NBER WORKING PAPER SERIES

SELECTIVE TURNOUT, VOTING POLICY, AND PARTISAN BIAS: EVIDENCE FROM MULTI-LEVEL DATA

Steven T. Berry
Christian Cox
Philip Haile

Working Paper 34149
http://www.nber.org/papers/w34149

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2025

This work was begun when Christian Cox was a postdoctoral associate at the Jackson School of Global Affairs at Yale University. We have benefited from the comments and suggestions of participants in many seminars and conferences. The authors have no relevant funding or financial relationships to disclose. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Selective Turnout, Voting Policy, and Partisan Bias: Evidence from Multi-Level Data
Steven T. Berry, Christian Cox, and Philip Haile
NBER Working Paper No. 34149
August 2025
JEL No. H10, L0, P0

ABSTRACT

We study voting in general elections for the U.S. House of Representatives. Our data set includes demographics and turnout of all registered voters for the years 2016–2020, as well as vote shares at the precinct and contest level. We estimate a Downsian voting model incorporating rich observed and unobserved heterogeneity at the voter and contest level. We find that voters with high perceived voting costs tend to favor Democrats, as do marginal voters in most districts. Variation in state voting policies accounts for a modest share of overall estimated voting costs but is sufficient to determine the majority party in some years. We also find that many states' district maps favor one party in converting votes to seats. On net, these biases favor Republicans. For example, we estimate that winning 50% of votes in every state would give Republicans a nine percentage point seat advantage in the House.

Steven T. Berry
Yale University
Department of Economics
and NBER
steven.berry@yale.edu

Philip Haile
Yale University
Department of Economics
and NBER
philip.haile@yale.edu

Christian Cox
University of Arizona
christiancox@arizona.edu

1 Introduction

We combine a model of individual voting decisions with data on general elections for the U.S. House of Representatives to study how election outcomes are affected by voter heterogeneity, selective turnout, state voting policies, and structural biases in the design of congressional districts. Voting plays an essential role in representative democracies. But because voting is typically voluntary, election outcomes may not reflect the preferences of the electorate; preferences of particular sociodemographic groups may be over- or under-represented; policies making it easier/harder to vote can alter not just turnout but election outcomes; and the effects of such policies depend on which voters (and how many of them) are on the relevant turnout margins. Likewise, the partisan effects of district design and gerrymandering depend not just on the mix of preferences in each district, but also on the associated heterogeneity across these potential voters in the likelihood of turning out.

To quantify the preferences and selective turnout that drive voting outcomes, we require a model that allows rich voter heterogeneity while capturing both turnout and candidate choice in one coherent framework. We posit a Downsian model of voting in two-party elections. Each registered voter has a cost of voting and preferences over which candidate wins. Voting costs can be negative for some voters and can be scaled by idiosyncratic beliefs about vote efficacy, idiosyncratic intensity of preference between candidates, or idiosyncratic taste for expressing one's preference. This is a discrete choice model, and the roles of voter heterogeneity and multidimensional selection here have connections to the roles of consumer heterogeneity and flexible "substitution patterns" in discrete demand models (e.g., Berry, Levinsohn, and Pakes (1995, 2004), Berry and Haile (2021)). However, both the form of the voting model and the nature of the data available to us require some new (nonparametric) identification results and a different estimation approach.

Our data set comes from several sources and covers general elections for the U.S. House in 2016, 2018, and 2020. For every registered voter in the country, we observe home location, turnout for each election, and a rich set of sociodemographic measures ("demographics") including party affiliation. At the "contest" (district×year) level, we observe total turnout and vote shares. We also observe vote shares at the precinct level for a large fraction of precincts. At the state level, we observe voting policies and up-ballot factors such as whether there is also a governor's race or Senate race.

Our empirical approach links the data to our voting model. In our empirical specification, each voter's preferences depend on her observed demographics and two fixed effects: one representing the mean relative attractiveness of the two candidates (signed to represent the "benefit" of electing the Republican),

and one representing the mean perceived effective voting cost ("cost"). These fixed effects are analogs of the product-specific mean utilities in typical discrete choice demand models (e.g., Berry, Levinsohn, and Pakes (1995, 2004)). In addition, each voter has a pair of jointly normal shocks—one to benefit and one to cost. In each contest, voter types are thus two-dimensional conditional on demographics.

We first estimate the contest-level fixed effects and parameters governing individual preferences using a likelihood-based approach applied to individual- and precinct-level data, along with the district-level vote shares. These estimates alone suffice for many of our questions. Our analysis of state voting policies requires that we estimate how changes in these policies alter the fixed effects. For this purpose, we estimate reduced forms for the fixed effects. Although this approach has limitations, here it allows counterfactual predictions to capture both direct effects of policy on voting costs and indirect effects on benefits and costs. These indirect effects may involve mediating factors—candidate gender, charisma, policy positions, campaign spending, etc.—that respond to voting policy changes. Moreover, our approach does not require us to observe all mediating factors or have sufficient structure and sources of variation to identify a more complex model that includes their equilibrium determination. When we address nonparametric identification, our results will cover both this approach and the more typical alternative in which one does observe, model, and instrument for endogenous factors. [1]

Our estimates reveal substantial variation in voter preferences and perceived voting costs. Individuals with high costs tend to prefer Democrats—an association reflecting both observables and unobservables at the individual level. As a result, election outcomes overall are heavily skewed toward Republican candidates relative to the majority preferences of each district. We find that marginal voters in most districts also have a strong tendency to prefer the Democrat. This supports the conventional wisdom that, in our data period, more restrictive voting policies tend to benefit Republicans. This remains the case even after accounting for countervailing responses that are implied by our contest-level reduced forms. Groups most affected by changes in voting policies include Blacks, Hispanics, younger voters, low-education voters, and recently registered voters.

The magnitudes of these policy impacts are potentially significant but nuanced. For example, if states with more permissive policies shifted to more

[1] Examples of this more traditional approach include Gerber (1998), Gillen, Moon, Montero, and Shum (2019), Gordon and Hartmann (2013), Cox (2024), Iaryczower, Montero, and Kim (2022), and Longuet-Marx (2025). These studies limit attention to a small number of contest-level endogenous variables—e.g., the level of campaign spending and/or candidates' positions in a one-dimensional policy space—that can be measured and instrumented.

restrictive policies (matching the 90th percentile of each policy measure), we predict a total loss of Democratic votes sufficient to reverse the Democratic majorities in the House in 2018 and 2020. However, the potential gains to Democrats from more permissive policies (shifting to 10th percentile policies in more restrictive states) would yield only modest gains in the number of Democratic seats won. This asymmetry reflects a finding that states in which partisan outcomes are most sensitive to voting policies tend to already have more permissive policies.

We also use our estimates to examine the "partisan bias" of each state's congressional district maps. Across states we find a mix of biases, some favoring each party. However, biases favoring Republicans dominate, both in number and magnitude. We estimate that if, in each state, the relative appeal of Republican and Democratic candidates were adjusted in each district to achieve a 50-50 balance in the statewide vote share for the two parties, Republicans would win a majority of House seats in each of the three ye ars we study, with nearly a 55-45 advantage on average.

Our work connects to several literatures in economics and political science. Most closely related is other work using turnout and/or vote shares to estimate voting models and construct counterfactual voting outcomes.[2] One approach is to model voters' choice of candidate conditional on voting (e.g., Alvarez and Nagler (1998), Glasgow (2001), Jessee (2010), Merlo and de Paula (2017), Huang and He (2021), Longuet-Marx (2025)). Other work models turnout and voting jointly as a standard random utility discrete choice problem: a potential voter abstains if no candidate is sufficiently appealing, otherwise choosing the candidate she likes best. Examples include Gillen, Moon, Montero, and Shum (2019), Gordon and Hartmann (2013), Iaryczower, Montero, and Kim (2022), and Cox (2024).[3]

Our "calculus of voting" approach differs in specifying utilities over election outcomes and the perceived cost of voting. Abstention (for voters with positive voting costs) then reflects an insufficiently strong preference between candidates rather than a dislike of the available options. An important example in this line is Kawai, Toyama, and Watanabe (2021) ("KTW"), who studied preference aggregation in the 2004 U.S. presidential

[2]As typical in this literature, we do not model registration. However, given reliable data on unregistered eligibles, our approach could be extended to account for registrations as an additional cost of turning out for those not already registered.

[3]A variation considered by Ujhelyi, Chatterjee, and Szabó (2021) accounts for a participating voter's option (in their application) to select "none of the above" candidates.

election.[4] We build on their model, generalizing by incorporating contest-level observables; allowing contest-level factors (observed or not) to respond to counterfactual interventions; and allowing for (correlated) two-dimensional voter-level unobservables—thus, more flexible selection into voting. Our results indicate that this multidimensional selection is important.[5] Despite these generalizations, we show that the parametric assumptions exploited by KTW are not necessary for identification. Our nonparametric identification results build on those of Berry and Haile (2024) for differentiated products demand, although our model and observables differ in ways that necessitate new results.[6]

Our multi-level data structure is similar to that used by Ainsworth (2020) to study gerrymandering in North Carolina. He noted the extreme complexity of the likelihood for observed vote shares in such a setting and proposed a normal approximation that we adopt in the first step of our estimation. His data covers all questions on each ballot, and his model accounts for all of them. This is potentially important (e.g., Knight (2017)) but leads to some compromises in the links between turnout and the preferences driving decisions in the voting booth. By focusing exclusively on voting for the U.S. House, we make a different set of modeling compromises, relying on a combination of fixed effects, indicators for other races on the ballot, and contest-level unobservables to capture the effects of other ballot questions.[7]

Substantively, our analysis connects to prior work studying the effects of voting policies, typically using the empirical methods of program evaluation to estimate certain causal effects. This literature alone is extensive, and an excellent survey is offered by Cantoni, Pons, and Schäfer (2024). We take a complementary empirical approach, exploiting a unified model of individual turnout and voting that allows us to capture multiple forces and outcomes within one coherent framework; to study multiple policies at once; and to simulate counterfactual outcomes at local, national, and group levels. This

[4]Alternative models of turnout and/or candidate choice are estimated by, e.g., Degan and Merlo (2011), Coate and Conlin (2004), Coate, Conlin, and Moro (2008), and Merlo and de Paula (2017).

[5]Multidimensional selection has been shown to be important in other contexts, including insurance markets (e.g., Einav, Finkelstein, and Cullen (2010), Bundorf, Levin, and Mahoney (2012)).

[6]Merlo and de Paula (2017) provided conditions (using precinct vote shares) for nonparametric identification in an n-party ideal point model in which voting is compulsory, party positions are exogenous, and voters within a precinct differ only in their positions in ideological space.

[7]Our substantive focus is also complementary although we overlap in predicting U.S. House seat shares for North Carolina when its state vote is split 50-50. His prediction of 74% is very similar to our prediction of 72%.

approach exploits both within- and cross-state policy variation, controlling for rich observed and unobserved heterogeneity at the individual level, as well as a rich set of observables at the district, contest, and state levels. Our broad findings regarding these policies are also complementary, demonstrating that restrictive voting policies discourage participation by minorities and others more likely to favor Democrats, but can have somewhat modest impacts on election outcomes, depending on the particular policy changes and states considered.[8] As noted already, however, we identify a potentially important asymmetry between parties, resulting from the cross-state heterogeneity in the number and preferences of voters on the relevant margins in each state.

There is also a large literature evaluating states' congressional district designs by measures of "partisan symmetry" in the implied mappings from votes to seats.[9] Such measures—including the "partisan bias" measure we study—typically require one to predict the distribution of district-level vote shares at particular hypothetical state-level vote shares. We accomplish this by adjusting the relative attractiveness of the parties' candidates in each state, letting the model reveal how turnout and vote shares in each district would respond. This contrasts with typical approaches that directly specify district-level vote share regression functions, creating counterfactual vote shares by adding appropriate constants (the so-called "uniform partisan swing"). Although there is no single correct way to create hypothetical statewide vote shares, ours allows clear interpretation and accounts for the multi-dimensional selection underlying actual outcomes.

2 Data

2.1 Individual-Level Data

Our individual-level data are drawn from the voter data set of the commercial data provider L2. This data set covers all registered voters—roughly 170 million in each of the three election years we study. Key elements of the L2 voter data reflect information taken directly from state voter files, such as each registered voter's name, address, party registration (if any), and turnout (voted or not) for each election. The L2 data set also includes an extensive set of demographic measures. These reflect information originating from official state voter files, commercial sources, and census data.

[8] See, e.g., Fraga and Miller (2021), Thompson, Wu, Yoder, and Hall (2020), and Cantoni and Pons (2021a).

[9] See, e.g., Tufte (1973), Grofman and King (2007), King (1990) King and Browning (1987), Gelman and King (1994), Katz, King, and Rosenblatt (2020a), Coate and Knight (2007), and Ainsworth (2020).

Table 1 summarizes the individual-level variables we use in our analysis. Across all contested races,[10] there are more than 500 million individual-level observations on demographics and turnout. Most of the demographic variables are binary.[11] For others, we scale the variable to have a range of approximately 0 to 1 in order to ease interpretation of parameter estimates later. In addition to demographic measures, we construct an indicator for newly registered voters—those who registered within the same calendar year.

Voters may be influenced by their neighbors, as through peer effects or perceived competitiveness (Shachar and Nalebuff (1999)). We use our data on party affiliation to construct measures of the partisanship of each individual's neighbors within two concentric "rings"—the individual's own census tract and the next 10 closest census tracts. We construct three measures: (a) $R/(R+D)$, (b) $\%I$, and (c) $|R-D|/(R+D)$, where R and D represent the number of registered Republicans and Democrats, respectively, and $\%I$ is the percentage of registered independents. Here, (a) is a measure of the partisan skew among partisans; (b) measures the share of nonpartisans; and (c) is a measure of partisan homogeneity. These measures may also serve—e.g., via residential sorting—as additional proxies for relevant heterogeneity across voters.

In Table 1 we saw that a little more than 70% of registered voters turn out to vote. Table 2 provides summary statistics for the selected sample of these actual voters. Actual voters tend to be older, wealthier, whiter, and more Republican. Those who vote are less likely to be recent registrants.

Although the L2 data represent the current state of the art and are widely used by practitioners, there are important caveats. First, state voter files differ across states, both in the information collected and in completeness (e.g., due to variations in privacy regulations and update frequencies).[12] Second, these differences may affect L2's success in linking state voter files to commercial data sources to obtain additional demographics.

Third, some measures are imputed for some or even all registered voters, using proprietary models and algorithms (performed by the L2's data consulting firm Haystaq DNA). For example, only 31 states have partisan registration,

[10]Contests that do not have a Republican or Democrat running as non-write-ins (which the FEC reports in their elections results tables), are considered uncontested. There were 64 uncontested races in 2016, 41 in 2018, and 27 in 2020. Republicans won 29, 3, and 7 of these, respectively. We exclude uncontested races when estimating the model (we discuss their treatment in counterfactuals below). Summary statistics are similar for the full sample.

