p < 0.05$) and large (>4pp) for both outcome variables. The treatment effect also grows markedly over the post-*Shelby County* period. In 2022, both turnout gaps were eight points higher than we would have expected if the preclearance regulatory environment had continued.

Next, we consider the models in which we pre-process our control units using entropy balancing. Although we use a different approach here (entropy balancing rather than genetic matching), the time series plots in Figure A.5.3 is very similar to that presented in Chapter 5. The coefficient plots in Figure A.5.4 further corroborate that our central conclusion—that *Shelby County* increased the turnout gap—is not due simply to our decision to use genetic matching. In sum, this finding holds whether we use genetic matching, entropy balancing, or condition the parallel trends assumption on covariates.

Figure A.5.3: Turnout Gap Time Series (Control Counties Entropy Balanced)



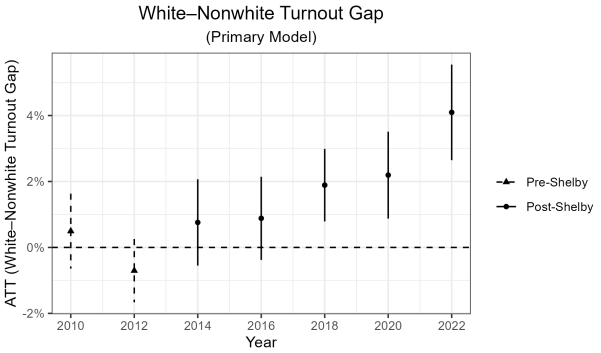
(a) White–Nonwhite Turnout Gap

(b) White–Black Turnout Gap

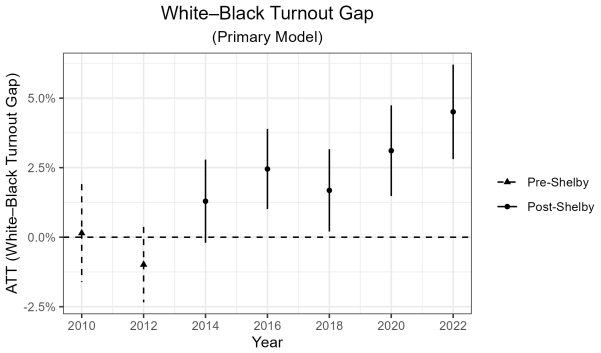
Discussion of Population Weighting

Because we are principally interested in the role of local election administrators in shaping the electoral experience of their voters, we allow all treated counties to be counted equally—regardless of their population. As we have shown, the overwhelming majority of policies proposed under Section 5 were local practices; so too were the majority of the policies that

Figure A.5.4: Entropy-Balanced Models, Unconditional on Covariates



(a) White–Nonwhite Turnout Gap



(b) White–Black Turnout Gap

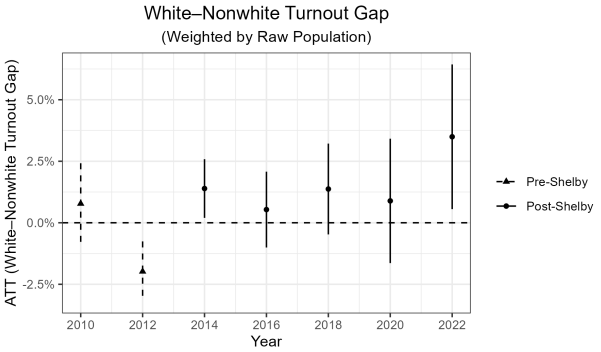
were blocked by the Department of Justice under the expectation that they would have racially disparate effects. If *Shelby County* resulted in an increase in discriminatory policies, we therefore expect that these policies will be widespread across the many local jurisdictions. Weighting all counties equally allows us to keep election administrators — the individuals or agencies with the most latitude to change voting policies in the aftermath of *Shelby County* without attracting attention — as the unit of observation.

Weighting each county equally, however, has some drawbacks. Most importantly, it allows states like Georgia and Texas, which have 159 and 254 counties respectively, to contribute considerably more to our estimates than states like Arizona, with just 15 counties. Because state policy matters for ballot access, we might instead want to weight each state equally. We do that here, showing that our results are broadly consistent.

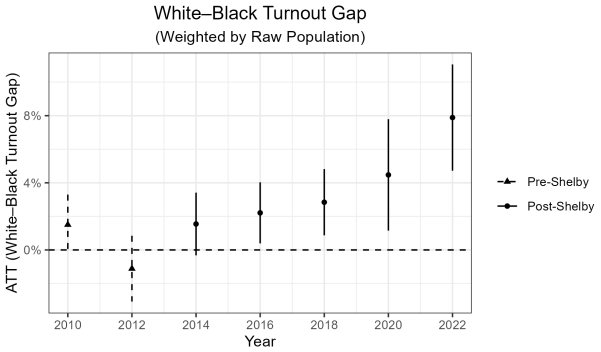
To address these potential issues, the models in Figure A.5.5 weight each county within its home state and year using the raw population of interest (either nonwhite or Black CVAP). This means that large counties matter within states (20% of Texas' weight comes from Harris County, for instance), and that all state-years are weighted equally. We return to the estimator conditioning the parallel trends assumption on covariates (dropping population as a covariate as it is incorporated into the weights), as combining genetic matching and population weights is not particularly intuitive. The ATT for both the white–nonwhite and

white–Black turnout gaps remain positive, substantively large (though smaller than those presented in Figure A.5.2), and statistically significant.

Figure A.5.5: Conditional on Covariates, Weighted by Population within States



(a) White–Nonwhite Turnout Gap



(b) White–Black Turnout Gap

Conditional Average Treatment Effects: Estimating County Heterogeneity

In addition to estimating the average treatment effect for all counties potentially impacted by *Shelby County*, we estimate conditional average treatment effects (CATEs) conditioned on the characteristics detailed above. Doing so allows us to account for the county heterogeneity when our analyses require us to do so. For instance, the model exploring whether local coverage under Section 5 mattered above and beyond state coverage takes the following form:

$$\begin{aligned} Y_{it} =& \beta_1 State\ Covered_i \times Post\ Shelby\ County_t + \\ & \beta_2 State\ and\ County\ Covered_i \times Post\ Shelby\ County_t + \\ & \alpha_i + \lambda_t + \varepsilon_{it} \end{aligned} \tag{1}$$

Where Y_{it} is the turnout gap in county i in year t; *State Covered*$_i$ $\times$ *Post-Shelby County*$_t$ indicates whether the election year is 2014 or later and a county was in a state

at least partially covered by Section 5; *State and County Covered*$_i$ $\times$ *Post-Shelby County*$_t$ identifies counties that were locally covered in the post-treatment period; and $_i$ and $_t$ are vectors of county and year fixed effects, respectively. We use the `fixest` package in `R` to improve computational efficiency; this approach drops the un-interacted *Post-Shelby County*, *State and County Covered*, and *State Covered* variables because they are collinear with the year and county fixed effects. Substantively, $_1$ will measure the effect of *Shelby County* in uncovered counties in partially covered states; $_2$ will capture any local effect beyond the state effect.

The models testing whether the CATE was larger in counties with an objection letter and where Obama won take the same general form. However, in the objection-letter models, we drop the uncovered counties in the partially covered states. Since these counties were not eligible to receive objection letters due to their formal non-coverage under Section 4b, including them risks objection letters simply re-identifying the counties with local coverage.

Recent advances in difference-in-difference estimators make clear that when treatment timing is staggered or when the parallel trends assumption is conditional on covariates, traditional two-way fixed effects (TWFE) estimates return biased estimates. We do not have staggered treatment timing in this case (all covered counties were removed from coverage in 2013) and, as we have shown, using genetic matching to pre-process the data means our parallel trends assumption is credible without conditioning on covariates. In other words, the genetic matching approach allows us to use traditional TWFE estimators for our models, which we do because the new difference-in-differences estimators do not yet allow for conditional treatment effects.

In Chapter 5 we presented the time series plots for these CATE models. Here, we present the regression tables showing that the CATTs were in fact statistically significant

Table A.5.9: Treatment Moderated by Local Coverage

	White-Nonwhite	White-Black
State Covered × Post Shelby County	-0.007	0.006
	(0.005)	(0.006)
State and County Covered × Post Shelby County	0.031*	0.013*
	(0.004)	(0.004)
County Fixed Effects	✓	✓
Year Fixed Effects	✓	✓
Num.Obs.	14159	11234
R2	0.851	0.843
R2 Adj.	0.830	0.820

* $p < 0.05$

Robust standard errors clustered by county.

Table A.5.10: Treatment Moderated by Objection Letter

	White-Nonwhite	White-Black
Covered × Post Shelby County	0.018*	0.013*
	(0.005)	(0.006)
Objection Letters × Post Shelby County	0.003*	0.004*
	(0.001)	(0.001)
County Fixed Effects	✓	✓
Year Fixed Effects	✓	✓
Num.Obs.	9624	8035
R2	0.845	0.805
R2 Adj.	0.822	0.777

* $p < 0.05$

Robust standard errors clustered by county.

Table A.5.11: Treatment Moderated by Obama 2012 Vote Share

	White-Nonwhite	White-Black
Post Shelby County × Obama 2012 Vote Share	0.025	0.062
	(0.027)	(0.033)
Covered × Post Shelby County	0.034*	0.051*
	(0.013)	(0.017)
Covered × Post Shelby County × Obama 2012 Vote Share	-0.045	-0.083*
	(0.030)	(0.037)
County Fixed Effects	✓	✓
Year Fixed Effects	✓	✓
Num.Obs.	14155	11234
R2	0.850	0.843
R2 Adj.	0.829	0.821

* $p < 0.05$

Robust standard errors clustered by county.

Alternative Modeling Approaches

Our results are also robust to a wide variety of other robustness checks, such as using other BISG approaches, limiting the analysis to the former confederacy, and using different statistical approaches for incorporating pre-*Shelby* characteristics. Interested readers can find these many other tests in the supplemental materials of our academic paper, "Did *Shelby County v. Holder* Increase the Racial Turnout Gap?" (Morris and Miller, 2025).

Estimating "Lost" Seats

At the end of Chapter 5, we present a table showing that the number of lower state legislative districts in which minority and Black voters made up a majority of participants in the 2020 election would have been higher had *Shelby County* not been handed down. Here, we explain in greater detail our approach for estimating these numbers.

We begin from our county-level data. We already have estimates, using BISG and the voter files, of how many ballots were cast in each county by voters of each race. These are the data that are used for the empirics throughout this chapter.

Determining the *counterfactual* — that is, how many ballots *would have been cast* by voters of color if preclearance were still operative — is more complicated. This is driven in part because we know that the effect of *Shelby County* varied according to the population of the county: The decision had smaller effects in larger counties.

To estimate the counterfactual in each county, we estimate another two-way fixed effects model, but allow it to be more fully specified than any presented in the chapter. That's because here, coming up with credible estimates of the turnout gap for each of the covered counties is of paramount concern. As with other models including objection letters, we drop the uncovered counties in the partially covered states. The equation takes the following form:

$$\begin{aligned} Turnout\,Gap_{it} = & \beta_1 Covered_i \times Year_t + \\ & \beta_2 Covered_i \times Post\,Shelby\,County_t \times Population_{it} + \\ & \beta_3 Covered_i \times Post\,Shelby\,County_t \times Objection\,Letters_i + \\ & \beta_4 Covered_i \times Post\,Shelby\,County_t \times Obama\,2012\,Vote\,Share_i + \\ & \alpha_{it} + \gamma_{it} + \varepsilon_{it} \end{aligned} \tag{2}$$

This allows the effect of *Shelby County* to vary by year; by population; by objection letters the county received; and by the Democratic vote share. As we showed in the chapter, each of these factors mattered. $_{\text{it}}$ contains a vector of the uninteracted version, and two-way interactions between, each of the above variables. $_{\text{it}}$ includes the time-varying county-level characteristics used in the matching procedure: share non-Hispanic white; share non-Hispanic Black; median income; median age; and share with at least a Bachelor's Degree.

We use this model to come up with two estimates for each county: the turnout gap (whether white–nonwhite, or white–Black, depending on the model) as it actually was, and what the turnout gap *would have been* if preclearance were still in effect. The difference between these two estimates is the estimated treatment effect of *Shelby County* in each county. We use this estimated effect to calculate how many more ballots would have been cast by either nonwhite voters as a whole, or specifically Black voters, if *Shelby County* had not

been handed down. We do this by multiplying the estimated effect by the relevant groups' citizen voting age population. We impose one constraint: Because it is implausible that *Shelby County* reduced the turnout gap anywhere, we force the number of "counterfactual" votes to be at least as high as the number of votes that were actually cast.

At this point, we know how many ballots were cast by voters of color in each county in 2022, and we have estimates of how many *would* have been cast if preclearance were still in effect. Unfortunately, however, counties do not map neatly onto state legislative districts. To determine the number of ballots cast by voters of different races in each state legislative district, we begin by decomposing the votes cast into the census blocks within the county. We allocate the number of white votes to each block according to the share of a county's white residents that live in each block, and do the same for Black and other voters (this assumes that turnout was the same across all blocks in the county). We also do this for the number of counterfactual Black and other nonwhite votes. We then determine the state legislative district that each block falls within, and aggregate up to the state legislative district level. This allows us to see the districts in which Black and nonwhite voters *actually* made up a majority of participants in the 2022 election, and in which they *would* have made up a majority under the counterfactual.

It is important to note that, because these models account for treatment effect heterogeneity by population, these estimates are not driven by large counties. In fact, the models allocate very few "missing" votes to these counties: the estimated effect of *Shelby County* on the white–Black turnout gap is zero in 12 of the 15 largest counties by Black CVAP, and the effect on the white–nonwhite gap is 0 in all 15 of the largest nonwhite counties. The models increase the overall Black turnout turnout rate by 1.4 percentage points across the covered region, and the nonwhite turnout by 1.5 points. Nevertheless, because the effects are concentrated in rural areas, we estimate that Black and other minority voters would have comprised a majority in a handful of additional seats in 2022 — perhaps, given racial polarization, enough to change the outcomes of a few of them.

Chapter 6

Defining Annexations and Section 5 Coverage

In Chapter 6, we do not include all block groups that were ever brought into an incorporated "place" over the time period we study. Instead, we limit our analysis in two ways. First, to be defined as annexed, the block groups must be incorporated into a municipality that already existed in 2009. We do this because we are interested in how annexations in *existing* municipalities changed in response to *Shelby County*, as opposed to the incorporation of new municipalities. About 66% of the block groups that are incorporated between 2010 and 2023 were incorporated into municipalities that existed in 2009. Secondly, we exclude municipalities that more than doubled in size via annexation in any year in our study period, as the concept of vote dilution becomes muddier when the "annexation" so dramatically changes the character of the municipality. Of the 1,019 municipalities we identify as incorporating populated block groups between 2010 and 2023, 68 (6.7%) doubled their population via annexation in at least one year and are thus excluded. We also exclude municipalities in the 9 states (Alaska, Connecticut, Hawaii, Massachusetts, Maine, New Hampshire, New Jersey, Rhode Island, and Vermont) where no annexations occurred over the 2010–2023 period, usually because all land is already incorporated. See Durst (2018) for more on this.

We consider a municipality to be "covered" by preclearance if it is inside a state that was fully covered under the coverage formula, or if its geographical center was inside a covered county.

Further Discussion of Modelling Choices

The model testing the effect of *Shelby County* on the annexation rate relies on a constant set of observations across time; we include every municipality that was incorporated as of 2009, asking whether it incorporated a populated block group in each year. As such, this lends itself to a traditional two-way fixed effects / difference-in-differences design. We condition

the parallel trends assumption on covariates using the approach detailed by Callaway and Sant'Anna (2021).

However, the tests of the effects of *Shelby County* on annexation size and racial characteristics are *conditional on an annexation occurring*. As such, they do not include a consistent set of municipalities across time, and so do not allow analysis in the traditional two-way fixed-effects framework that we have employed throughout this book. For instance, in the period where we have data available, the city of Houston, Texas annexed populated block groups in 2010, and then again in 2015, 2016, and 2017. Houston is not included in the other years when it failed to annex any territory. That said, the underlying assumptions for causal inference are identical: if we assume that the character of the annexations that occurred in both regions would have evolved in parallel, any deviation from that parallelity can be inferred as a causal effect. Like in a more typical difference-in-differences approach, we can assess the plausibility of such an assumption by examining whether the parallel trends condition holds in the pre-treatment period.

Regression Tables

Chapter 7

Calculating Purge Rate

Some readers will note that the EAVS data does not directly report removal *rates*; instead, counties report the raw number of voters that they removed over the two-year period ending each federal election. In an ideal world, we would know how many voters were registered at the *beginning* of the time period, and calculate the purge rate by dividing the number of removals by that count. Instead, however, we know how many voters were registered at the *end* of the period. We thus calculate the purge rate by dividing the number of removals by the sum of the removals plus the total number of registered voters. Put differently, the

Table A.6.1: Estimated Effect of *Shelby County* on Annexation Rate

Time Period	Annexation Rate
ATT(-3)	-0.004 (0.002)
ATT(-2)	0.002 (0.002)
ATT(-1)	-0.002 (0.002)
ATT(0)	0.001 (0.001)
ATT(1)	0.002 (0.002)
ATT(2)	-0.003 (0.003)
ATT(3)	0.001 (0.002)
ATT(4)	0.001 (0.002)
ATT(5)	-0.001 (0.002)
ATT(6)	0.004 (0.002)
ATT(7)	0.001 (0.002)
ATT(8)	0.004 (0.002)
ATT(9)	0.002 (0.002)
Aggregate ATT	0.001 (0.001)
Municipality Fixed Effects	✓
Year Fixed Effects	✓

Note:
* $p < 0.05$
Standard errors clustered by county.
Parallel trends assumption conditional on median income, median age, share with a bachelor's degree or higher, log(population), share white, and share Black.

denominator for our purge rate is the number of registered voters there would have been at the end of the period had no purges occurred.