[11]Income is binned, with eleven bins for incomes up to $250,000. We use the midpoint of each bin as the associated income level. For the top-coded (12th) bin, we use an income level of $275,000.

[12]See, e.g., Cao, Kim, and Alvarez (2022) and Kim and Fraga (2022). A summary of information collected by each state is available at `https://www.eac.gov/sites/default/files/voters/Available_Voter_File_Information.pdf`.

Table 1: Summary Statistics: Registered Voters
2016-2018-2020

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**
Votes	0.700	0.458	0	1
Age/100	0.505	0.184	0.180	1.010
Income/$250	0.364	0.235	0	2.010
Male	0.468	0.499	0	1
Education/16	0.798	0.074	0	1.125
White	0.649	0.477	0	1
Hispanic	0.111	0.314	0	1
Black	0.098	0.298	0	1
Family size/4	0.514	0.239	0.250	2.500
Recent registration	0.08	0.272	0	1
Republican	0.323	0.468	0	1
Democrat	0.387	0.487	0	1
Urban	0.313	0.464	0	1
Suburban	0.435	0.496	0	1
Nearby tract Republican	0.307	0.151	0	0.967
Nearby tract Democrat	0.395	0.174	0	1
Own tract Republican	0.323	0.169	0	1
Own tract Democrat	0.387	0.186	0	1
Near tract $R/(R+D)$	0.442	0.203	0	1
Near independent share	0.298	0.129	0	1
Near tract $\lvert R-D\rvert/(R+D)$	0.340	0.250	0	1
Own tract $R/(R+D)$	0.457	0.219	0	1
Own tract independent share	0.290	0.132	0	1
Own tract $\lvert R-D\rvert/(R+D)$	0.363	0.258	0	1
Observations		513,171,413		

The sample covers registered voters in district-years with a contested House race.

and individual party affiliation is imputed for registered voters in the remaining states.[13] Such imputations may be especially troubling when a primary goal is to assess the causal effects of particular demographics. However, we employ demographics primarily as proxies for underlying voter heterogeneity, adding flexibility to the model and limiting the roles played by unobserved individual heterogeneity. Even imputed proxies can serve this purpose, and our approach is to use the best available micro data while being cognizant

[13]Our results are robust to treating the party registration measures as different variables in the registration and non-registration states.

Table 2: Summary Statistics: Actual Voters
2016-2018-2020

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**
Votes	1	0	1	1
Age/100	0.529	0.178	0.180	1.010
Income/$250	0.381	0.243	0	2.010
Male	0.459	0.498	0	1
Education/16	0.803	0.075	0	1.125
White	0.680	0.467	0	1
Hispanic	0.095	0.293	0	1
Black	0.086	0.280	0	1
Family size/4	0.528	0.233	0.250	2.500
Recent reg.	0.070	0.255	0	1
Republican	0.371	0.483	0	1
Democrat	0.396	0.489	0	1
Urban	0.300	0.458	0	1
Suburban	0.447	0.497	0	1
Nearby tract Republican	0.314	0.149	0	0.912
Nearby tract Democrat	0.387	0.169	0	1
Own tract Republican	0.334	0.166	0	1
Own tract Democrat	0.377	0.178	0	1
Near tract $R/(R+D)$	0.452	0.199	0	1
Near independent share	0.299	0.126	0	0.957
Near tract $\lvert R-D\rvert/(R+D)$	0.330	0.242	0	1
Own tract $R/(R+D)$	0.471	0.213	0	1
Own tract independent share	0.289	0.127	0	1
Own tract $\lvert R-D\rvert/(R+D)$	0.351	0.250	0	1
Observations		359,183,410		

The sample covers all actual voters in district-years with a contested House race.

of the potential limitations. A secondary role of demographics is to describe some of the heterogeneity in responses to counterfactuals. Imputation implies some qualifications for that purpose. For example, when we find that restrictive voting policies disproportionately affect Black registered voters, a precise statement will be that the disproportionate effect is on those who are either Black or viewed as likely to be Black according to the L2 imputations.

2.2 Precinct-Level Data

Although the L2 micro data provide individual-level turnout, individual votes are, of course, not observed. Thus, we supplement the L2 data with vote shares at the precinct and contest levels, primarily sourced from the Harvard Election Database.[14] We also obtained precinct shape files from the Voting and Election Science Team and Redistricting Data Hub. Mapping individual addresses from the L2 data to the corresponding precincts allows us to link the individuals turning out to the corresponding precinct-level vote shares.[15] We are able to match registered voters to precincts for most precincts in most states. However, some voters cannot be mapped to precincts due to missing maps or geographic misidentification. For each contest, we put all such voters into a separate "super-precinct," treated as a standard precinct for estimation. For districts where digitized precinct maps are not available at all, the entire district is treated as one super-precinct.

2.3 Contest-Level Data

Table 3 summarizes our data at the contest level. We use the L2 data to construct district demographic averages (and the measures of partisanship discussed above) for each year. We obtained state-level measures of policies affecting voting costs from the Cost of Voting Index (COVI) database (Li, Pomante, and Schraufnagel (2018), Pomante (2025)). This source documents a set of voting laws and related administrative policies across states. Because we model voting by registered voters, we focus on the voting cost measures, leaving aside COVI measures related to voter registration. We use voting cost measures for two broad categories: voting "inconvenience" (e.g., polling station food/drink, paid postage, wait times, number of polling locations), and voter identification requirements.[16]

We collected two advertising cost measures commonly used as instruments for candidate spending (e.g., Snyder and Strömberg (2010), Gordon and Hartmann (2013), Spenkuch and Toniatti (2018), Wang, Lewis, and Schweidel

[14]Baltz, Agadjanian, Chin, Curiel, DeLuca, Dunham, Miranda, Phillips, Uhlman, Wimpy et al. (2022) discuss the data quality issues with precinct-level data.

[15]There are on average, roughly 1,000 registered voters per mapped precinct. Congressional districts in our sample period were targeted to have approximately 750,000 residents.

[16]COVI measures are collected for each presidential election year. We use the 2016 values for the 2018 midterm elections. Because the set of COVI measures collected grows over time, we use the individual components to construct consistently defined category-level indices the election years in our sample.

Table 3: Contest-Level Variables

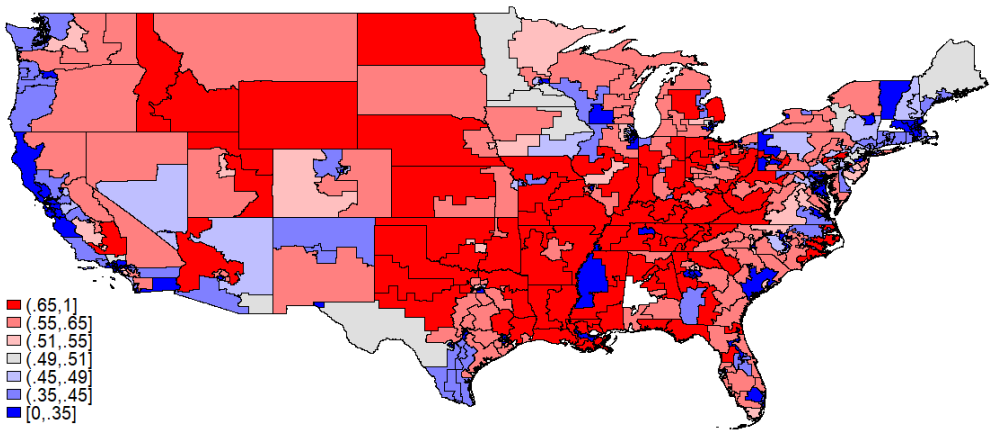
Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**
District Average Demographics (by Year)				
Republican	0.323	0.129	0.039	0.726
Democrat	0.389	0.147	0.090	0.837
Independent	0.289	0.115	0.047	0.653
Male	0.468	0.013	0.402	0.511
Age/100	0.504	0.025	0.431	0.601
White	0.641	0.185	0.100	0.932
Black	0.096	0.138	0.000	0.669
Hispanic	0.121	0.146	0.009	0.793
Education/16	0.796	0.038	0.675	0.936
Income/$250k	0.360	0.088	0.134	0.68
Urban	0.319	0.256	0.000	1.000
Suburban	0.426	0.277	0.000	1.000
Family size/4	0.514	0.032	0.392	0.630
Recent registration	0.081	0.027	0.021	0.219
$R/(R+D)$	0.457	0.171	0.047	0.849
$\|R-D\|/(R+D)$	0.279	0.217	0.001	0.905
State Voting Policy Indices and Up-Ballot Factors (by Year))				
Voting inconveniences	5.265	2.247	0	10
Voter ID laws	1.262	1.401	0	4
Governor's race has open seat	0.159	0.365	0	1
State has governor's race	0.354	0.478	0	1
Senate race has open seat	0.090	0.286	0	1
State has Senate race	0.666	0.472	0	1
Other Contest-Level Variables				
Local ad price (CPP)	888	1,049	50	3,860
Media market overlap	83	21	22	100
Registered	437,486	69,307	196,057	726,298
Vote	306,209	70,912	111,377	595,727
Votes D	154,560	58,696	32,405	396,274
Votes R	151,649	63,697	5,240	335,909
Number of Competitive Contests			1,173	

The sample includes all district-years with a contested House race.

(2018)). One is a measure of local media advertising cost.[17] The second is the overlap between the congressional district and the media market (DMA), obtained from the Daily Kos. State-level measures of up-ballot factors for each election year were obtained from Dave Leip's U.S. Election Atlas. These measures are binary indicators for whether there is a governor's race or senate race on the same ballot, and whether these involve open seats.

Contest-level vote shares were taken from the Harvard Election Database.[18] The map in Figure 1 shows average outcomes (across the three years) in house elections, with red indicating Republican wins.

Figure 1: Republican Vote Shares
3 year average



3 Model

3.1 Random Utilities

We posit a "calculus of voting" model (Downs (1957), Riker and Ordeshook (1968)), building on that in KTW. For each contest dt there are two candidates $j \in \{1, 2\}$.[19] For clarity, all random variables include dt among their subscript indices. There is a population of registered voters ("voters") for each contest. Voters have heterogeneous preferences over contest outcomes—i.e., over which candidate wins—and may have beliefs about the efficacy of their vote. Voters

[17] We use CPP for the local TV late news time slot all-adults demographic, for the October one year prior to the election. These data were obtained from Guideline Solutions, Inc.

[18] For districts with non-standard plurality rule voting (majority rule in GA, MS, and LA, and ranked-choice voting in ME since 2018), we use the final run-off vote tally.

[19] In our empirical work we adopt the convention that candidate 1 is the Republican.

also face heterogeneous costs of voting and this cost may sometimes be negative. Each voter chooses one of three mutually exclusive options: voting for candidate 1, voting for candidate 2, and not voting.

Let $V_{ijdt} \in \mathbb{R}$ denote the utility voter i would obtain from election of candidate $j \in \{1, 2\}$ in contest dt. Let $K_{idt} \in \mathbb{R}$ denote i's cost of voting. This cost of voting accounts for the utility voter i obtains from voting itself and so it need not be positive.[20] If we suppose that $P_{idt} > 0$ represents voter i's perceived likelihood of pivotality in contest dt, then voter i's expected utility from voting for candidate 1 takes the form[21]

$$\tilde{U}_{i1dt} = P_{idt} \left(V_{i1dt} - V_{i2dt} \right) - K_{idt},$$

Similarly, the expected utility from voting for candidate 2 is

$$\tilde{U}_{i2dt} = -P_{idt} \left(V_{i1dt} - V_{i2dt} \right) - K_{idt}.$$

Without loss, we normalize the location of voter i's utilities by setting the expected utility $\tilde{U}_{i0dt}$ from not voting (labeled option 0) to zero. Each voter selects the option $j \in \{0, 1, 2\}$ with the highest expected utility.

We have followed KTW in describing P_{idt} as a voter's perceived pivotality, and one plausible model of voting is that many voters act as if their pivotality were larger than its true value. But this interpretation is unnecessarily narrow. We do not impose rational expectations or even require $P_{idt} \in [0, 1]$. Broader interpretations include treating P_{idt} as a shifter of preference intensity, a measure of i's political engagement, a shifter of perceived civic duty, or a shifter of the utility i gains from expressing her preference for her favored candidate.

Some of these possibilities become clearer if we define

$$B_{idt} = V_{i1dt} - V_{i2dt} \tag{1}$$

and

$$C_{idt} = \frac{K_{idt}}{P_{idt}}. \tag{2}$$

[20] As in Riker and Ordeshook (1968), a voter may obtain satisfaction from "compliance with the ethic of voting, ... affirming allegiance to the political system ... affirming a partisan preference ... deciding, going to the polls, etc. ... affirming one's efficacy in the political system ... [or] other satisfactions that do not occur to us at the moment."

[21] This representation requires a technical assumption when P_{idt} is interpreted as i's perceived pivotality. It is sufficient to assume that voters place probability zero on ties, so that voting when pivotal implies that one's chosen candidate wins. Alternatively, one may assume each voter treats as equally likely the following outcomes, absent her own vote: (i) 1 and 2 will tie, leading to a coin toss; (ii) 1 will win by one vote; (iii) 2 will win by one vote. See, e.g., KTW, Riker and Ordeshook (1968), and Myerson and Weber (1982).

Then the expected utilities

$$\begin{aligned} U_{i1dt} &= B_{idt} - C_{idt} && (3) \\ U_{i2dt} &= -B_{idt} - C_{idt} && (4) \\ U_{i0dt} &= 0 && (5) \end{aligned}$$

provide a representation of voter preferences equivalent to that implied by the utilities $\left(\tilde{U}_{i0dt}, \tilde{U}_{i1dt}, \tilde{U}_{i2dt}\right)$. Here B_{idt} (respectively, $-B_{idt}$) can be interpreted as the "benefit" of voting for 1 (respectively, 2), while C_{idt} is the effective "cost" of voting. From (2) it is clear that P_{idt} modulates the effective cost of voting C_{idt}. When $K_{idt} > 0$, the K_{idt}/P_{idt} determines the threshold level of preference intensity–i.e., magnitude of $|B_{idt}|$—necessary for i to vote. When $K_{idt} \leq 0$, voter i votes for her preferred candidate regardless of her preference intensity or perceived pivotality.