Table A.6.2: Effect of *Shelby County* on Annexation Characteristics

	Number of Annexed Residents	Relative Whiteness of Incorporated Block Group
ATT(-4)	69.777	0.088*
	(284.149)	(0.032)
ATT(-3)	-1230.781*	-0.087
	(147.396)	(0.062)
ATT(-2)	-813.004*	-0.033
	(182.029)	(0.024)
ATT(0)	21.192	-0.057
	(154.647)	(0.057)
ATT(1)	1624.430*	-0.033
	(322.128)	(0.030)
ATT(2)	687.505*	0.171*
	(201.588)	(0.052)
ATT(3)	911.432*	0.134*
	(332.221)	(0.053)
ATT(4)	1277.453*	0.153*
	(172.037)	(0.028)
ATT(5)	-386.362*	0.141*
	(136.504)	(0.019)
ATT(6)	327.562*	0.160*
	(118.275)	(0.030)
ATT(7)	97.115	-0.012
	(169.773)	(0.033)
ATT(8)	80.477	0.012
	(130.395)	(0.019)
ATT(9)	453.107	-0.020
	(300.487)	(0.036)
Log(Population)	378.454*	0.021
	(97.771)	(0.011)
Share White	-1285.649	
	(684.613)	
Share Black	-1344.154	
	(734.705)	
Median Income	0.011	-0.000
	(0.008)	(0.000)
Median Age	-5.738	0.000
	(7.010)	(0.003)
Bachelor's Degree or Higher	1545.516	-0.194
	(1358.353)	(0.095)
Num.Obs.	1226	1604
R2	0.127	0.034
R2 Adj.	0.102	0.015
Year Fixed Effects	✓	✓

* $p < 0.05$

Robust standard errors clustered by year and municipality.

Difference in Differences

The structure of the EAVS data mean that we cannot use quite the same approach here as we used in Chapter 5. When it came to the racial turnout gap, we were able to estimate the turnout gap for every single county in the country from 2008 to the present. The EAVS data, however, have holes: not all counties reported their removal numbers in all years. In 2008, for instance (the first year in which there was relatively widespread participation), we have comprehensive data for 2,846 counties; in 2018, we had data for 3,051 counties; and in 2022 (the latest date for which data are available), the number declined to 2,977. No Texas counties, for instance, reported their data in 2012 — but we still want to understand whether *Shelby County* influenced purge rates there, and their collective reporting rate across the other years is 94%. Missingness like this means that matching is not ideal for this context; a county that serves as a good control in, say, 2014 might not have submitted its purge numbers to EAVS in 2012 or 2018. To get around this issue, we instead use a statistical model developed by Callaway and Sant'Anna (2021) that allows us to make the assumption that covered and uncovered counties would have moved in parallel, *after* adjusting for differences in counties' characteristics. As before, we can test the validity of this assumption by examining whether the two groups of counties moved in parallel in the years prior to *Shelby County*. We note that throughout this chapter, we retain all county observations, regardless of whether they reported their information every year, or in only one. Our results are consistent (unless otherwise noted) if we instead limit the analysis to only the counties that reported their removal numbers every year.

For the purposes of visual clarity, however, Figure 7.1 uses a matching model wherein we keep the counties that report in every cycle. This allows us to present the time-series of purge rates for treated and control units *after* accounting for the differences in their socioeconomic composition. This means that the plot is not directly comparable to the regression estimates, since it uses both a different empirical approach (matching instead of conditioning the parallel trends assumption on covariates) and set of observations. Nevertheless, it

corroborates the broad story that *Shelby County* had an effect on the voter purge rate.

Regression Tables

Table A.7.1: Estimated Effect of *Shelby County* on Removal Rate

Time Period	State and Local Practices Covered	At Least State Practices Covered
ATT(-1)	-0.023	-0.009
	(0.013)	(0.013)
ATT(-2)	-0.019	-0.028
	(0.009)	(0.016)
ATT(0)	0.052*	0.030
	(0.013)	(0.014)
ATT(1)	0.020	0.002
	(0.014)	(0.019)
ATT(2)	0.041*	0.018
	(0.014)	(0.017)
ATT(3)	0.022	0.010
	(0.011)	(0.014)
ATT(4)	0.017	0.006
	(0.012)	(0.012)
Aggregate ATT	0.030*	0.013
	(0.011)	(0.014)
County Fixed Effects	✓	✓
Year Fixed Effects	✓	✓

Note:
* $p < 0.05$
Standard errors clustered by county.
Parallel trends assumption conditional on median income, median age, share with a bachelor's degree or higher, population, share white, and share Black share voting for Obama in 2012, and share moved in past 12 months.

Table A.7.2: Estimated Effect of *Shelby County* on Removal Rate by Objection Letters

Time Period	Objection Letter	No Objection Letter
ATT(-1)	-0.041	-0.017
	(0.018)	(0.009)
ATT(-2)	-0.018	-0.005
	(0.013)	(0.007)
ATT(0)	0.054*	0.048*
	(0.022)	(0.009)
ATT(1)	0.026	0.022*
	(0.016)	(0.008)
ATT(2)	0.044*	0.040*
	(0.017)	(0.009)
ATT(3)	0.020	0.026*
	(0.014)	(0.007)
ATT(4)	0.023	0.016
	(0.014)	(0.009)
Aggregate ATT	0.033*	0.030*
	(0.014)	(0.008)
County Fixed Effects	✓	✓
Year Fixed Effects	✓	✓

Note:
* $p < 0.05$
Standard errors clustered by county.
Parallel trends assumption conditional on median income, median age, share with a bachelor's degree or higher, population, share white, and share Black share voting for Obama in 2012, and share moved in past 12 months.

Chapter 8

Party Loyalty Scores

In Chapter 8 we analyze the partisan composition of the population of voters who did and did not file a reasonable impediment declaration. Because Texas is an open primary state that does not require partisan registration, this means that we must estimate voters' partisanship using information from the voter file. Following Dowling, Miller and Morris (2025), we do this using voters' past participation in primary elections. We begin by summing the number of federal Democratic and Republican primaries in which each voter participated in the

ten years prior to the treatment year. Participation in a Democratic primary is coded as -1, while participation in a Republican one is coded as 1. If a voter participated in all 5 federal primaries as a Democrat, this number would thus be -5; if they did the same but instead in Republican primaries, it would be 5. If they instead participated in either a mix of Democratic and Republican primaries, this number would fall somewhere in between -5 and 5, as voting in a Democratic primary in one year and a Republican one in another year "cancel out." Similarly, if a voter did not participate in all 5 primaries, this number would be somewhere between -5 and 5. Finally, we divide these scores by the number of primaries in which the voter actually participated. Thus, a voter who voted in all 5 Democratic primaries would have a score of -1; so too would a voter who only voted in a single primary, but did so in a Democratic one. The roughly 40% of voters who did not participate in a recent federal primary are assigned a partisan score of 0.

Difference in Differences

We argue that if the pre-treatment turnout of individuals who used an RID in election$_{t=0}$ moved in parallel with the turnout of individuals who also participated that year but did *not* use an RID, any deviation from parallel trends in the post-treatment period can be attributed to the RID in election$_{t=0}$. Of course, as Fraga and Miller (2022) make clear, there are real differences between the types of voters who use RIDs and those who do not. RID filers are, for instance, more likely to be voters of color, and typically have lower pre-treatment turnout rates. We further show here (see Table A.8.1) that RID users are considerably younger than other voters; are more likely to be women; and are slightly more Democratic. These underlying differences therefore probably make the full set of non-RID users a poor comparison group, and might lead to violations of the parallel trends assumption.

To ensure that our control group is as similar to our treatment group as possible and thus, to minimize the likelihood that the parallel trends assumption would be violated, we match each RID-using voter to another voter using a genetic matching algorithm (Sekhon,

Table A.8.1: Balance Table for Genetic Matching

Variable	RID Voters	Matched Non-RID Voters	All Non-RID Voters
Share White	53.0%	53.2%	55.2%*
Share Black	20.1%	20.1%	15.2%*†
Share Latino	22.3%	22.3%	24.7%*
Share Asian	2.7%	2.6%	3.2%*
Share AIAN	0.2%	0.2%	0.2%
Share Female	58.8%	58.8%	54.7%*
Share Male	41.0%	41.0%	45.3%*
Partisan Score	-0.1	-0.1	0.1*†
Age	41.3	41.3	51.6*†
Registration Year	2013	2013	2005*†

Notes:

* Mean different from RID users (t-test, $p < 0.05$).

† Distribution different from RID users (Kolmogorov–Smirnov test, $p < 0.05$).

2011). Because using an RID is conditional on actually turning out and we therefore cannot identify the non-participants that would have used an RID if they had turned out, we match each RID filer to someone who also voted in 2016, but did not use an RID.[4] We also force matches to come from the same county. Inexact matching is done on racial probabilities, gender, partisan score, age, registration year, and pre-treatment turnout.[5] Matching is done with replacement, and ties are randomly broken. Table A.8.1 indicates that the matching was successful: The means and distributions of all covariates are well-balanced for the treated voters who filed an RID and their matched controls. This process is repeated for voters using RIDs in 2018 and 2020.

Once we create our set of treated and matched control voters, we run two-way fixed effects models. We include fixed effects for both individual and year, with robust standard errors clustered at the individual level. Our approach is different from the "staggered"

[4]If a voter used an RID in multiple years, they are only included in the first election in which the RID was used.

[5]As Daw and Hatfield (2018) explain, using pre-treatment outcomes in a matching procedure risks selecting controls who will regress to the mean outcome level in the post-treatment period. This is less likely to be a problem in our case than others: voting is highly autocorrelated within individuals (Green and Shachar, 2000; Coppock and Green, 2016), which Daw and Hatfield (2018) argue minimizes the threat of regression to the mean.

difference-in-difference method that has received considerable attention in the literature in recent years (see, for instance Roth et al., 2023). Staggered difference-in-differences designs feature a constant set of units over time, some of which are "treated" by an intervention at different times. Our set of units, however, is not constant: because using an RID in, say, 2018 is conditional on turning out that year, inclusion in the set of potential control is also conditioned on voting that year. The same is true of all years. Thus, if a (non-RID) voter participated in 2018 but not 2016, they are considered "available" to serve as a control only in the second election (of course, if they voted in both years, they could serve as a control for voters using an RID in both years). For this reason, we report the effects of RIDs on subsequent turnout for each treatment year separately, as well as the "overall" or combined effects.

Regression Tables

Here, we include the full regression tables that correspond to the confidence interval plots we include in the main text. Table A.8.2 shows the causal estimates returned by the two-way fixed effects models testing the average treatment effect of an RID on subsequent turnout for all RID-filers.

In Table A.8.3, we break these effects out by whether the voter in 2016 used a hardship or a non-hardship RID.

Chapter 9

Data

We joined the rejection records to a snapshot of the Texas registered voter file dated May 5th, 2022, obtained from data provider L2.

Many counties did not report rejected ballot applications to the state. Across all of Texas' 254 counties, 88 reported zero application rejections to the state; another (Travis)

Table A.8.2: Effect of an RID on Subsequent Turnout

	2016	2018	2020	Overall
Treated × First Post-Treatment Election	-0.110***	-0.104***	-0.042***	-0.080***
	(0.007)	(0.011)	(0.007)	(0.004)
Treated × Second Post-Treatment Election	-0.139***	-0.090***		-0.121***
	(0.007)	(0.012)		(0.005)
Treated × Third Post-Treatment Election	-0.111***			-0.104***
	(0.007)			(0.006)
Voter Fixed Effects	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓
Voter-Year Fixed Effects	✓	✓	✓	✓
Num.Obs.	228624	76594	173028	478246
R2	0.480	0.534	0.483	0.504
R2 Adj.	0.406	0.457	0.380	0.422

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Standard errors clustered by voter.

Point estimates for pre-treatment years not shown.

Coefficients measured relative to turnout in final pre-treatment election.

Table A.8.3: Effect of a RID on Subsequent Turnout by RID Type

	Hardship	Non-Hardship
Treated × 2018	-0.139***	-0.096***
	(0.011)	(0.008)
Treated × 2020	-0.168***	-0.125***
	(0.012)	(0.008)
Treated × 2022	-0.132***	-0.101***
	(0.011)	(0.008)
Voter Fixed Effects	✓	✓
Year Fixed Effects	✓	✓
Num.Obs.	75056	153568
R2	0.454	0.492
R2 Adj.	0.376	0.419

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Standard errors clustered by voter.

Point estimates for pre-treatment years not shown.

reported an improbably low number given its size. The example of Bexar County further corroborates our assumption that many of the large counties are missing data. Although Bexar County reported no application rejections to the state, a county administrator said under oath in a deposition that some applications were rejected. The transcript of the deposition is available from the authors upon request.

Unfortunately, it is impossible to distinguish counties that *rejected applications but did not report them* from counties that *rejected zero applications*. While we can be sure that some large counties with zero reported application rejections like Hidalgo—home to more than 860,000 residents—are probably missing data, it is more difficult to know whether small counties with zero rejections like Loving (population 117) are missing rejections.

To rectify these data issues, we sent public records requests to each of the 11 counties with populations over 50,000 with no rejected applications (Bell, Bexar, Ector, El Paso, Fort Bend, Hays, Hidalgo, Nueces, Potter, Waller, and Webb), as well as Travis County. Despite multiple attempts to receive the data, just three (El Paso, Travis, and Webb) gave us responsive documents.

In Chapter 9, we exclude the nine remaining large counties from our analyses under the assumption that they are missing data; smaller counties are included under the assumption that they are not missing data. Tables A.9.1 and A.9.2 indicate that our results are not sensitive to these missing data. In Table A.9.1, we include all counties. In other words, we assume that all counties with no reported application rejections really had a 0% rejection rate; all voters who submitted applications in these counties are assumed to have done so successfully. Table A.9.2 takes the opposite approach, assuming that all counties with no reported application rejections are missing data. Under either assumption, application and ballot rejection rates for SB1 related reasons are higher for racial and ethnic minorities.

Table A.9.1: Rejection Rates, 2022 Primary

	Application Rejections		Ballot Rejections	
	SB1	Non-SB1	SB1	Non-SB1
P(Black)	0.008***	0.013***	0.062***	0.001
	(0.002)	(0.003)	(0.009)	(0.001)
P(Latino)	0.008***	0.006**	0.050***	0.001
	(0.002)	(0.002)	(0.009)	(0.001)
P(Asian)	0.015***	0.006*	0.058***	0.006
	(0.003)	(0.003)	(0.010)	(0.005)
P(Other Race)	0.002	0.005	0.009	-0.004
	(0.004)	(0.005)	(0.010)	(0.005)
Age	0.001***	0.000***	0.003***	0.000
	(0.000)	(0.000)	(0.001)	(0.000)
Male	-0.002***	-0.002**	-0.008***	-0.001
	(0.001)	(0.001)	(0.002)	(0.001)
Partisan Score	0.006***	0.006***	0.004	0.002*
	(0.001)	(0.001)	(0.006)	(0.001)
Past Primary Turnout	-0.002***	-0.006**	-0.006***	-0.001*
	(0.000)	(0.002)	(0.001)	(0.001)
County Fixed Effects	✓	✓	✓	✓
Num.Obs.	257915	257915	195016	195016
R2	0.024	0.033	0.067	0.079
R2 Adj.	0.023	0.032	0.066	0.078

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

"P(White)" is reference category for race.

Partisan score determined by partisan primary participation,
2010–2020. It ranges from -1 (most Democratic) to 1 (most Republican).

Past primary turnout is the number of midterm primaries
a voter participated in between 2010 and 2018.

Standard errors clustered by county.

Table A.9.2: Rejection Rates, 2022 Primary

	Application Rejections		Ballot Rejections	
	SB1	Non-SB1	SB1	Non-SB1
P(Black)	0.010***	0.015***	0.062***	0.002
	(0.002)	(0.003)	(0.010)	(0.002)
P(Latino)	0.009***	0.007**	0.045***	0.002
	(0.002)	(0.002)	(0.011)	(0.002)
P(Asian)	0.018***	0.006	0.051***	0.007
	(0.003)	(0.003)	(0.007)	(0.006)
P(Other Race)	0.002	0.006	0.013	-0.004
	(0.005)	(0.006)	(0.010)	(0.006)
Age	0.001***	0.000***	0.003***	0.000
	(0.000)	(0.000)	(0.001)	(0.000)
Male	-0.002***	-0.003**	-0.009***	-0.001
	(0.001)	(0.001)	(0.003)	(0.001)
Partisan Score	0.007***	0.008***	0.003	0.002*
	(0.002)	(0.001)	(0.008)	(0.001)
Past Primary Turnout	-0.003***	-0.007***	-0.007***	-0.002*
	(0.001)	(0.002)	(0.001)	(0.001)
County Fixed Effects	✓	✓	✓	✓
Num.Obs.	208397	208397	155396	155396
R2	0.019	0.028	0.061	0.083
R2 Adj.	0.018	0.027	0.060	0.082

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

"P(White)" is reference category for race.

Partisan score determined by partisan primary participation,
2010–2020. It ranges from -1 (most Democratic) to 1 (most Republican).

Past primary turnout is the number of midterm primaries
a voter participated in between 2010 and 2018.

Standard errors clustered by county.

Rejections by Type

Table A.9.3 reports the share of applications and ballots rejected (in the retained counties) by reason.