Although we refer to C_{idt} as a cost of voting, this is shorthand for the broader notion of a voter's perceived effective cost of voting. As the definition (2) makes clear, it will not be possible to distinguish between the roles of P_{idt} and those of K_{idt} without additional restrictions.[22] However, many important positive and normative questions can be addressed without decomposing the effective voting cost C_{idt} into the components K_{idt} and P_{idt}.[23]

3.2 Model Variables

We model B_{idt} and C_{idt} as functions of contest-level observables, voter-level observables, contest-level unobservables, and voter-level unobservables:

$$\begin{aligned} B_{idt} &= B\left(z_{idt}, x_{dt}, y_{dt}, \xi_{Bdt}, \epsilon_{iBdt}\right) && (6) \\ C_{idt} &= C\left(z_{idt}, x_{dt}, y_{dt}, \xi_{Cdt}, \epsilon_{iCdt}\right). && (7) \end{aligned}$$

Here, individual-level demographics are denoted by z_{idt}. Unobserved shocks to voter-level benefit and cost are denoted by ϵ_{iBdt} and ϵ_{iCdt}. Thus, voter-level heterogeneity within a contest is represented by the vector $(z_{idt}, \epsilon_{iBdt}, \epsilon_{iCdt})$. At the contest level, exogenous observables are denoted by x_{dt}, with endogenous observables denoted by y_{dt}. Contest-level unobservables are represented by the two scalars ξ_{Bdt} and ξ_{Cdt}, one each for benefit and cost. The exogeneity of x_{dt} is defined by an assumption that it is mean independent of the contest-level unobservables (ξ_{Bdt}, ξ_{Cdt}).

[22] KTW provide one set of such restrictions.

[23] One important implication is that anything affecting B_{idt} could also affect C_{idt} through P_{idt}, ruling out certain exclusion restrictions.

The assumption that unobservables at the choice set (contest) level are represented by two scalars is restrictive but standard in discrete choice models when unobservables at the level of the choice set are acknowledged at all (see e.g., Berry, Levinsohn, and Pakes (1995), Berry and Haile (2021)). We allow for dependence and serial correlation of ξ_{Bdt} and ξ_{Cdt} within each district. We assume $(\epsilon_{iBdt}, \epsilon_{iCdt})$ are independent of $(z_{idt}, x_{dt}, y_{dt}, \xi_{Bdt}, \xi_{Cdt})$. This specifies that $(z_{idt}, x_{dt}, y_{dt}, \xi_{Bdt}, \xi_{Cdt})$ alter the distribution of (B_{idt}, C_{idt}) through their role as arguments of the functions B and C rather than through effects on the joint distribution of $(\epsilon_{iBdt}, \epsilon_{iCdt})$.[24] Although our model and identification results require no restriction on the dimension of ϵ_{iBdt} or ϵ_{iCdt}, in our empirical specification these will be two scalars.

3.3 Voting Choice Functions

Define $\xi_{dt} = (\xi_{Bdt}, \xi_{Cdt})$, and let F denote the joint distribution of $(\epsilon_{iBdt}, \epsilon_{iCdt})$. In contest dt, candidate 1 is chosen by voter i when

$$B_{idt} - C_{idt} > \max\{0, -B_{idt} - C_{idt}\},$$

whereas candidate 2 is chosen when

$$-B_{idt} - C_{idt} > \max\{0, B_{idt} - C_{idt}\}.$$

Thus, given a particular realization (z, x, y, ξ) of $(z_{idt}, x_{dt}, y_{dt}, \xi_{dt})$, choice probabilities for each option are given by the voting choice functions

$$\sigma_1(z,x,y,\xi) = \int 1\{B(z,x,y,\xi_B,\epsilon_B) > \max\{0, C(z,x,y,\xi_C,\epsilon_C)\}\} \, dF(\epsilon_B,\epsilon_C) \tag{8}$$

$$\sigma_2(z,x,y,\xi) = \int 1\{-B(z,x,y,\xi_B,\epsilon_B) > \max\{0, C(z,x,y,\xi_C,\epsilon_C)\}\} \, dF(\epsilon_B,\epsilon_C) \tag{9}$$

$$\sigma_0(z,x,y,\xi) = 1 - \sigma_1(z,x,y,\xi) - \sigma_2(z,x,y,\xi). \tag{10}$$

Figure 2 illustrates the determination of these choice probabilities, representing voters by points in the space of the random variables (B_{idt}, C_{idt}). Voters in the grey region do not vote; those in the pink region vote for candidate 1, and those in the blue region vote for candidate 2. The choice probabilities in (8)–(10) correspond to the probability measure on each region conditional on a given value of $(z_{idt}, x_{dt}, y_{dt}, \xi_{dt}) = (z, x, y, \xi)$.[25] Figure 3 shows the same

[24] If $(z_{idt}, x_{dt}, y_{dt}, \xi_{Bdt}, \xi_{Cdt})$ could freely alter the joint distribution of $(\epsilon_{iBdt}, \epsilon_{iCdt})$, there would be no need to include $(z_{idt}, x_{dt}, y_{dt}, \xi_{Bdt}, \xi_{Cdt})$ as arguments of B and C: one could set $(B_{idt}, C_{idt}) = (\epsilon_{iBdt}, \epsilon_{iCdt})$ without loss.

[25] Observe that if C_{idt} is set to a constant C_d for all voters in district d with the same observables, as in KTW, choices vary across voters (conditional on observables) only with the value of B_{idt}. In that case, one obtains an ordered choice problem, represented by a single horizontal line segment (at height C_d) in Figure 2.

information when we represent voters as points in the space of the expected utilities (U_{i1dt}, U_{i2dt}) defined by (3) and (4). Again, voters in the grey region do not vote; those in the pink region vote for candidate 1, and those in the blue region vote for candidate 2. The probability masses on each region are, by construction, identical to those in Figure 2. Indeed, Figure 3 is simply a rotation (225 degrees clockwise) of Figure 2.

The graphical representation of choice outcomes in Figure 3 is identical to that for a standard random utility discrete choice model (e.g., Thompson (1989)). Nonetheless, there are some important differences. First, although a voter can choose not to vote for either candidate, one of the two will still be elected. Thus, even the decision of *whether* to vote depends in part on the difference between the utilities associated with the two contest outcomes, not on the level of these utilities relative to a normalized outside option. Second, the two utilities U_{i1dt} and U_{i2dt} are both formed as linear combinations of B_{idt} and C_{idt}. Thus, the model necessarily lacks independence and exclusivity conditions built into typical random utility discrete choice models: here, U_{i1dt} and U_{i2dt} are each affected by all factors—including the latent $(\xi_{Bdt}, \xi_{Cdt}, \epsilon_{iBdt}, \epsilon_{iCdt})$—that affect B_{idt} and C_{idt}. We will see some implications of these differences as we discuss estimation and identification.

3.4 Empirical Specification

We use a more restrictive specification for our empirical work, imposing both an index structure linking to the identification results of Berry and Haile (2024) and an assumption that the individual-level observables and taste shocks enter linearly, as in standard random utility models. We specify

$$B_{idt} = z_{idt}\alpha_B + \psi_B(x_{dt}, y_{dt}) + \xi_{Bdt} + \epsilon_{iBdt} \tag{11}$$

$$C_{idt} = z_{idt}\alpha_C + \psi_C(x_{dt}, y_{dt}) + \xi_{Cdt} + \epsilon_{iCdt}. \tag{12}$$

We assume $(\epsilon_{iBdt}, \epsilon_{iCdt})$ to be multivariate normal, with mean zero and (normalized) covariance matrix denoted by Σ.

Observe that in (11) and (12) the terms

$$\delta_{Bdt} \equiv \psi_B(x_{dt}, y_{dt}) + \xi_{Bdt}, \tag{13}$$

and

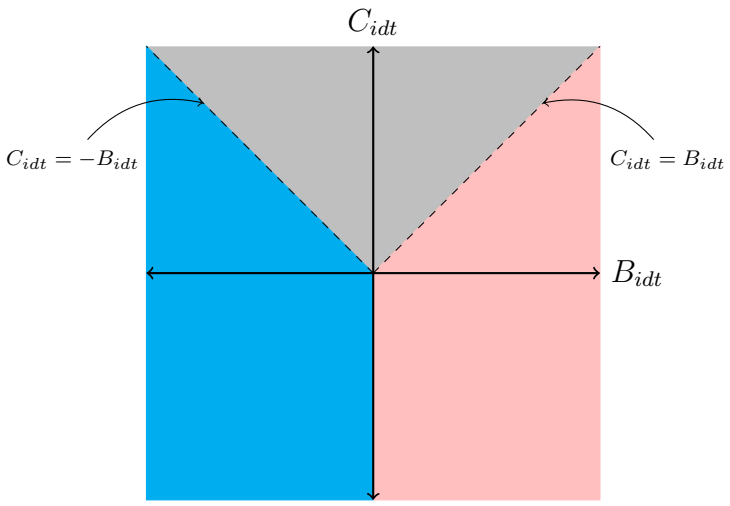
$$\delta_{Cdt} \equiv \psi_C(x_{dt}, y_{dt}) + \xi_{Cdt} \tag{14}$$

form two contest-level "fixed effects" that capture all observed and unobserved heterogeneity across contests. Thus, we have

$$B_{idt} = z_{idt}\alpha_B + \delta_{Bdt} + \epsilon_{iBdt} \tag{15}$$

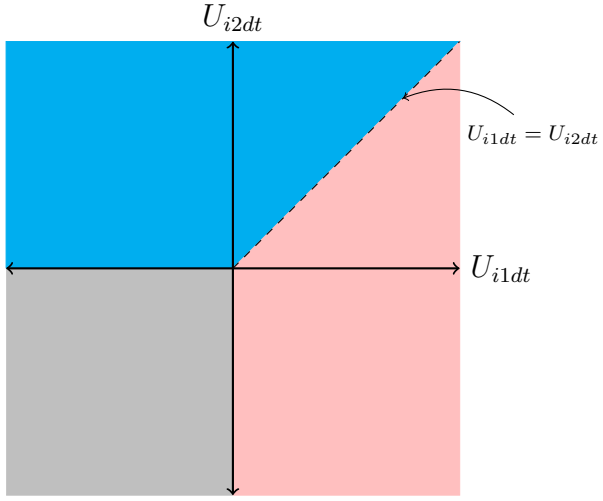
$$C_{idt} = z_{idt}\alpha_C + \delta_{Cdt} + \epsilon_{iCdt}. \tag{16}$$

Figure 2



The origin is the point $(0, 0)$. Voters in the grey region do not vote; those in the pink region vote for candidate 1; and those in the blue region vote for candidate 2.

Figure 3



$U_{i1dt} = B_{idt} - C_{idt}$ and $U_{i2dt} = -B_{idt} - C_{idt}$. The origin is the point $(0, 0)$. Voters in the grey region do not vote; those in the pink region vote for candidate 1; and those in the blue region vote for candidate 2.

Defining

$$\theta_1 \equiv (\alpha_B, \alpha_C, \Sigma),$$

we can (abusing notation slightly) also write the choice probability functions in (8)–(10) more simply as

$$\sigma_j\left(z_{idt}; \theta_1, \delta\right), \quad j = 0, 1, 2.$$

4 Contest-Level Reduced Forms

For most of the questions we explore, it suffices to have estimates only of the parameters $(\alpha_B, \alpha_C, \Sigma)$ and fixed effects $(\delta_{Bdt}, \delta_{Cdt})$ appearing in (15)–(16). However, our analysis of state voting policies requires that we quantify their effects on the contest-level fixed effects. A natural approach, then, is to also estimate the structural equations (13) and (14), along with any additional equations necessary to characterize the responses of endogenous contest-level factors y_{dt} to counterfactual policies. This could be interpreted as estimating the "supply" of candidates/candidate characteristics in addition to the "demand" for candidates.

Such an approach is both natural and necessary for some questions. But it presents significant challenges in our setting. The overall effects of changes in voting policies can reflect not only direct effects on voting costs, but also indirect effects arising through endogenous "supply" responses—e.g., changes in party or candidate spending patterns, policy positions, campaign efforts, or even the identity of candidates selected in primaries. Accounting for these mediated effects is possible when modeling and estimating the equilibrium determination of the mediating factors. But in our setting, this introduces significant challenges, due to the large number of endogenous factors and ambiguities regarding appropriate supply-side modeling.

A more fundamental challenge is that we do not observe all relevant mediating factors. We have data on several endogenous contest-level variables that could respond to voting policy changes—e.g., candidate gender, age, incumbency, and campaign spending. But we lack data on many others—e.g., the charisma of each candidate, their positions on various policy questions, the sizes of their volunteer campaign staffs, and the content and targeting of their ad campaigns. This obviously precludes estimating structural models of these factors. To the extent that omitted factors are affected by the observed exogenous variables x_{dt} (or by candidate instruments for endogenous variables that are observed), these omitted variables introduce problems for the identification of any (even partial) causal effects of the voting policies.

Given these challenges, we instead focus on reduced forms for the fixed effects. Reduced forms, by definition, exclude all endogenous variables but

represent the map from exogenous variables (and reduced-form errors) to ultimate outcomes. We assume the reduced forms follow the linear specification

$$\delta_{Bdt} = \beta_{0B} + x_{dt}\beta_{xB} + w_{dt}\beta_{wB} + \lambda_{Bdt} \tag{17}$$
$$\delta_{Cdt} = \beta_{0C} + x_{dt}\beta_{xC} + w_{dt}\beta_{wC} + \lambda_{Cdt}, \tag{18}$$

where λ_{Bdt} and λ_{Bdt} represent contest-level unobservables and w_{dt} are exogenous observables (instruments) that would enter the structural model for y_{dt} (or for unobserved endogenous factors) but not that for the fixed effects δ_{dt} directly. Linearity of the reduced forms is assumed only for parsimony, reflecting the relatively modest number of contests in our sample. But the restriction to a single scalar error in each equation is important and restrictive. This is analogous to our assumption on contest-level unobservables in the structural model, following standard discrete choice specifications. Such restrictions are often imposed without comment, but they are restrictive. Here this structure, as well as the linearity in (x_{dt}, w_{dt}), can (for example) be derived by assuming linearity of ψ_B, ψ_C, and the reduced forms for all endogenous contest-level factors.[26]

The model we estimate then combines the individual-level voting choice model based on the benefit and cost specification (11)–(12) with the reduced forms (17)–(18) for the contest-level fixed effects. We emphasize that this approach requires assumptions, has limitations, and will not always be appropriate.[27] For our purposes, this approach has the advantage of allowing us to characterize the total effects of counterfactual changes in voting policies without observing all mediating factors and imposing structure sufficient to ensure identification of an expanded structural model.

5 Estimation

We follow a two-step estimation approach, broadly similar to that of Berry, Levinsohn, and Pakes (2004). In the first step we match the model predictions to the data on individual turnout, precinct-level vote shares, and district-level vote shares. This step is based on maximum likelihood and yields estimates

[26] Abusing notation by letting y_{dt} now represent all endogenous contest-level factors (observed or not), suppose the reduced form for each component $y_{dt}^{(k)}$ takes the form $y_{dt}^{(k)} = \chi^{(k)}(x_{dt}, w_{dt}) + \nu_{dt}^{(k)}$, where $\chi^{(k)}$ is linear and $\nu_{dt}^{(k)}$ is potentially correlated with (ξ_{Bdt}, ξ_{Bdt}). Substituting this into (11) and (12) and imposing the assumed linearity of ψ_B and ψ_C yields (17) and (18).