Table A.9.3: Rejections by Reason, 2022 Primary

Rejection Reason	Applications Rejected	Ballots Rejected
SB1 (Incorrect or Missing Number)	45.1%	88.3%
Received Late	31.7%	8.9%
No Signature	4.6%	1.5%
Other	4.4%	1.3%
Party Missing	4.6%	
Not Eligible for VBM	4.2%	
Not Registered	3.9%	
Improperly Delivered	1.5%	
Total	100.0%	100.0%

Modeling Approach

We construct linear probability models wherein the binary indicators of an application or ballot rejection serve as the dependent variable, and the various probabilities of a voter's racial or ethnic identity are the independent variables of interest.

We include p(Black), p(Latino), p(Asian), and p(other) in the model, omitting p(white). The coefficients on these variables operate similarly to discrete dummy variables: the coefficient on P(Latino), for instance, measures the rejection rate of a voter who is 100% likely to be Latino, relative to a voter who is 100% likely to be white. If the coefficients for these variables are positive and significant, it indicates that relative to voters with a higher probability of being white, those with a higher probability of being Black, Latino, Asian, or other had a greater likelihood of experiencing a rejection—which would be consistent with Hypotheses 1a and 1b.

We include as covariates voters' age, gender, and past turnout, which we also obtained from the L2 voter file. We follow the approach detailed in the Appendix to Chapter 8 for

calculating partisan loyalty. Finally, because county administrators had some discretion in terms of helping voters cure their ballots and applications, we include county (and county-year) fixed effects.

We note that whenever we are testing the disparities in, or effects of, an *application* rejection, the study population is comprised of all individuals who submitted a mail ballot application. For the ballot rejection models, however, we include only those individuals who actually cast a mail ballot, excluding individuals whose applications were rejected as well as those who successfully requested a mail ballot but did not cast it. These voters are excluded because we cannot know whether their ballot would have been accepted or rejected had they cast it.

Rejection Regression Table

Here, we show the full regression table for application and ballot rejections.

White–Nonwhite Regressions

In Chapter 9 we present regression tables where distinct racial categories are used as predictors of the likelihood of a ballot or application rejection. Here, we test the overall rejection rate for nonwhite voters (calculated by subtracting the predicted likelihood of being white from 1).

Table A.9.4: Effect of a Primary Rejection on General Turnout

	Application Rejected		Ballot Rejected	
	(1)	(2)	(3)	(4)
Treated × 2022	-0.161***	-0.163***	-0.015***	-0.014***
	(0.007)	(0.007)	(0.003)	(0.003)
Voter Fixed Effects	✓	✓	✓	✓
Year Fixed Effects	✓		✓	
County Fixed Effects	✓		✓	
County-Year Fixed Effects		✓		✓
Num.Obs.	861920	861920	645304	645304
R2	0.471	0.483	0.466	0.480
R2 Adj.	0.295	0.310	0.287	0.306

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Standard errors clustered by individual.
Includes 2010–2022 midterm general elections.

Difference-in-Differences Regression Tables

The following tables present the econometric results of the two-way fixed effects models used to estimate the causal effect of an application or ballot rejection on subsequent behavior.

Turnout in primary elections is somewhat more complicated to model than in general elections: a given primary might be competitive for only a single party, for instance. Moreover, the characteristics of individuals who experienced an application or ballot rejection in 2022 might have shaped their turnout in 2024 in ways independent of the rejection. These voters were older than other voters, and were slightly more likely to have been removed from the rolls by the time of the 2024 election (perhaps due to higher mortality rates). Whenever a voter is no longer on the rolls, we code them as not participating. Our results are consistent if we limit the pool to individuals still registered as of the 2024 primary election. As such, in the models presented in Table A.9.6, we include year fixed effects interacted with various other characteristics (county, age, and party score). By way of reminder, we use past partisan primary participation to identify voters' partisan scores; voters with a score below (above) 0 participated in more (few) Democratic than Republican primaries. This variable

Table A.9.5: Effect of a Primary Rejection on Voting In-Person in General (Conditional on Voting)

	Application Rejected		Ballot Rejected	
	(1)	(2)	(3)	(4)
Treated × 2022	0.350***	0.356***	0.173***	0.146***
	(0.009)	(0.009)	(0.004)	(0.004)
Voter Fixed Effects	✓	✓	✓	✓
Year Fixed Effects	✓		✓	
County Fixed Effects	✓		✓	
County-Year Fixed Effects		✓		✓
Num.Obs.	688943	688943	524437	524437
R2	0.642	0.656	0.653	0.668
R2 Adj.	0.480	0.500	0.500	0.520

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Standard errors clustered by individual.
Includes 2010–2022 midterm general elections.

is a continuous one, running from -1 to 1. In these party fixed effects models, we group voters into 3 discrete categories depending on whether their partisan score is less than 0 (Democrats), 0 (Independents), or greater than 0 (Republicans). This approach effectively compares the 2024 turnout of voters who were disenfranchised in 2022 to non-disenfranchised voters with similar characteristics, ensuring valid inference.

Table A.9.6: Effect of a Primary Rejection on 2024 Primary Turnout

	Application Rejected				Ballot Rejected			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treated × 2024	-0.130***	-0.137***	-0.094***	-0.099***	-0.097***	-0.107***	-0.068***	-0.068***
	(0.007)	(0.008)	(0.007)	(0.007)	(0.005)	(0.005)	(0.004)	(0.004)
Voter Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Year-County Fixed Effects		✓	✓	✓		✓	✓	✓
Year-Age Fixed Effects			✓	✓			✓	✓
Year-Party Fixed Effects				✓				✓
Num.Obs.	861920	861920	861092	861092	645304	645304	644748	644748
R2	0.435	0.444	0.465	0.488	0.427	0.438	0.458	0.481
R2 Adj.	0.246	0.258	0.285	0.316	0.236	0.249	0.276	0.306

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Standard errors clustered by individual.

Includes 2012–2024 presidential primaries

Table A.9.7: Effect of a Primary Rejection on Voting In-Person in 2024 Primary (Conditional on Voting)

	Application Rejected				Ballot Rejected			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treated × 2024	0.145***	0.154***	0.147***	0.142***	0.158***	0.145***	0.127***	0.128***
	(0.012)	(0.012)	(0.011)	(0.011)	(0.007)	(0.007)	(0.007)	(0.007)
Voter Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Year-County Fixed Effects		✓	✓	✓		✓	✓	✓
Year-Age Fixed Effects			✓	✓			✓	✓
Year-Party Fixed Effects				✓				✓
Num.Obs.	496562	496562	496167	496167	389557	389557	389262	389262
R2	0.603	0.620	0.628	0.630	0.595	0.612	0.622	0.623
R2 Adj.	0.340	0.366	0.380	0.383	0.338	0.364	0.379	0.381

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Standard errors clustered by individual.

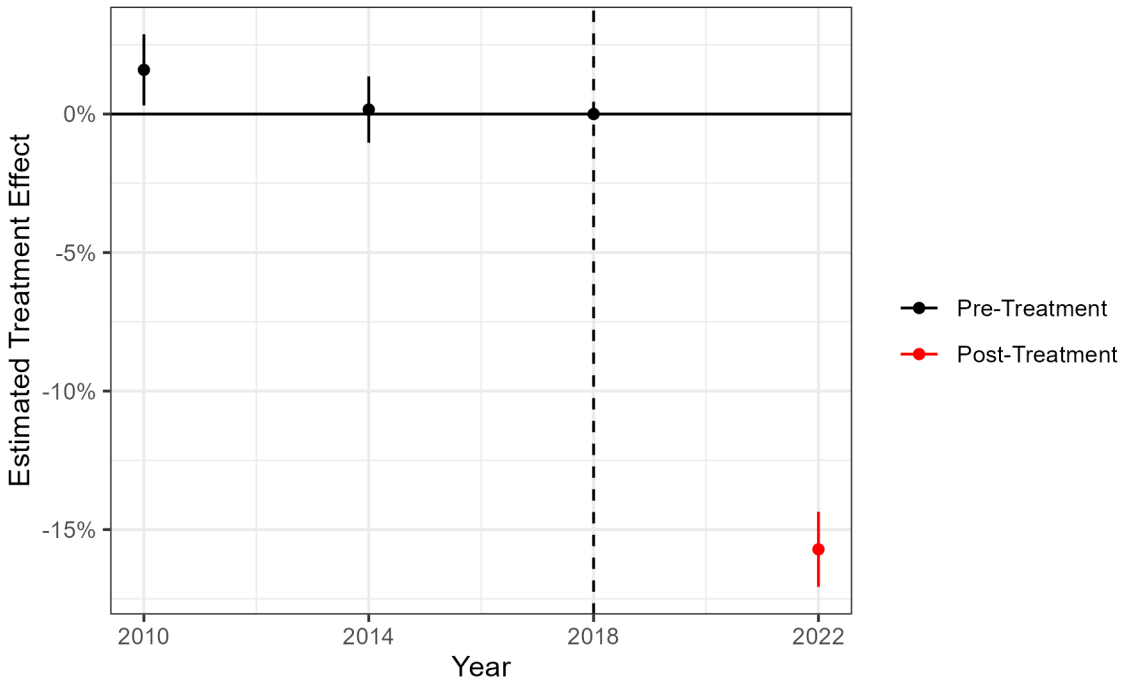
Includes 2012–2024 presidential primaries

Event Study Plots

2022 Election: Turnout

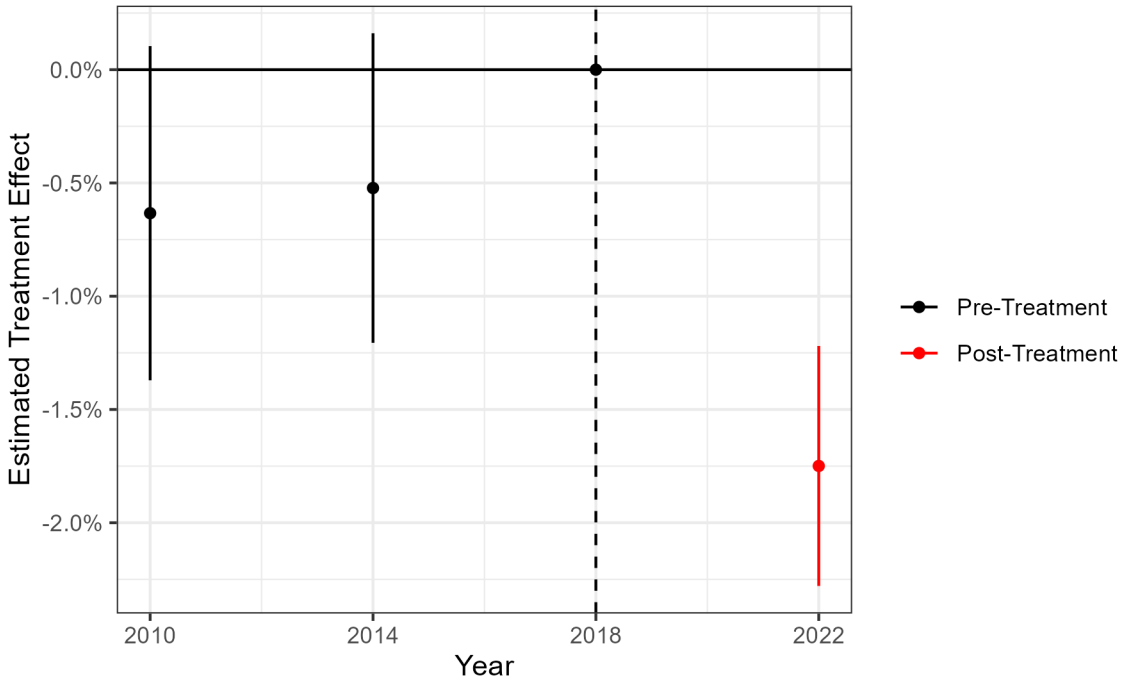
In Chapter 9 we present several general difference-in-differences models testing the effects of application and ballot rejections on turnout in the 2022 general and 2024 primary elections. Here, we present the event study plots for each DID, using in each case the model with the most fully specified set of fixed effects. We note that here, the coefficient on "treated" is measured relative to the final pre-treatment year, and not relative to the entire base period (as the estimations in the regression tables in the manuscript are done).

Figure A.9.1: Event Study, Application Rejection, 2022 General Election Turnout



All years measured relative to 2018 (final pre-treatment midterm election). Includes voter and county-year fixed effects. Standard errors clustered by voter.

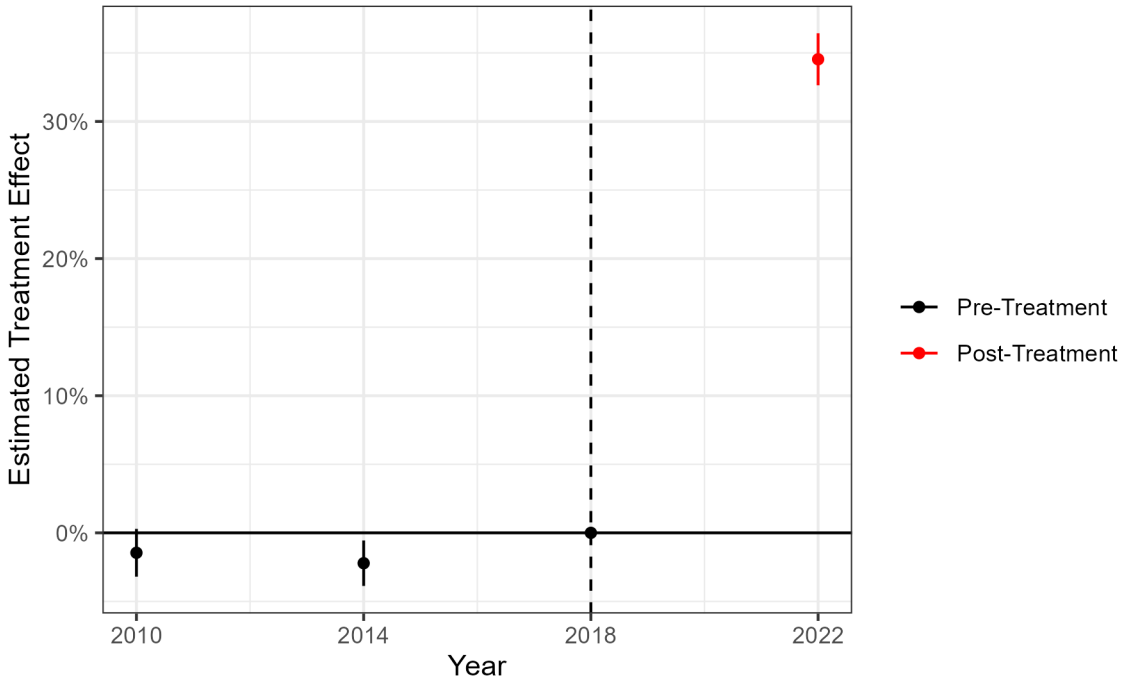
Figure A.9.2: Event Study, Ballot Rejection, 2022 General Election Turnout



All years measured relative to 2018 (final pre-treatment midterm election). Includes voter and county-year fixed effects. Standard errors clustered by voter.

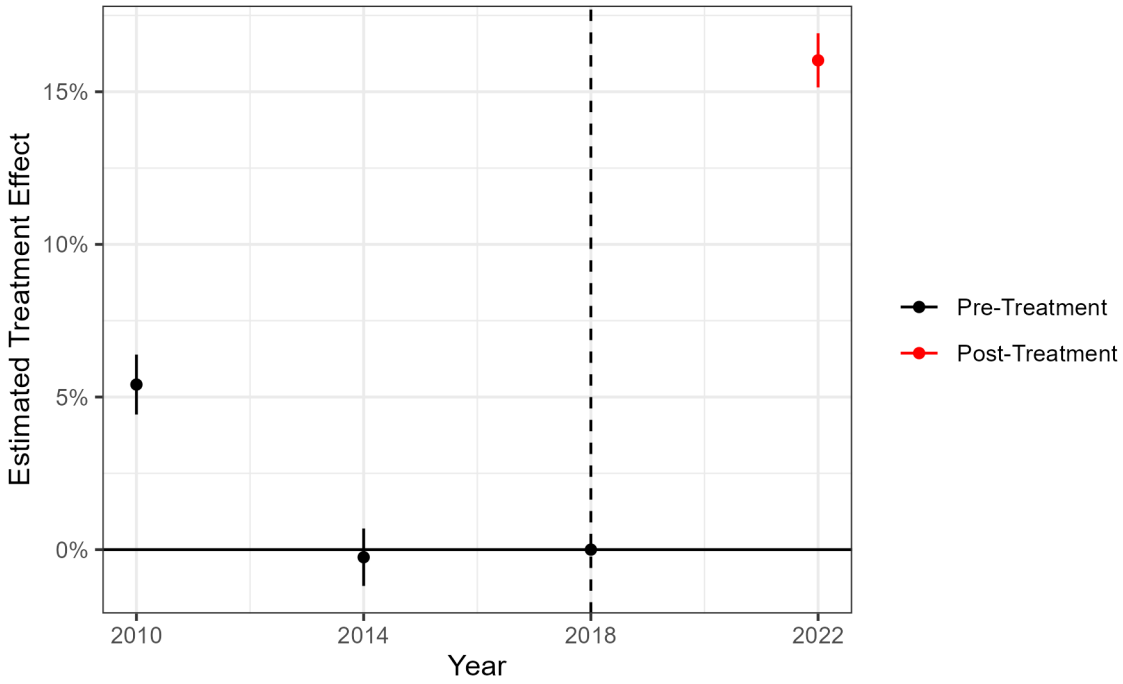
2022 Election: Voting in Person

Figure A.9.3: Event Study, Application Rejection, 2022 General Election In-Person Voting (Conditional on Voting)



All years measured relative to 2018 (final pre-treatment midterm election). Includes voter and county-year fixed effects. Standard errors clustered by voter.