[27] An analogous approach typically will not be helpful in applications to discrete choice demand because primary structural features of interest involve the demand responses to exogenous changes in endogenous choice set characteristics—namely prices.

of the micro parameters θ_1 and contest-level fixed effects $\delta \equiv \{\delta_{Bdt}, \delta_{Cdt}\}_{(d,t)}$. We give additional detail below.

In the second step, we estimate the reduced-form parameters

$$\theta_2 \equiv (\beta_{0B}, \beta_{0C}, \beta_{xB}, \beta_{wB}, \beta_{xC}, \beta_{wC})$$

by GMM, stacking the OLS normal equations for the two reduced forms. The second-step is required only when we examine state-level voting costs in section 9 below.[28] The GMM specification assumes independence across districts of the reduced form errors $(\lambda_{Bdt}, \lambda_{Cdt})_{t=1,2,3}$ but leaves the associated six-by-six covariance matrices for each district fully flexible. This clustering allows arbitrary cross-district heteroskedasticity, contemporaneous within-district correlation, and serial correlation within each district.

In the first-step, the contest-level fixed effects are chosen in a nested fixed point routine that fits district-level turnout and vote shares. This follows the usual practice in discrete choice demand estimation of solving for mean utilities that allow the model to match market shares exactly in large markets. Here, contests are the analogs to markets, and the number of registered voters per contest (i.e., per pair of fixed effects) is roughly 400,000. More generally, the first step utilizes hundreds of millions of individual turnout decisions and hundreds of thousands precinct-level vote shares. This contrasts with 1,173 contests, which is the number of observations in the second step.

To describe the first-step objective function in more detail, let $\mathbb{I}_{pdt}$ denote the set of registered voters in precinct p. Let $\mathbb{I}^A_{pdt}$ denote the subset of $\mathbb{I}_{pdt}$ who turn out (actual voters), with $n^A_{pdt} = \left|\mathbb{I}^A_{pdt}\right|$. For $j \in \{0, 1, 2\}$ let

$$s_{ijdt} = 1\{i \text{ chooses option } j\}$$

and recall the model prediction

$$\sigma_j(z_{idt}; \theta_1, \delta) = E[s_{ijdt}|z_{idt}; \theta_1, \delta] = \Pr(s_{ijdt} = 1|z_{idt}; \theta_1, \delta).$$

Let

$$\sigma^A_j(z_{idt}; \theta_1, \delta) = \sigma_j(z_{idt}; \theta_1, \delta) / (1 - \sigma_0(z_{idt}; \theta_1, \delta))$$

denote the model's predicted probability that individual i votes for candidate j conditional on turning out. Let

$$\bar{s}_{1pdt} = \frac{1}{n^A_{pdt}} \sum_{i \in \mathbb{I}^A_{pdt}} s_{i1dt}$$

[28] Like Berry, Levinsohn, and Pakes (2004) but in contrast to Berry, Levinsohn, and Pakes (1995), the second-step orthogonality conditions are required only for identification of θ_2.

denote the vote share (among actual votes) for candidate 1 in precinct p.

The first-step likelihood can be written as the likelihood of the turnout decisions generating $\mathbb{I}^A_{pdt}$, multiplied by the precinct vote share likelihoods conditional on that turnout. This likelihood is

$$L(\theta_1, \delta) = \prod_t \prod_d \prod_p L^0\left(\mathbb{I}^A_{pdt}; \theta_1, \delta\right) \times L^1\left(\bar{s}_{1pdt} | \mathbb{I}^A_{pdt}; \theta_1, \delta\right)$$

where

$$L^0\left(\mathbb{I}^A_{pdt}; \theta_1, \delta\right) = \prod_{i \in \mathbb{I}_{pdt}} \sigma_0\left(z_{idt}; \theta_1, \delta\right)^{s_{i0dt}} \left(1 - \sigma_{0dt}\left(z_i; \theta_1, \delta\right)\right)^{1 - s_{i0dt}}$$

and

$$L^1_{pdt}\left(\bar{s}_{1pdt} | \mathbb{I}^A_{pdt}; \theta_1, \delta\right) = \sum_{\substack{\mathcal{I} \subset \mathbb{I}^A_{pdt}: \\ |\mathcal{I}| = \bar{s}_{1pdt} \times n^A_{pdt}}} \left(\prod_{i \in \mathcal{I}} \sigma^A_1\left(z_{idt}; \theta_1, \delta\right) \prod_{i' \in \{\mathbb{I}^A_{pdt} - \mathcal{I}\}} \sigma^A_2\left(z_{idt}; \theta_1, \delta\right) \right).$$

Each term $L^1_{pdt}\left(\bar{s}_{1pdt} | \mathbb{I}^A_{pdt}; \theta_1, \delta\right)$ above takes the form of a Poisson-Binomial probability and is computationally intractable. For example, for a precinct with 500 actual voters and a 50% Republican vote share in a given contest, the likelihood contribution involves $\binom{500}{250}$—roughly 10^{150}—terms. However, we can exploit the fact that the Poisson-Binomial is well approximated by a normal distribution and, deriving an approach similar to Ainsworth (2020), replace each $L^1_{pdt}\left(\bar{s}_{1pdt} | \mathbb{I}^A_{pdt}; \theta_1, \delta\right)$ with its normal approximation. We provide details in Appendix C. We construct standard errors for the resulting estimate of θ_1 using a GMM analog of this estimator in which the score (with respect to θ_1) of the quasi-likelihood forms the set of moment conditions.

6 Nonparametric Identification

Although practical considerations dictate the use of a parametric model for estimation, such restrictions are not necessary for identification of the model. Developing formal nonparametric identification results requires a significant detour. Thus, despite the importance of these results, we provide only a high-level summary here, leaving the details to Appendix A.

A key result addresses identification of functions

$$\tilde{\sigma}_j(z_{idt}, x_{dt}, w_{dt}, \lambda_{dt}) \equiv \Pr(i \text{ chooses option } j | z_{idt}, x_{dt}, w_{dt}, \lambda_{dt}) \tag{19}$$

for all j, along with the reduced-form errors λ_{dt}. The functions $\tilde{\sigma}_j$ represent the structural relationship between demographics and voting choice probabilities

conditional on the exogenous contest-level variables (x_{dt}, w_{dt}) and the reduced-form errors λ_{dt}. Thus, their identifiability provides a nonparametric foundation for our empirical approach combining our model of individual-level choice with contest-level reduced forms. We also provide conditions for identification of the joint distribution of (B_{idt}, C_{idt}) conditional on $\{z_{idt}, x_{dt}, w_{dt}, \lambda_{dt}\}$. Although not necessary for identification of the counterfactual quantities we examine below, this aids interpretation by ensuring nonparametric identification of the underlying random utility model.[29]

Some of our identification results rely on those obtained by Berry and Haile (2024) for demand models using micro data linking consumer characteristics to the goods they choose. However, two features of the present setting prevent applying those results directly. One is the type of micro data available: we observe characteristics of registered voters and whether they vote, but not the candidate selected by any voter. We overcome this shortcoming by exploiting the observability of demographic distributions and vote shares at the precinct level.[30] The second distinction reflects the nature of the choice problem: in contrast to goods in standard discrete choice demand models, the available options here are not "weak gross substitutes." For example, an improvement in the quality of candidate 1 can (all else equal) drive *up* the share of nonparticipation, due to the presence of voters whose relative preference for candidate 2 now becomes insufficient to overcome their costs of voting. This implies violation of properties typically used to demonstrate invertibility of choice probability mappings (see, e.g, Berry, Gandhi, and Haile (2013)).[31] However, we show that natural conditions on the voting model imply invertibility.

With these two issues resolved, the key results can be obtained following Berry and Haile (2024). Broadly speaking, these results combine two sources of "clean" variation in the voting context: (a) variation across voters within a given contest (i.e., where all contest-level unobservables are fixed) and (b) cross-contest variation driven by exogenous contest-level observables. This combination of within-contest and cross-contest variation is essential: without additional *a priori* structure, data from a single contest (no matter how large the population of precincts and potential voters) does not suffice for identification (see Berry and Haile (2024)).

---

[29]The appendix also covers identification of the functions σ_j in (8)–(10) (and the corresponding conditional distribution of (B_{idt}, C_{idt})) when one observes and instruments for all endogenous variables.

[30]This result extends to other discrete choice or demand settings in which one observes share data at a finer ("sub-market") level than that at which the choice set varies.

[31]Invertibility is a key property exploited to obtain identification of models in which multiple structural errors determine each observed outcome. See, e.g., Berry (1994), Berry, Levinsohn, and Pakes (1995), Matzkin (2008, 2015), and Berry and Haile (2014, 2018, 2021).

7 Estimates

7.1 Micro Parameters and Contest Level Fixed Effects

Tables 4–6 display our estimates of the first-step "micro" parameters θ_1. Most signs are consistent with conventional wisdom about the association between voter demographics and preferences, even though that conventional wisdom may reflect partial correlations rather than the *ceteris paribus* relations estimated here. Voters are more likely to prefer the Republican candidate when they are men, older, white, or richer. In contrast, the Democrat tends to be preferred by voters who are non-white, more educated, or non-rural.

Table 4: Micro Coefficients: Benefit

Variable	Coefficient	Standard Error
Republican	0.648	0.044
Democrat	−0.132	0.025
Male	0.096	0.005
Age/100	0.416	0.046
White	0.003	0.005
Black	−0.259	0.015
Hispanic	−0.080	0.007
Education/16	−0.537	0.026
Income/$250k	0.076	0.012
Urban	−0.067	0.005
Suburban	−0.042	0.004
Family Size/4	0.065	0.012
Recent Registration	−0.156	0.012
Near tract $R/(R+D)$	0.075	0.013
Near independent share	0.004	0.019
Near tract $\lvert R-D\rvert/(R+D)$	−0.018	0.007
Own tract $R/(R+D)$	0.306	0.019
Own tract independent share	−0.081	0.024
Own tract $\lvert R-D\rvert/(R+D)$	−0.005	0.006

In Table 5, we see that several of the observed factors associated with voting Republican are also associated with lower voting costs. All else equal, registered Republicans have lower voting costs than registered Democrats (whose costs are lower than those of independents). Older voters and richer voters also have lower voting costs, as do White voters and non-urban residents. Running the opposite direction, more highly educated voters have lower voting costs, while men have somewhat higher voting costs, all else equal.

Table 5: Micro Coefficients: Cost

Variable	Coefficient	Standard Error
Republican	−0.621	0.040
Democrat	−0.324	0.021
Male	0.031	0.006
Age/100	−1.482	0.011
White	−0.102	0.003
Black	0.196	0.018
Hispanic	0.115	0.005
Education/16	−0.922	0.070
Income/$250k	−0.353	0.006
Urban	0.023	0.006
Suburban	−0.013	0.004
Family size/4	−0.423	0.007
Recent registration	0.106	0.012
Near tract $R/(R+D)$	0.074	0.013
Near independent share	−0.189	0.023
Near tract $\lvert R-D\rvert/(R+D)$	0.044	0.007
Own tract $R/(R+D)$	−0.413	0.017
Own independent share	0.339	0.030
Own tract $\lvert R-D\rvert/(R+D)$	0.143	0.008

Table 6: Micro Parameters of Joint Normal Errors

Variable	Coefficient	Standard Error
Std Dev of ϵ_{Cidt}	1	—
Std Dev of ϵ_{Bidt}	0.496	0.023
Corr$(\epsilon_{Bidt}, \epsilon_{Cidt})$	−0.515	0.073

In Table 6, we see a similar pattern for voter-level unobservables: the normal shocks ϵ_{Bidt} and ϵ_{Cidt} are negatively correlated, enhancing the association between turning out and preferring the Republican candidate. However, the correlation is far less than perfect, implying that a single dimension of unobserved heterogeneity would not be sufficient to describe voting behavior.

The contest-level fixed effects are also estimated in the QMLE first step. Figure 4 displays the mean (over the three election years) estimated benefit fixed effect within each district. Compared to the map of electoral wins in Figure 1, this figure shows significantly more Democratically leaning districts. This points to the role costs in driving vote shares through selective turnout. Figure 5 displays the mean estimated cost fixed effects for each district. One feature that stands out is that costs tend to be higher in the South.

Figure 4: Mean Benefit Fixed Effects ($B - \epsilon_B$) by District

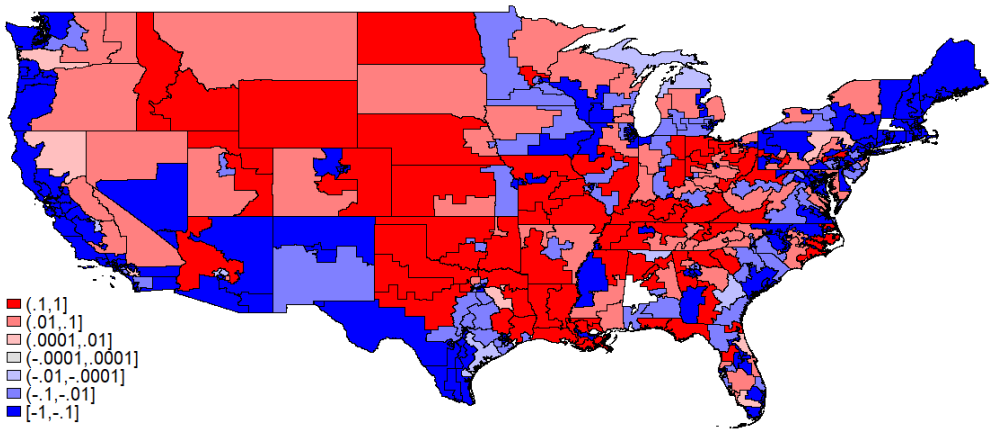
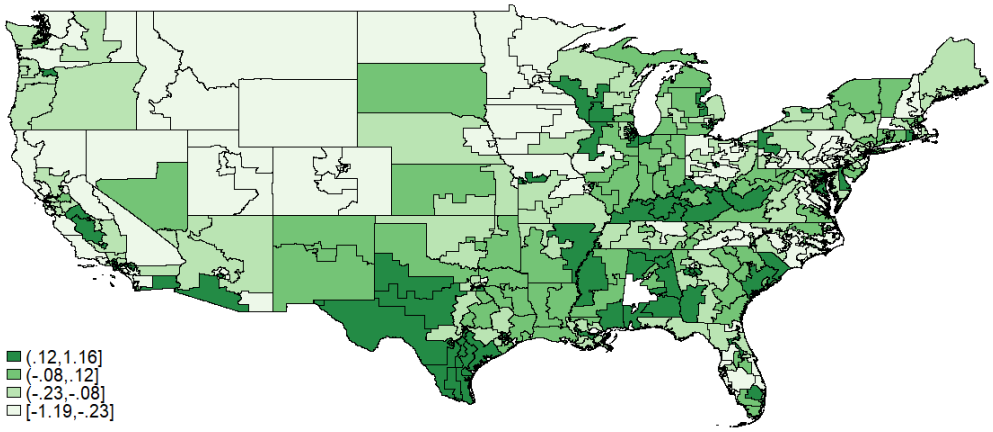


Figure 5: Mean Cost Fixed Effects ($C - \epsilon_C$) by District



7.2 Reduced-Form Parameters

Table 7 shows our estimates of the reduced-form parameters θ_2.[32] Recall that the micro portion of our model already accounts for the demographics of voters themselves, their close neighbors (same census tract), and their more distant neighbors (nearby census tracts), assuming that these are the demographic measures that either directly affect voter preferences or proxy for latent factors that do. The role of the district-level demographics here is different. District-wide demographics shape the preferences of voters in the entire district and, therefore, affect things like which candidates run (or win in the primaries), their positions on various policies, their campaign spending, or social norms with respect to turnout. However, while the potential influence of these demographics on the benefit and cost fixed effects is clear,[33] the expected signs of the coefficients are not, in part because the unmodeled endogenous contest-level measures like campaign spending or policy positions are equilibrium outcomes in a competitive election.

Although a similar difficulty applies when predicting signs of some other measures in Table 7, the expected signs are clearer (and confirmed) for several of these. For example, if advertising is thought to intensify voter preferences or to enhance perceived vote efficacy, the coefficient on media advertising cost would be positive in the voting cost fixed effect column. The 2018 fixed effects reveal both the typical midterm backlash against the presidential party (e.g., Erikson (1988)) and an increase in voting cost consistent with the a reduction in incentives for turnout in midterm elections, all else equal. The two measures of state voting policies affect mean voting costs in the expected direction. In contrast, the expected effects of these measures on the mean benefit (i.e., preference for the Republican candidate) are not clear, and our results suggest a mix of small effects, among which the only positive coefficient on voter identification policies is significant. This suggests the potential for equilibrium responses to soften the impacts of changes in these policies.

[32] Shares of registered Independents are omitted from the reduced forms, since the shares of Democrats, Republicans, and Independents sum to one.

[33] In addition to possible direct effects, recall that anything altering the benefit potentially affects the perceived vote efficacy and, therefore, the perceived effective voting cost.

Table 7: Contest-Level Reduced Forms

Variable	Benefit FE Estimate	Benefit FE SE	Cost FE Estimate	Cost FE SE
Contest-Level Average Demographics				
Republican	−0.297	0.107	1.073	0.311
Democrat	0.008	0.079	0.419	0.165
Male	0.545	0.442	−3.282	1.175
Age/100	−0.647	0.185	1.090	0.528
White	0.258	0.089	−0.186	0.181
Black	0.309	0.087	−0.375	0.190
Hispanic	0.272	0.091	−0.184	0.178
Education/16	−0.448	0.257	−0.327	0.581
Income/$250k	0.224	0.144	−0.315	0.363
Urban	−0.013	0.032	0.045	0.060
Suburban	−0.013	0.029	−0.112	0.053
Family size	0.208	0.199	1.244	0.527
Recent registration	−0.367	0.177	−2.104	0.458
$R/(R+D)$	0.299	0.123	−0.685	0.326
$\lvert R-D\rvert/(R+D)$	−0.100	0.022	0.051	0.050
State Voting Cost Indices and Up-Ballot Measures				
Voting inconveniences	0.000	0.002	0.007	0.004
Voter ID Laws	0.006	0.003	0.036	0.007
Gov. race has open seat	0.016	0.008	−0.138	0.017
State has governor's race	−0.012	0.007	0.055	0.015
Senate race has open seat	−0.008	0.012	−0.003	0.018
State has Senate race	0.010	0.004	−0.002	0.006
Other Contest-Level Variables				
Log(ad price)	−0.011	0.006	0.029	0.010
Media market overlap	−0.055	0.024	0.052	0.058
Constant	−0.063	0.353	2.658	0.764
Year=2018	−0.129	0.008	0.339	0.020
Year=2020	0.003	0.007	−0.104	0.014
Observations	1,173		1,173	
R^2	0.487		0.677	

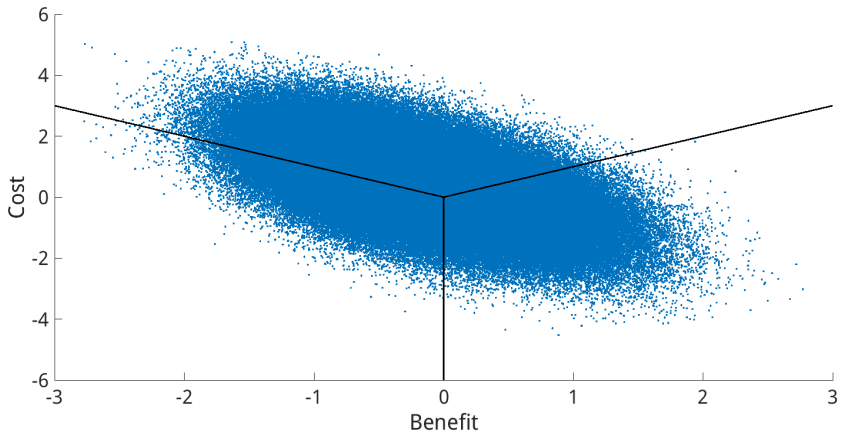
8 Selection, Preference Aggregation, and the Preferences of Marginal Voters

8.1 Selective Turnout and Preference Aggregation

Several important questions concern the effects of selective turnout on election outcomes and representation. One can already see evidence of the potential role of selection in the maps above. For example, consider Texas District 23—the distinctively large district at the southwest edge of the state. We saw in Figure 1 that votes in this district were split almost evenly—e.g., 50.38% for Republicans in 2018. In that same year we estimate that only 35.17% of registered voters preferred the Republican (see Figure 4). Figure 5 suggests why the discrepancy is possible: this district has high average effective voting costs.

Of course, the difference between the preferences of registered voters and those of actual voters is not explained by high voting costs alone: the strong positive association between voting cost and preference for Democrats is also important. Our estimates reveal that this dependence reflects both voter-level observables and voter-level unobservables; thus, our rich demographics allow this association to vary substantially across districts. In this district, Figure 6, shows all registered voters in (B_{idt}, C_{idt})-space for 2018, using their observed demographics, the estimated model, and simulated draws of $(\epsilon_{Bidt}, \epsilon_{Cidt})$. Here we also show the "Y"-shaped partition of (B_{idt}, C_{idt})-space into the implied voting choices (recall Figure 2). As illustrated here, the model predicts that a substantial fraction of registered voters will fail to turn out, and that these are disproportionately registered voters who prefer Democrats.

Figure 6: Benefit & Cost, Individual Level
TX23, 2018



The Texas 23rd is, of course, just one district, although it illustrates the potentially important role of multidimensional selection in determining vote outcomes. One way to generalize this to the entire country is to consider a hypothetical setting in which all voting costs are zero—i.e., when every registered voter turns out and votes for their preferred candidate.[34] Table 8 compares the resulting Republican seat count in the House to the actual count for each election year. All of our counterfactual confidence intervals use percentile 95% intervals from a parametric bootstrap with 1000 replications.[35]

Table 8: Republican Seats Under 100% Turnout

Year	Baseline Seats	Counterfactual Seats
2016	241	196
		[162,220]
2018	200	76
		[51,107]
2020	213	155
		[140,167]

95% bootstrap confidence intervals in brackets.

The gap between the baseline and counterfactual seats is striking, with Republicans losing a large fraction of seats under 100% turnout. However, the results in this table should be interpreted with some caution. Although zero (even negative) voting costs are within the support of our estimates, driving costs to zero for all voters requires a prediction far from the sample. Furthermore, by altering cost but not benefit, this exercise cannot be interpreted as an equilibrium counterfactual in which, for example, candidates might take different policy positions knowing that all registered voters will turn out. Nonethe-

[34]A complication in any analysis of counterfactual vote outcomes is the existence of uncontested races in the data (recall that uncontested races are excluded in estimation). We follow a common convention in the literature of imputing a vote share of 75% for the uncontested party (see, e.g., Katz, King, and Rosenblatt (2020b) and Gelman and King (1994)). Setting the reduced-form cost error λ_{Cdt} to its unconditional mean (zero), we then back out the reduced-form benefit error λ_{Bdt} that would rationalize the imputed vote share. We substitute these imputed values of $(\lambda_{Bdt}, \lambda_{Cdt})$ for their missing estimates before proceeding to the counterfactual computations. Our results are virtually unchanged if, instead of a 3–1 imputed vote ratio we use 2–1 or 4–1. We also obtain virtually identical average results (but with more year-to-year variability) if we take the more extreme approach of holding vote outcomes fixed in uncontested races—equivalent to imputing an 100% vote share for the uncontested party and, thus, an infinitely large (positive or negative) λ_{Bdt}.

[35]Our bootstrap procedure follows Nevo (2001). For each bootstrap replication we draw a θ_1 vector from its estimated asymptotic normal distribution and calculate the implied fixed effects $\delta(\theta_1)$. Using these values we then calculate the counterfactual quantity of interest.

less, the results again suggest a strong tendency for Democratic preference to be underrepresented among actual voters.

8.2 The Preferences of Marginal Voters

A less extreme way to examine the importance of selective turnout is to isolate the preferences of voters close to the turnout margin. We do this here by examining the estimated model's predicted votes of marginal voters in a district-level 5% swing in turnout. To construct the set of marginal voters, for each district we first calculate the change in mean voting cost that would reduce turnout by 2.5%. Similarly, we find the reduction in mean voting cost that would increase turnout by 2.5%. We then examine the model's predicted vote share for the voters whose turnout status changes in this simulated -2.5% to +2.5% swing in turnout.

Figure 7: Republican Vote Share among 5% Marginal Voters

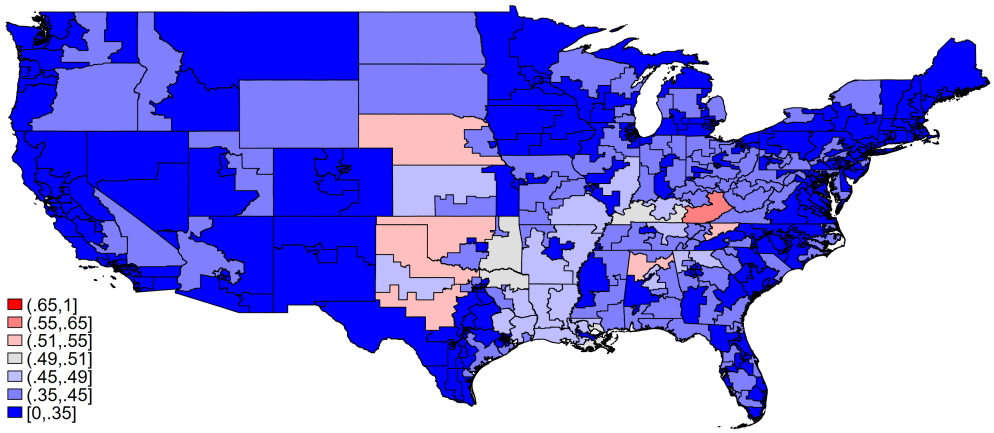


Figure 7 illustrates the results.[36] In most districts, Democratic votes dominate among marginal voters, although by varying margins. The Republican candidate is preferred by a majority of marginal voters in just 7 out of 435 districts. Furthermore, the systematic preference for Democrats among marginal voters is strong: 72% of districts are shaded dark blue, indicating that Republicans are predicted to receive less than 35% of the votes among the 5% marginal voters. This is consistent with conventional wisdom that making it easier to vote tends to benefit Democrats, all else equal.

In Table 9 we compare the average demographics of the 5% marginal voters to those of actual voters and of all registered voters. We saw in Tables 1 and

[36] We provide the underlying point estimates and confidence intervals in Appendix D Table 14. For 2016, the mean share for marginal is 0.286, [0.233, 0.336], for 2018 is 0.273, [0.225, 0.319], and for 2020 is 0.269, [0.216, 0.320].

2 that actual voters differ from the population of registered voters in these observable dimensions. Table 9 shows that marginal voters differ from the pool of actual voters. For example, voters on the margin tend to be younger, less white, recently registered, and less likely to be registered Republicans. In most dimensions, the mean demographics of the 5% marginal voters are very similar to those of the full population of registered voters. A notable exception is party affiliation: marginal voters are somewhat more likely to be registered Democrats, and substantially less likely to be registered Republicans.

Table 9: Demographics of Marginal Voters

Demographic	5% Marginal	Actual	Registered
Age/100	0.475	0.529	0.505
Income/$250	0.342	0.381	0.364
Male	0.467	0.459	0.468
Education/16	0.790	0.803	0.798
White	0.593	0.680	0.649
Hispanic	0.143	0.095	0.111
Black	0.117	0.086	0.098
Family size/4	0.499	0.528	0.514
Recent reg.	0.098	0.070	0.080
Republican	0.230	0.371	0.323
Democrat	0.414	0.396	0.387
Urban	0.341	0.300	0.313
Suburban	0.415	0.447	0.435

The sample covers district-years with a contested House race.

9 The Effects of State Voting Policies

The results above suggest that perceived voting costs have significant effects on election outcomes. Some of these voting costs are affected by state-level policies, although the effects of particular policies depend on who is marginal. To explore the role of voting policies, we exploit our model's implication that states can be ranked by the two policy indices—voter inconveniences and voter ID requirements. We focus on two counterfactuals. The first examines outcomes with policies that yield low voting costs in every state. Here we set the policy indexes in each state to the minimum of their actual values and the 10th percentile values among all states. Thus, in the low-cost counterfactual all states have voting policies at or below the component-wise 10th percentiles. Similarly, we consider a high-cost simulation in which state voting cost policies are all at or above the component-wise 90th percentile values.

In both counterfactuals, we allow the change in policies to alter both δ_C and δ_B through the estimated reduced forms. Unlike our examination of marginal voters, this allows a variety of potential equilibrium responses to the policy changes.[37] In addition, while our analysis of marginal voters considered a 5% swing in every district, here the turnout responses will differ across districts. This reflects the cross-state differences in baseline policies, differences in how many voters are close to the turnout margin, and the nonlinearity in turnout responses to changes in voting costs.