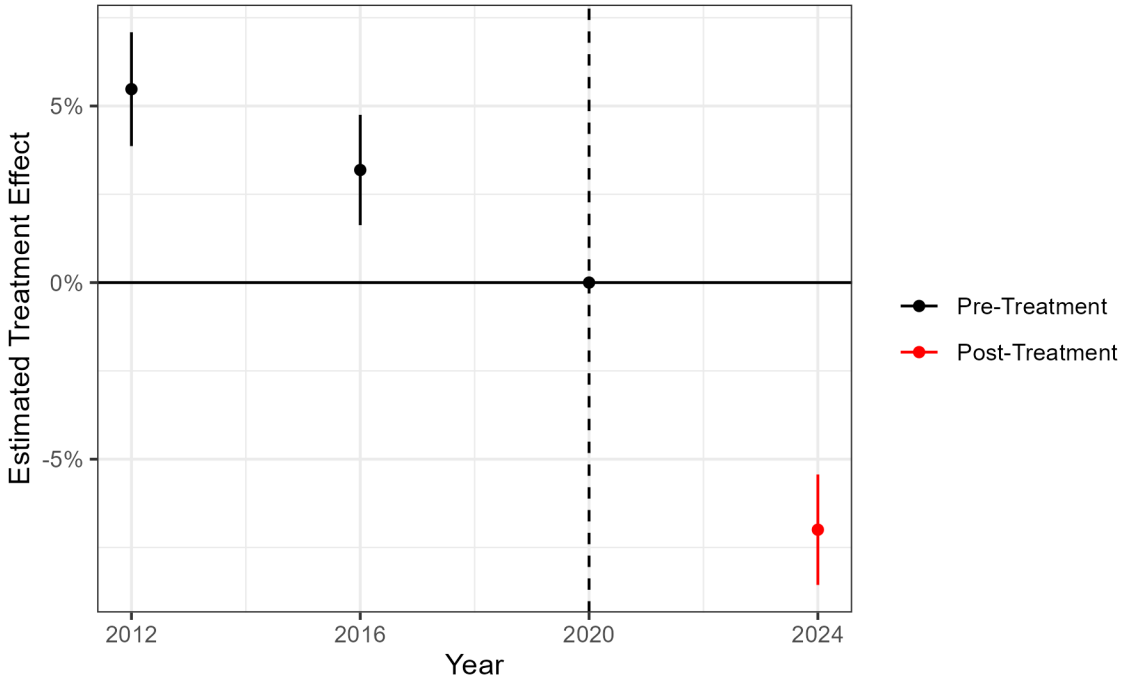
Figure A.9.4: Event Study, Ballot Rejection, 2022 General Election In-Person Voting (Conditional on Voting)



All years measured relative to 2018 (final pre-treatment midterm election). Includes voter and county-year fixed effects. Standard errors clustered by voter.

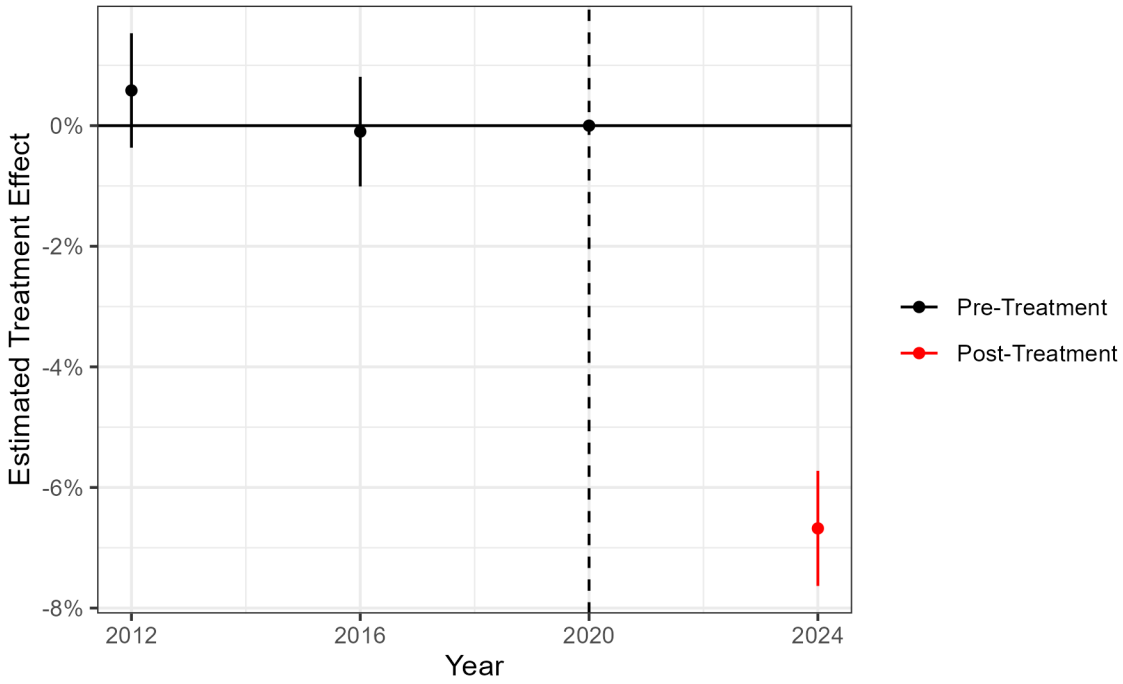
2024 Election: Turnout

Figure A.9.5: Event Study, Application Rejection, 2024 Primary Election Turnout



All years measured relative to 2020 (final pre-treatment primary election). Voter, year-county, year-age, and year-party fixed effects included. Standard errors clustered by voter.

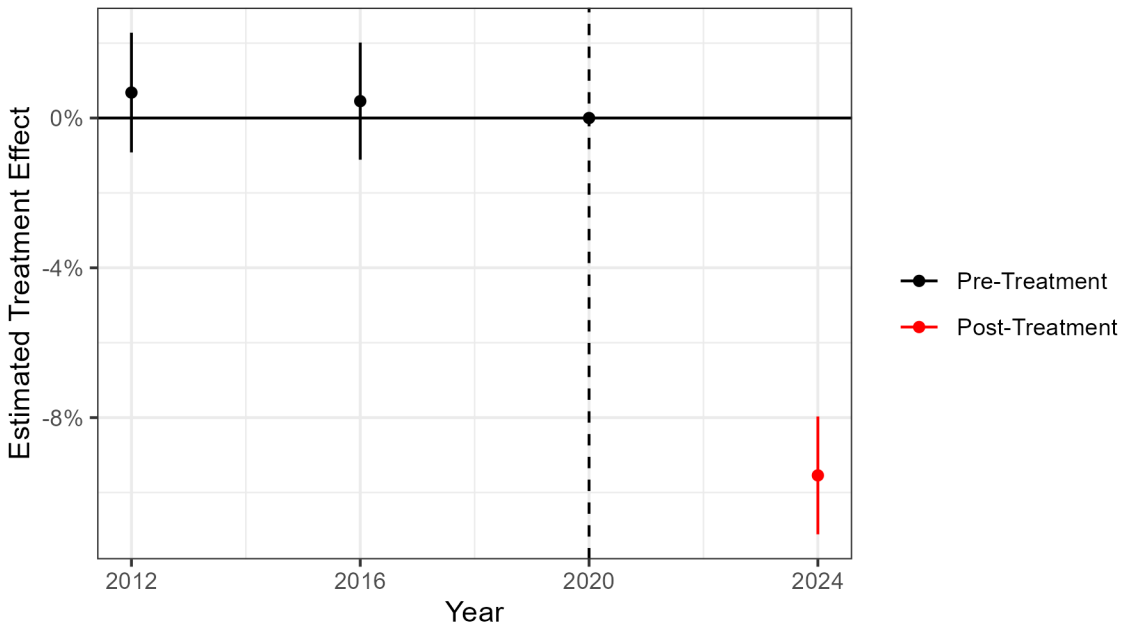
Figure A.9.6: Event Study, Ballot Rejection, 2024 Primary Election Turnout



All years measured relative to 2020 (final pre-treatment primary election). Voter, year-county, year-age, and year-party fixed effects included. Standard errors clustered by voter.

With the exception of the model testing the effect of an application rejection on turnout in the 2024 primary election, the event study plots indicate that the parallel trends assumption likely holds. In the case of that one model, however, there is some reason for concern about pre-trends. We therefore re-estimate this model, but pre-preprocess the control units using entropy balancing (Hainmueller, 2012). Entropy balancing assigns control units unique weights such that, in aggregate, they look similar to the treated ones. Here, we weight using past primary turnout (2012, 2016, 2020); predicted race; age; de-registration status as of the 2024 primary; party; and county. The event study plot in Figure A.9.7 indicates that, net of entropy balancing, the parallel trends assumption is valid. However, using pre-treatment outcomes of interest in a pre-processing strategy raises some concerns about reversion to the mean (Daw and Hatfield, 2018); as such, we interpret these results more cautiously than the other DID results.