9.1 Voting Policies and Turnout

Figures 8 and 9 illustrate the changes in turnout (abstention) by state under the low and high cost counterfactuals, respectively. In all states, the direct effect of the policy-controlled voting costs dominate, implying larger turnout when voting policies are more lenient. The effects are heterogeneous across states, both because states differ in their baseline policies and because different numbers of voters will be on the turnout margin in each state. Tables 15 and 16 in Appendix D shows the point estimates and confidence intervals.[38]

We can also use our model to show how the turnout responses in these counterfactuals differ across demographic groups. In Table 10 we see that the predicted effects of the policy are largest for Blacks, Hispanics, younger voters, low education voters, and recently registered voters—all groups with notably low baseline turnout. Registered Republicans (or, in states without partisan registration, voters imputed to have Republican affiliation) are both the group with the highest baseline turnout and the group whose turnout responds least to the policy changes.

9.2 Voting Policies and Election Outcomes

Figures 10 and 11 illustrate the changes in vote shares by state. Restrictive (high costs) policies help Republicans, while more permissive (low cost) policies help Democrats. This is consistent with conventional wisdom and our analysis of marginal voters in section 8. Note that we estimated a modest positive coefficient on the voter ID policy measure in the reduced form for the benefit (of electing the Republican), working against the conventional wisdom. This

[37] Note also that any equilibrium effects of voting policies on voters' perceived vote efficacy would be captured by the reduced form for δ_C and, thus, accounted for here.

[38] Here we extend our parametric bootstrap procedure (see footnote 35), drawing both θ_1 and θ_2 from their asymptotic normal distribution in each replication. These values of the parameters and the implied fixed effects $\delta(\theta_1)$ and reduced-form errors are then used to construct the bootstrap replication of the counterfactual quantity of interest.

Table 10: Turnout by Demographics
High- and Low-Cost Counterfactuals

Demographic	Base-line	High Cost	%Δ	Low Cost	%Δ
Overall	0.70	0.65	-6.9%	0.72	3.6%
		[0.639, 0.667]		[0.710, 0.725]	
Republican	0.80	0.77	-3.5%	0.82	2.0%
		[0.765, 0.784]		[0.810, 0.821]	
Democrat	0.71	0.66	-7.6%	0.73	3.5%
		[0.645, 0.678]		[0.725, 0.741]	
Male	0.68	0.64	-7.1%	0.70	3.7%
		[0.627, 0.654]		[0.696, 0.711]	
White	0.73	0.70	-5.7%	0.75	3.2%
		[0.682, 0.707]		[0.746, 0.760]	
Black	0.60	0.56	-8.4%	0.64	6.6%
		[0.540, 0.572]		[0.626, 0.649]	
Hispanic	0.59	0.54	-10.9%	0.61	3.5%
		[0.518, 0.555]		[0.606, 0.620]	
Urban	0.66	0.62	-8.3%	0.69	3.8%
		[0.600, 0.631]		[0.677, 0.693]	
Suburban	0.72	0.67	-6.7%	0.74	3.2%
		[0.660, 0.687]		[0.729, 0.744]	
Recent Reg.	0.61	0.56	-9.4%	0.63	4.7%
		[0.539, 0.571]		[0.622, 0.639]	
Age > Median	0.78	0.74	-5.0%	0.80	2.8%
		[0.728, 0.752]		[0.789, 0.802]	
Age < Median	0.62	0.57	-8.8%	0.64	4.5%
		[0.551, 0.581]		[0.630, 0.648]	
Ed > Median	0.75	0.71	-6.3%	0.77	2.8%
		[0.694, 0.721]		[0.761, 0.774]	
Ed < Median	0.54	0.48	-11.5%	0.56	4.7%
		[0.466, 0.613]		[0.550, 0.675]	
Inc > Median	0.74	0.70	-6.3%	0.76	2.7%
		[0.680, 0.710]		[0.745, 0.764]	
Inc < Median	0.65	0.61	-7.5%	0.68	4.5%
		[0.595, 0.628]		[0.668, 0.690]	

95% percentile bootstrap confidence intervals in brackets.

Figure 8: Low Cost Simulation: Effects on Abstention

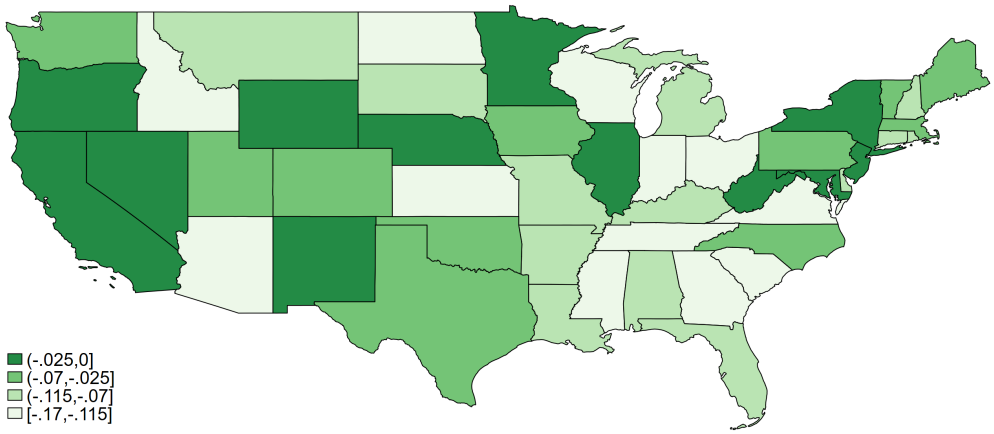
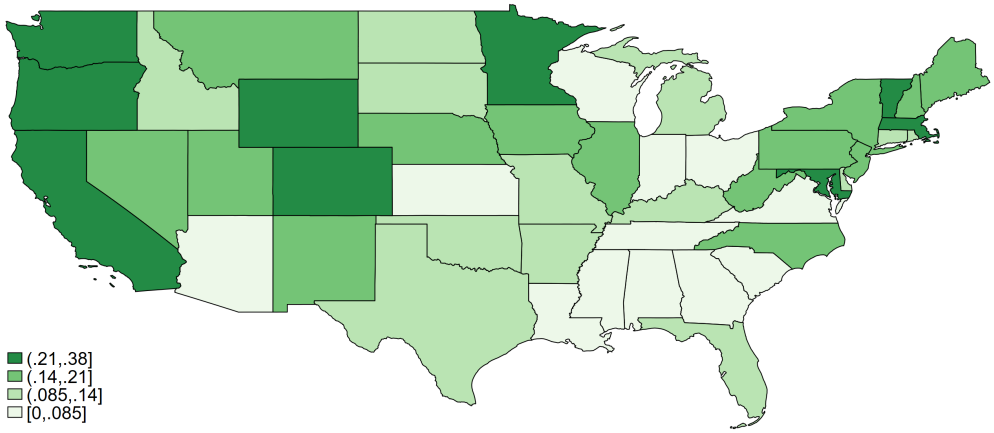


Figure 9: High Cost Simulation: Effects on Abstention



the conventional wisdom in any state.

The magnitudes of the effects differ across states in the two counterfactuals. This reflects a combination of (i) differences in baseline policies, (ii) differences in turnout responses, and (iii) differences in the preferences of the marginal voters. We explore this asymmetry further below.

Figures 10 and 11 show only the change in vote shares, and only for contested races. We can also look at predicted election outcomes, accounting for uncontested races as well (recall footnote 34). Table 11 shows the resulting changes in the composition of the House under both counterfactuals. There we see that the high-cost counterfactual predicts nontrivial gains for Republicans: gains of 10, 24, or 18 seats depending on the election year. In both 2018 and 2020, our point estimates imply a predicted Republican gain that would have reversed the Democratic majority (ignoring independents, 218 seats are needed for a majority). Averaging over the three years, the mean predicted

Figure 10: Low Cost Simulation: Changes in Vote Share

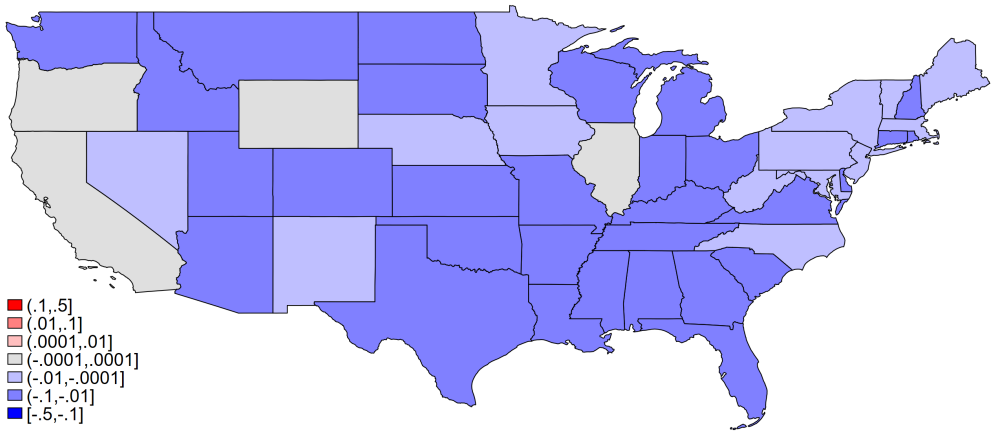
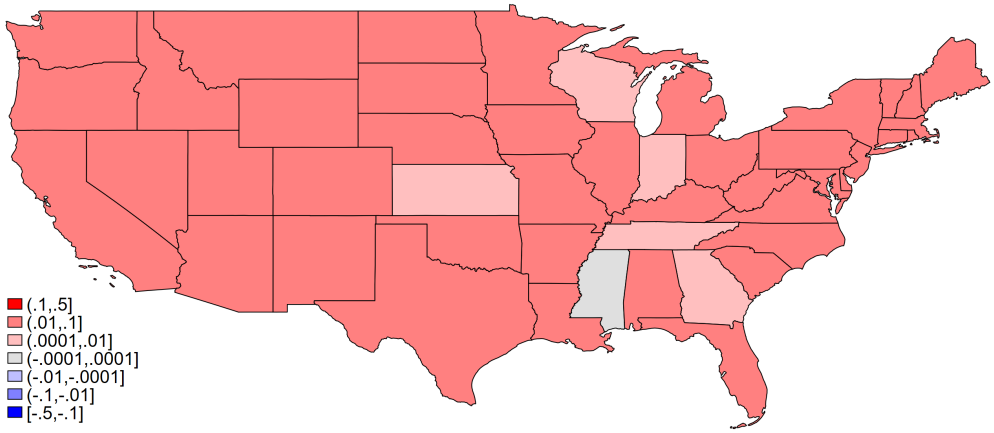


Figure 11: High Cost Simulation: Changes in Vote Share



gain of 17 seats for Republicans represents 3.9 percent of House seats. Such a swing from one party to the other would have been sufficient to reverse the majority in the House in 8 of the 13 Congresses since 2001.

The low-cost counterfactual shows somewhat more modest effects. Democrats gain from more permissive voting policies, but only 5 seats on average across years. This asymmetry is possible in part because baseline voting policies are not randomly assigned. For example, it is consistent with voting policies already being relatively permissive in the states whose seat outcomes are most sensitive to policy-driven voting costs.

Figure 12 shows that this is indeed the case. Here the horizontal axis is the state-level cost policy index (the sum of the two policy-driven terms in the estimated reduced form for voting cost). The vertical axis is the predicted seat change when moving from the low-cost to high-cost counterfactual. Each point plotted in the figure represents a state. Here we see that states in which

changes voting policy generate the largest predicted change in seats are disproportionately those for which actual voting policies are less restrictive.[39] Even when accounting for the many states in which we predict no seat change, the relationship shows distinct downward slope (the solid line). To the extent that the observed range of voting policies reflects the feasible set, this indicates an asymmetry between the two parties in terms of incentives to alter state=level voting policies: Republicans have more to gain from making policies more restrictive than Democrats do from making them less restrictive.

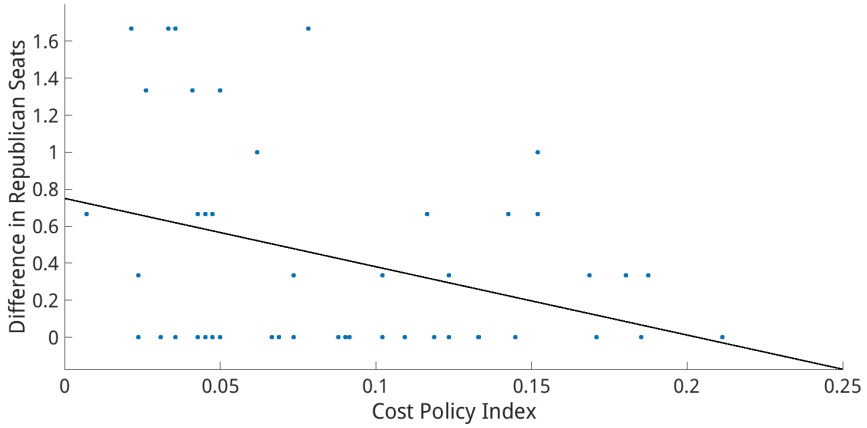
Table 11: Voting Policy and Prediced Seat Sensitivity

Year	Counterfactual Type	Baseline Seats	Counterfactual Seats	Percent Change
2016	High cost	241	251 [246,258]	4.15
2018	High cost	200	224 [211,234]	12.00
2020	High cost	213	231 [221,243]	8.92
2016	Low cost	241	240 [239,242]	-0.41
2018	Low cost	200	193 [185,197]	-3.50
2020	Low cost	213	206 [202, 211]	-7.98

95% percentile bootstrap confidence intervals in brackets.

[39]This offers a possible reconciliation of our findings with those of Cantoni and Pons (2021b), whose difference-in-difference analysis finds only small effects of stricter voter ID laws in the states that adopted them between 2008 and 2018.

Figure 12: Status Quo Policy Index and
Republican Seat Share Difference (Mean): High vs. Low Cost



Each dot represents a state average. A regression line is shown.

10 Partisan Bias of State District Designs

A major political and legal issue in the U.S. is how congressional district boundaries are drawn in states with more than one congressional seat.[40] District design can have significant effects on the representation of each state's voters' preferences and even on the overall composition of the U.S. House. Quantitative measurement of bias in favor of one party or the other is challenging: it requires both a measure of distortion relative to some "neutral" ideal and, often, prediction of counterfactual voting outcomes.

One standard measure for this purpose is "partisan bias," a measure of asymmetry between parties in transforming votes to seats. Take a given state and let τ denote its total number of seats (districts). For party $j \in \{D, R\}$, let $\tau_j(s)$ denote the seat share party j would win if the statewide vote share for party j were s. Partisan bias at the target vote share s is defined as

$$PB(s) = [\tau_R(s) - \tau_D(s)]/\tau.$$

Positive (negative) values indicate bias in favor of Republicans (Democrats). For example, at $s = 0.5$, a negative value means that votes are aggregated in a way allowing Democrats to win more than half the state's seats when statewide votes are split equally.

Partisan bias cannot be calculated from vote and seat outcomes alone. It is defined by counterfactuals involving the target statewide vote share s.

[40] Eguia (2022) discusses the legal status of partisan gerrymandering and competing notions of fairness for evaluating district designs.

There is, of course, no single "right" way to adjust statewide vote shares. But our model suggests an approach that accounts for differences in how districts respond to forces driving statewide changes in vote shares. Fix a state and election year t. For party j and target share s, we predict $\tau_j(s)$ by adding a constant $\kappa_t(s)$ to our estimated fixed effects $\hat{\delta}_{Bdt}$, choosing $\kappa_t(s)$ so that the statewide vote share for party j in year t is s.[41] We then use the model predictions of district-level outcomes, aggregating these to predict the fraction of seats won at the state level.

Figure 13: Partisan Bias at 50-50 Statewide Votes

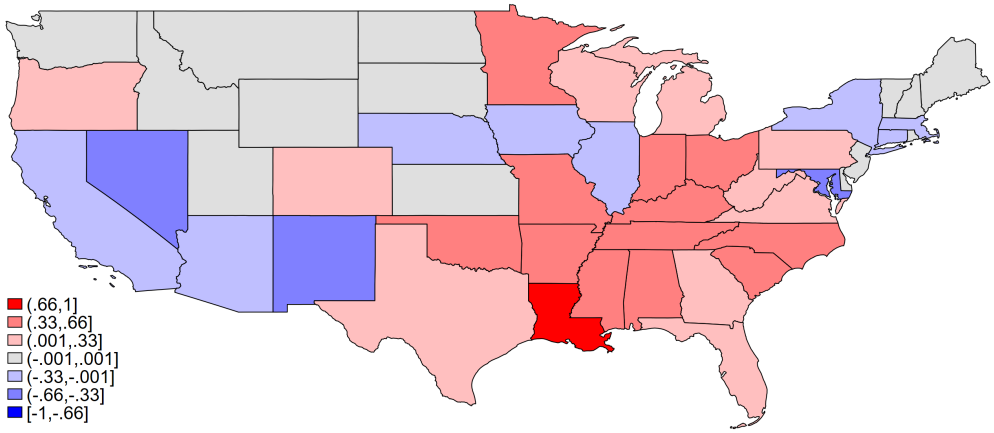


Figure 13 illustrates the partisan bias at a vote share of 50%, averaged over the three election years.[42] Because seats are lumpy, we should not expect the bias to be exactly zero even with neutral district designs.[43] Overall, however, states with a Republican bias outnumber those with a Democratic bias almost 2 to 1. This pattern can be seen more clearly in Figure 14, where we consider a weighted average of partisan bias measures centered at the 50% vote share.[44] The horizontal axis measures the number of seats in the state (we exclude single-district states). The vertical axis is the estimated partisan bias in absolute value, with colors indicating the direction of the bias. The units

[41] Although there is no need to estimate the reduced forms to perform this exercise, it can be interpreted as adding a shock of size $\kappa_t(s)$ to each of the state's district-level reduced-form errors λ_{Bdt}.

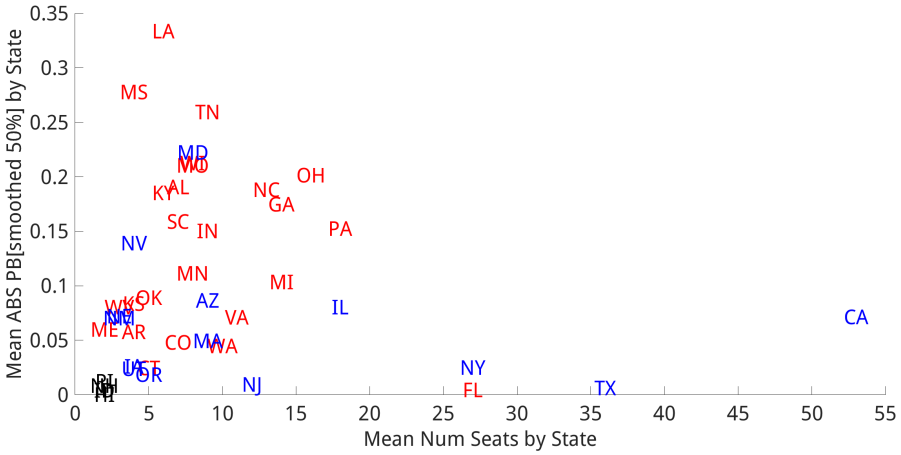
[42] For counterfactual quantities presented graphically in this section and those that follow, we provide the underlying point estimates and standard errors in Appendix D.

[43] In single-district states, partisan bias is always zero by definition. When considering House-level outcomes under 50-50 statewide votes, we allocate half a seat to each party in the single-district states.

[44] This is based on target vote shares for each election year ranging from 0.4 to 0.6 with weight 1/3 on 0.5 and 1/6 on each share 0.4, 0.45, 0.55, 0.6.

describe a percentage advantage for one party; for example, partisan bias of 0.05 favoring Republicans means that despite an equally split statewide vote, Republicans are predicted to win 5% more of the state's seats on average. There are many states with estimated bias close to zero, but also many with partisan bias much larger than 5% in absolute value. The vast majority of states with substantial partisan bias (e.g., above 0.05 in absolute value) favor Republicans. This includes 15 of the 17 states whose estimated bias exceeds 0.1 in absolute value, 6 of 7 states with absolute bias above 0.2, and all 3 states with absolute bias above 0.25.

Figure 14: Absolute Partisan Bias at (smoothed) 50-50 Statewide Votes



Although partisan bias allows a type of apples-to-apples comparison across states, the effect of a given state's bias on the composition of the House depends on its size (number of seats). In Figure 15 we rescale each state's partisan bias by its number of congressional seats. The resulting values represent the number of expected "excess" seats won by the advantaged party when the state vote share is split 50-50. Here we see that the relatively modest partisan bias in California favoring Democrats accounts for a significant number of excess Democratic seats. However, excess Republican seats still dominate.

We can see this dominance directly by examining the predicted makeup of the full House under the hypothetical 50-50 vote share in every state. Table 12 shows the results. Averaging over the three years, Republicans would win an estimated 54.5% of seats in the House (vs. 45.5 for Democrats) when votes are evenly split within each state. To put this 9.0 percentage point advantage in context, in the 13 Congresses this century, the average advantage of the majority party is approximately 7.5 percentage points.

Another way to describe the aggregate asymmetry between parties is to show, for each party, the predicted share of all House seats won as a function

Figure 15: Excess Seats at (smoothed) 50-50 Vote Shares

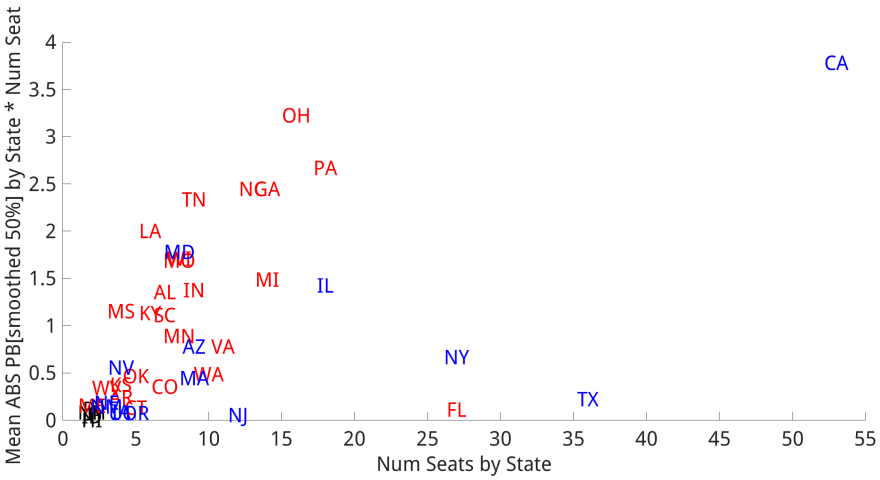


Table 12: Predicted Republican Seats at 50-50 Statewide Vote Shares

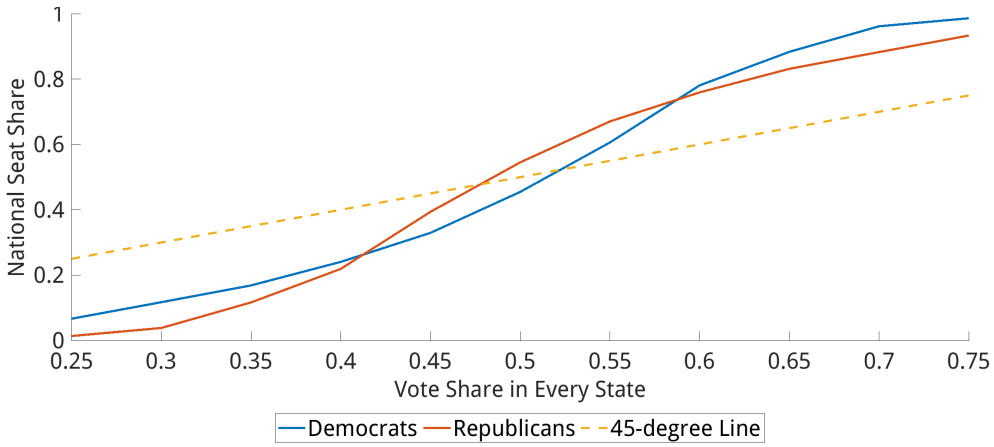
Year	Republican Seats Won	Share
2016	250.5	57.59
	[248.5, 250.5]	
2018	232.5	53.45
	[231.5, 234.5]	
2020	228.5	52.53
	[227.5, 229.5]	

95% bootstrap confidence intervals in brackets.

of statewide vote shares. Figure 16 shows these nationwide "votes-to-seats" curves (e.g., Niemi and Deegan (1978)), focusing on shares between 0.25 and 0.75.. These curves require predictions farther out of sample. However, for both parties the estimated votes-seats curve is S-shaped and steeper than 45 degree line in the most relevant range, especially for vote shares between 40% and 60%.[45] Over this range, the votes-seats curves for the two parties have similar slopes—similar "responsiveness" of seats to votes on the margin. However, Republicans enjoy an advantage through most of this range. This relative advantage is reversed at more extreme vote shares.

[45]Such a shape arises naturally from the lumpiness of representation for each state. Under our approach for simulating target statewide vote shares, there is also an automatic symmetry property: the Democratic seat share at a target vote share s equals the Republican seat share at target $(1-s)$.

Figure 16: Nationwide Votes-Seats Curves



11 Conclusion

We have examined how election outcomes are influenced by preferences, policy, and multi-dimensional selection into voting. We found that marginal voters have a strong tendency to prefer Democrats, and that the composition of the U.S. House over-represents preferences for Republicans. State voting policies have the potential to affect the overall balance of the House. But because states whose representation is most sensitive to voting policy already tend to have relatively permissive rules, Republicans have more to gain from making voting policies more restrictive than Democrats have to gain from making them less so—at least when considering policies in the range currently observed. We also found that many states have congressional maps exhibiting partisan biases in one direction or the other, but with an overall bias favoring Republicans. This compounds the effects of selective turnout in favoring Republicans on net.

Of course, any effort to characterize counterfactual voting outcomes or the preferences of non-voters requires a model. A Downsian model—no matter how flexible—imposes restrictions that surely do not characterize the motivations of very potential voter. And while our estimation approach has a number of advantages, it also relies on important assumptions and does not allow us to answer all questions. Thus, further complementary evidence from alternative voting models and research designs will be valuable. Other promising avenues for future work include analyzing district design beyond partisan bias and extending the analysis to include voter registration.

References

Ainsworth, R. (2020): "Measuring Gerrymandering by Recovering Individuals' Preferences and Turnout Costs," working paper, University of Florida.

Alvarez, R. M. and J. Nagler (1998): "When Politics and Models Collide: Estimating Models of Multiparty Elections," *American Journal of Political Science*, 55–96.

Andrews, D. W. K. (2017): "Examples of L^2-Complete and Boundedly Complete Distributions," *Journal of Econometrics*, 199, 213–220.

Baltz, S., A. Agadjanian, D. Chin, J. Curiel, K. DeLuca, J. Dunham, J. Miranda, C. H. Phillips, A. Uhlman, C. Wimpy, et al. (2022): "American Election Results at the Precinct Level," *Scientific Data*, 9, 651.

Berry, S. (1994): "Estimating Discrete Choice Models of Product Differentiation," *RAND Journal of Economics*, 23, 242–262.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 60, 889–917.

——— (2004): "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Vehicle Market," *Journal of Political Economy*, 112, 68–105.

Berry, S. T., A. Gandhi, and P. A. Haile (2013): "Connected Substitutes and Invertibility of Demand," *Econometrica*, 81, 2087–2111.

Berry, S. T. and P. A. Haile (2014): "Identification in Differentiated Products Markets Using Market Level Data," *Econometrica*, 82, 1749–1797.

——— (2018): "Nonparametric Identification of Simultaneous Equations Models with a Residual Index Structure," *Econometrica*, 86, 289–315.

——— (2021): "Foundations of Demand Estimation," in *Handbook of Industrial Organization*, ed. by K. Ho, A. Hortaçsu, and A. Lizzeri, Elsevier.

——— (2024): "Nonparametric Identification of Differentiated Products Demand Using Micro Data," *Econometrica*, 92, 1135–1162.

Bundorf, K., J. Levin, and N. Mahoney (2012): "Pricing and Welfare in Health Plan Choice," *American Economic Review*, 102, 3214–3248.

Cantoni, E. and V. Pons (2021a): "Does Context Outweigh Individual Characteristics in Driving Voting Behavior? Evidence from Relocations within the US," *American Economic Review (Forthcoming)*.

——— (2021b): "Strict ID Laws Don't Stop Voters: Evidence From a US Nationwide Panel, 2008–2018," *The Quarterly Journal of Economics*, 136, 2615–2660.

Cantoni, E., V. Pons, and J. Schäfer (2024): "Voting Rules, Turnout, and Economic Policies," *Annual Review of Economics*, 17, 217–239.

Cao, J., S.-y. S. Kim, and R. M. Alvarez (2022): "Bayesian Analysis of State Voter Registration Database Integrity," *Statistics, Politics and Policy*.

Chernozhukov, V. and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

Coate, S. and M. Conlin (2004): "A Group Rule-Utilitarian Approach to Voter Turnout: Theory and Evidence," *American Economic Review*, 94, 1476–1504.