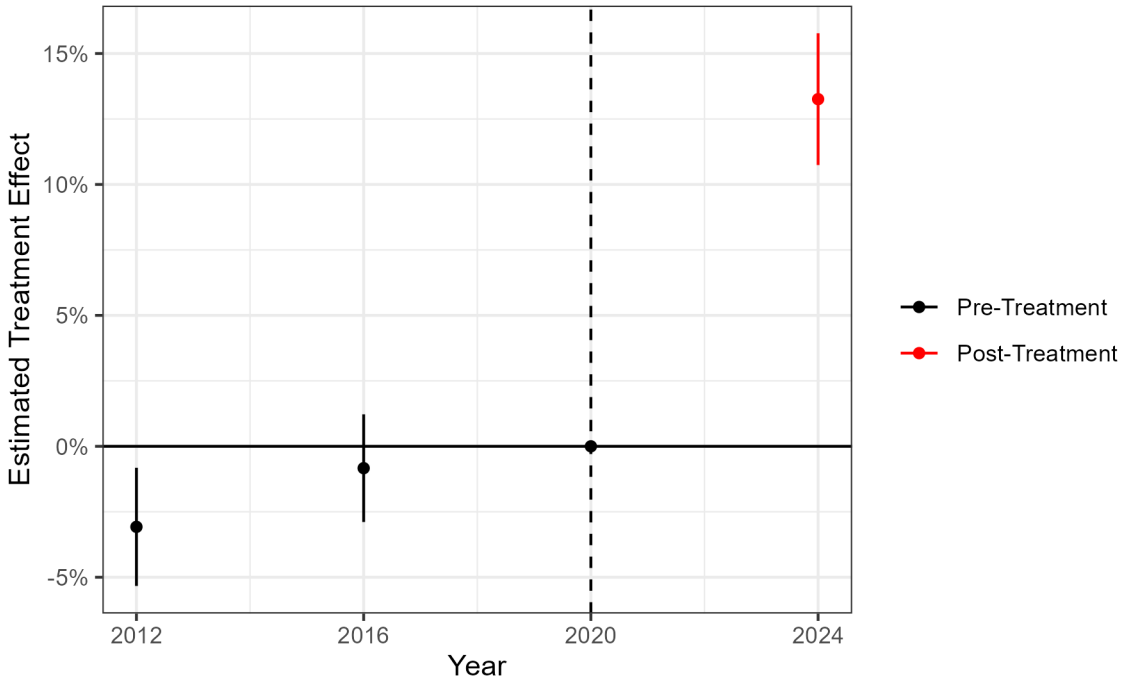
Figure A.9.7: Event Study, Application Rejection, 2024 Primary Election Turnout (Using Entropy Balancing)



All years measured relative to 2020 (final pre-treatment primary election). Voter, year-county, year-age, and year-party fixed effects included. Standard errors clustered by voter. Non-rejected voters entropy balanced using the following characteristics: predicted race; age; deregistration status; party; primary turnout (2012, 2016, 2020); county.

2024 Election: Voting in Person

Figure A.9.8: Event Study, Application Rejection, 2024 General Election In-Person Voting (Conditional on Voting)



All years measured relative to 2020 (final pre-treatment primary election). Voter, year-county, year-age, and year-party fixed effects included. Standard errors clustered by voter.

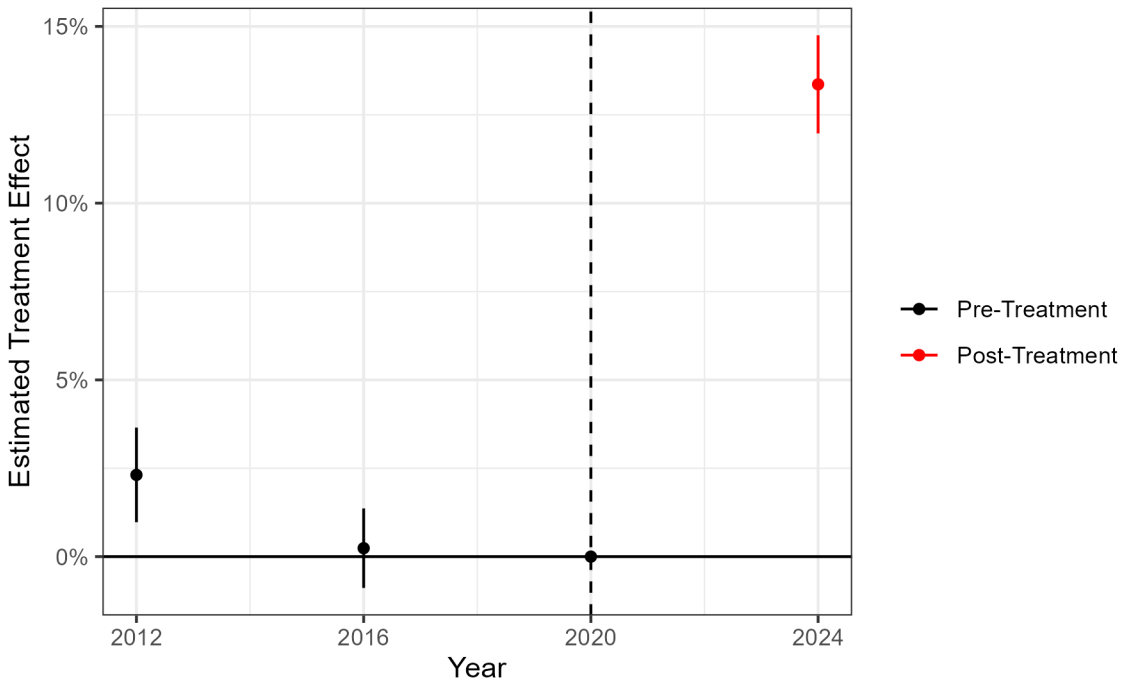
References

Billings, Stephen B., Noah Braun, Daniel B. Jones and Ying Shi. 2024. "Disparate Racial Impacts of Shelby County v. Holder on Voter Turnout." *Journal of Public Economics* 230:105047.

Callaway, Brantly and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics* 225(2):200–230.

Coppock, Alexander and Donald P. Green. 2016. "Is Voting Habit Forming? New Evidence from Experiments and Regression Discontinuities." *American Journal of Political Science* 60(4):1044–1062.

Figure A.9.9: Event Study, Ballot Rejection, 2024 General Election In-Person Voting (Conditional on Voting)



All years measured relative to 2020 (final pre-treatment primary election). Voter, year-county, year-age, and year-party fixed effects included. Standard errors clustered by voter.

Daw, Jamie R. and Laura A. Hatfield. 2018. "Matching and Regression to the Mean in Difference-in-Differences Analysis." *Health Services Research* 53(6):4138–4156.

Dowling, Conor M., Michael G. Miller and Kevin T. Morris. 2025. "Crossover Voting Rates in Partisan and Nonpartisan Elections: Evidence From Cast Vote Records." *Political Research Quarterly* 78(2):720–737.

Durst, Noah J. 2018. "Racial Gerrymandering of Municipal Borders: Direct Democracy, Participatory Democracy, and Voting Rights in the United States." *Annals of the American Association of Geographers* 108(4):938–954.

Fraga, Bernard L and Michael G Miller. 2022. "Who Do Voter ID Laws Keep from Voting?" *The Journal of Politics* 84(2):1091–1105.

Green, Donald P. and Ron Shachar. 2000. "Habit Formation and Political Behaviour: Evidence of Consuetude in Voter Turnout." *British Journal of Political Science* 30(4):561–573.

Hainmueller, Jens. 2012. "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies." *Political Analysis* 20(1):25–46.

Imai, Kosuke and Kabir Khanna. 2016. "Improving Ecological Inference by Predicting Individual Ethnicity from Voter Registration Records." *Political Analysis* 24(2):263–272.

Kim, Seo-young Silvia and Bernard Fraga. 2022. "When Do Voter Files Accurately Measure Turnout? How Transitory Voter File Snapshots Impact Research and Representation." *Working paper* .

Komisarchik, Mayya and Ariel White. 2025. "Throwing Away the Umbrella: Minority Voting after the Supreme Court's Shelby Decision." *Quarterly Journal of Political Science* 20(2):269–305.

McCartan, Cory, Jacob Goldin, Daniel E. Ho and Kosuke Imai. 2023. "Estimating Racial Disparities When Race Is Not Observed.".

Morris, Kevin and Michael G. Miller. 2025. "Did Shelby County v. Holder Increase the Racial Turnout Gap?" *The Journal of Politics* .

Roth, Jonathan, Pedro H. C. Sant'Anna, Alyssa Bilinski and John Poe. 2023. "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature." *Journal of Econometrics* 235(2):2218–2244.

Sekhon, Jasjeet S. 2011. "Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R." *Journal of Statistical Software* 42(7):1–52.