Coate, S., M. Conlin, and A. Moro (2008): "The Performance of Pivotal-Voter Models in Small-Scale Elections: Evidence from Texas Liquor Referenda," *Journal of Public Economics*, 92, 582–596.

Coate, S. and B. Knight (2007): "Socially Optimal Districting: A Theoretical and Empirical Exploration," *Quarterly Journal of Economics*, 122, 1409–1471.

Cox, C. (2024): "The Equilibrium Effects of Campaign Finance Deregulation on US Elections," *Working Paper*.

Degan, A. and A. Merlo (2011): "A Structural Model of Turnout and Voting in Multiple Elections," *Journal of the European Economic Association*, 9, 209–245.

Downs, A. (1957): *An Economic Theory of Democracy*, New York: Harper & Row.

Eguia, J. X. (2022): "A Measure of Partisan Advantage in Redistricting," *Election Law Journal: Rules, Politics, and Policy*, 21, 84–103.

Einav, L., A. Finkelstein, and M. R. Cullen (2010): "Estimating Welfare in Insurance Markets Using Variation In Price," *The Quarterly Journal of Economics*, 125, 877–921.

Erikson, R. S. (1988): "The Puzzle of Midterm Loss," *The Journal of Politics*, 50, 1011–1028.

Fraga, B. L. and M. G. Miller (2021): "Who Does Voter ID Keep From Voting?" *Journal of Politics*.

Gelman, A. and G. King (1994): "A Unified Method of Evaluating Electoral Systems and Redistricting Plans," *American Journal of Political Science*, 38, 514–554.

Gerber, A. (1998): "Estimating the Effect of Campaign Spending on Senate Election Outcomes Using Instrumental Variables," *The American Political Science Review*, 92, 401–411.

Gillen, B. J., H. R. Moon, S. Montero, and M. Shum (2019): "BLP-Lasso for Aggregate Discrete Choice Models of Elections with Rich Demographic Covariates," *The Econometrics Journal*, 22, 262—281.

Glasgow, G. (2001): "Mixed Logit Models for Multiparty Elections," *Political Analysis*, 9, 116–136.

Gordon, B. and W. R. Hartmann (2013): "Advertising Effects in Presidential Elections," *Marketing Science*, 32, 19–35.

Grofman, B. and G. King (2007): "The Future of Partisan Symmetry as a Judicial Test for Partisan Gerrymandering after LULAC v. Perry," *Election Law Journal*, 6.

Huang, Y. and M. He (2021): "Structural Analysis of Tullock Contests with an Application to US House of Representatives Elections," *International Economic Review*, 62, 1011–1054.

Iaryczower, M., S. Montero, and G. Kim (2022): "Representation Failure," Tech. rep., National Bureau of Economic Research.

Jessee, S. A. (2010): "Partisan Bias, Political Information and Spatial Voting in the 2008 Presidential Election," *The Journal of Politics*, 72, 327–340.

Katz, J. N., G. King, and E. Rosenblatt (2020a): "Theoretical Foundations and Empirical Evaluations of Partisan Fairness in District-Based Democracies," *American Political Science Review*, 114, 164–178.

——— (2020b): "Theoretical Foundations and Empirical Evaluations of Partisan Fairness in District-Based Democracies," *American Political Science Review*, 114, 164–178.

Kawai, K., Y. Toyama, and Y. Watanabe (2021): "Voter Turnout and Preference Aggregation," *American Economic Journals: Microeconomics*, 13, 548–586.

Kim, S.-y. S. and B. Fraga (2022): "When Do Voter Files Accurately Measure Turnout? How Transitory Voter File Anapshots Impact Research and Representation," working paper, American University.

King, G. (1990): "Electoral Responsiveness and Partisan Bias in Multiparty Democracies," *Legislative Studies Quarterly*, 15, 159—-181.

King, G. and R. X. Browning (1987): "Democratic Representation and Partisan Bias in Congressional Elections," *American Political Science Review*, 81, 1252—-1273.

Knight, B. (2017): "An Econometric Evaluation of Competing Explanations for the Midterm Gap," *Quarterly Journal of Political Science*, 12, 205–239.

Lehmann, E. L. and H. Scheffe (1950): "Completeness, Similar Regions, and Unbiased Estimation," *Sankhya*, 10, 305–340.

Li, Q., M. J. Pomante, II, and S. Schraufnagel (2018): "Cost of Voting in the American States," *Election Law Journal: Rules, Politics, and Policy*, 17, 234—-247.

Longuet-Marx, N. (2025): "Party Lines or Voter Preferences? Explaining Political Realignment," Tech. rep., Columbia University.

Matzkin, R. L. (2008): "Identification in Nonparametric Simultaneous Equations," *Econometrica*, 76, 945–978.

——— (2015): "Estimation of Nonparametric Models with Simultaneity," *Econometrica*, 83, 1–66.

Merlo, A. and Á. de Paula (2017): "Identification and Estimation of Preference Distributions when Voters are Ideological," *The Review of Economic Studies*, 84, 1238–1263.

Myerson, R. B. and R. J. Weber (1982): "A Theory of Voting Equilibria," *American Political Science Review*, 87, 102–114.

Nevo, A. (2001): "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69, 307–42.

Newey, W. K. and J. L. Powell (2003): "Instrumental Variable Estimation in Nonparametric Models," *Econometrica*, 71, 1565–1578.

Niemi, R. G. and J. Deegan, Jr. (1978): "A Theory of Political Districting," *The American Political Science Review*, 72, 1304–1323.

Pomante, II, M. J. (2025): "Cost of Voting in the American States: 2024," *Election Law Journal: Rules, Politics, and Policy*, 24, 2—-98.

Riker, W. H. and P. C. Ordeshook (1968): "A Theory of the Calculus of Voting," *American Political Science Review*, 62, 25—-42.

Shachar, R. and B. Nalebuff (1999): "Follow the Leader: Theory and Evidence on Political Participation," *American Economic Review*, 89, 525–547.

Snyder, Jr., J. M. and D. Strömberg (2010): "Press Coverage and Political Accountability," *Journal of Political Economy*, 118, 355–408.

Spenkuch, J. L. and D. Toniatti (2018): "Political Advertising and Election Results," *The Quarterly Journal of Economics*, 133, 1981–2036.

Thompson, D., J. Wu, J. Yoder, and A. Hall (2020): "Universal Vote-by-Mail Has No Impact on Partisan Turnout or Vote Share," *Proceedings of the National Academy of Sciences*, 117, 14052—-14056.

Thompson, T. S. (1989): "Identification of Semiparametric Discrete Choice Models," Discussion paper no. 249, Center for Economics Research, Department of Economics, University of Minnesota.

Tufte, E. R. (1973): "The Relationship Between Seats and Votes in Two-Party Systems," *American Political Science Review*, 67, 540—-554.

Ujhelyi, G., S. Chatterjee, and A. Szabó (2021): "None of the Above: Protest Voting in the World's Largest Democracy," *Journal of the European Economic Association*, 19, 1936–1979.

Wang, Y., M. Lewis, and D. A. Schweidel (2018): "A Border Strategy Analysis of Ad Source and Message Tone in Senatorial Campaigns," *Marketing Science*.

Appendices

A Nonparametric Identification

In this appendix we provide conditions sufficient for nonparametric identification of the key features of interest discussed in the text. For clarity we will (in this appendix only) use uppercase to denote random variables, with lowercase representing particular realizations.[46] Because we rely on results from Berry and Haile (2024) we focus initially on the case (analogous to theirs) in which one observes and instruments for all endogenous contest-level variables y_{dt}. This allows us to cite their results directly at key steps and leads to results for identification of the voting choice functions σ_j in (8)–(10). The extension to identification of the functions $\tilde{\sigma}_j$ in 19 is then straightforward.

The observables are, for each contest dt:

- z_{idt} and turnout for all registered voters;
- x_{dt}, y_{dt};
- w_{dt}, a set of excluded instruments for y_{dt};
- $\mathcal{P}_{dt}$, the set of precincts p;
- vote shares $s_{pdt} = (s_{1pdt}, s_{2pdt})$ and $s_{dt} = (s_{1dt}, s_{2dt})$ at the precinct and district level, respectively; and
- the distributions ζ_{pt} and ζ_{dt}, of Z_{idt} for each precinct p and district d, respectively.

We assume throughout that Z_{idt} has at least two components. When Z_{idt} has more than two components, the "extra" components can be treated fully flexibly.[47] Thus, we henceforth condition on any such extra components, suppress them from the notation, and let Z_{idt} now denote the two-dimensional voter-level observables. Let $\mathcal{Z}_{dt} = \operatorname{supp} Z_{idt}|dt$ and $\mathcal{Y} = \operatorname{supp} Y_{dt}$.

A.1 Index Structure

The following assumptions introduce the index structure.[48]

[46] Recall that Ξ and Λ are uppercase versions of ξ and λ, respectively. To avoid confusion with the expectations operator E, we use $\mathcal{E}$ as uppercase ϵ.

[47] Although these extra components are not required for identification, in practice their variation will contribute to the precision of estimates.

[48] Our empirical model provides an example satisfying this structure. See also the examples and discussion in Berry and Haile (2024).

Assumption 1. $B\left(Z_{idt}, X_{dt}, Y_{dt}, \Xi_{Bdt}, \mathcal{E}_{iBdt}\right) = B\left(\gamma_B\left(Z_{idt}, X_{dt}, \Xi_{Bdt}\right), X_{dt}, Y_{dt}, \mathcal{E}_{iBdt}\right)$, *with* $\gamma_B\left(Z_{idt}, X_{dt}, \Xi_{Bdt}\right) = g_B(Z_{idt}, X_{dt}) + \Xi_{Bdt}$.

Assumption 2. $C\left(Z_{idt}, X_{dt}, Y_{dt}, \Xi_{Cdt}, \mathcal{E}_{iCdt}\right) = C\left(\gamma_C\left(Z_{idt}, X_{dt}, \Xi_{Cdt}\right), X_{dt}, Y_{dt}, \mathcal{E}_{iCdt}\right)$, *with* $\gamma_C\left(Z_{idt}, X_{dt}, \Xi_{Cdt}\right) = g_C\left(Z_{idt}, X_{dt}\right) + \Xi_{Cdt}$.

Assumption 3. *B is strictly increasing in $\gamma_B\left(Z_{idt}, X_{dt}, \Xi_{Bdt}\right)$; C is strictly increasing in $\gamma_C\left(Z_{idt}, X_{dt}, \Xi_{Cdt}\right)$*

Assumptions 1 and 2 are nonparametric index restrictions. Assumption 1, for example, requires that voter characteristics Z_{idt} and the shock Ξ_{Bdt} affect B_{idt} only through an index that excludes the endogenous characteristics Y_{dt}. Assumptions 2 places the same type of structure on C_{idt}. An important implication is that, in terms of voter behavior, variation in voter characteristics Z_{idt} can compensate for contest-level variation in unobservables Ξ_{Bdt} and Ξ_{Cdt}. In particular, fixing (X_{dt}, Y_{dt}), any change in Ξ_{kdt} could be offset (in terms of conditional vote shares) by an equal-sized change in $g_k\left(Z_{idt}, X_{dt}\right)$. Assumption 3 specifies that the latent shocks Ξ_{Bdt} and Ξ_{Cdt} can be interpreted as "vertical" characteristics for voting options 1 and 0, respectively. That is, a higher value of ξ_{Bdt} makes candidate 1 more attractive for all voters in contest dt, while a higher value of ξ_{Cdt} makes nonparticipation more attractive.[49] By Assumptions 1 and 2, the same monotonicity then holds with respect to $g_B(Z_{idt}, X_{dt})$ and $g_C\left(Z_{idt}, X_{dt}\right)$, respectively.

For simplicity, we henceforth condition on X_{dt} and suppress it from the notation, treating it fully flexibly.[50] Let $\gamma(Z_{idt}, \Xi_{dt})$ represent the index vector $(\gamma_B(Z_{idt}, \Xi_{dt}), \gamma_C(Z_{idt}, \Xi_{dt}))$. Given $\gamma(Z_{idt}, \Xi_{dt}) = \gamma$ and $Y_{dt} = y$, vote shares can then be written as

$$\sigma_1(\gamma, y) = \int 1\left\{B\left(\gamma_B, y, \epsilon_B\right) > \max\left\{0, C\left(\gamma_C, y, \epsilon_C\right)\right\}\right\} dF\left(\epsilon_B, \epsilon_C\right) \tag{A.1}$$

$$\sigma_2(\gamma, y) = \int 1\left\{-B\left(\gamma_B, y, \epsilon_B\right) > \max\left\{0, C\left(\gamma_C, y, \epsilon_C\right)\right\}\right\} dF\left(\epsilon_B, \epsilon_C\right) \tag{A.2}$$

$$\sigma_0\left(\gamma, y\right) = 1 - \sigma_1\left(\gamma, y\right) - \sigma_2\left(\gamma, y\right). \tag{A.3}$$

[49] Note that an increase in the value of Ξ_{Cdt} can be interpreted as (1) raising voting cost, (2) reducing the perceived likelihood of pivotality, or (3) reducing the intensity of preference between the candidates.

[50] Formally the remainder of the discussion is to be interpreted conditional on an arbitrary value of X_{dt}, and can be repeated at all such values.

A.2 Injectivity

Define the set of interior choice probability vectors

$$\Delta^* = \{(s_1, s_2) : s_1 > 0, s_2 > 0, 1 - s_1 - s_2 > 0\}.$$

For $y \in \mathcal{Y}$, let $\mathcal{G}_y^*$ denote the pre-image of Δ^* under $\sigma(\cdot, y)$. We will demonstrate, for all y, injectivity of $\sigma(\cdot, y)$ on $\mathcal{G}_y^*$ under the following assumptions.

Assumption 4. *Conditional on any $y \in \mathcal{Y}$ and $\gamma \in \mathcal{G}_y^*$, (B_{idt}, C_{idt}) are continuously distributed and have support with non-empty convex interior $\mathcal{I}(\gamma, y)$.*

Assumption 5. *Conditional on any $y \in \mathcal{Y}$ and $\gamma \in \mathcal{G}_y^*$,*
$\Pr(B_{idt} > C_{idt} \mid B_{idt} > 0, \gamma(Z_{idt}, \Xi_{Bdt}) = \gamma, Y_{dt} = y) < 1$, *and*
$\Pr(-B_{idt} > C_{idt} \mid B_{idt} < 0, \gamma(Z_{idt}, \Xi_{Bdt}) = \gamma, Y_{dt} = y) < 1$.

Assumption 4 requires continuously distributed (B_{idt}, C_{idt}) conditional on all values of $(Z_{idt}, \Xi_{Bdt}, \Xi_{Cdt})$. It implies that in every district and at all values of Z_{idt}, there are voters on each margin of indifference. This is a type of non-degeneracy condition, as is Assumption 5, which requires that the probability of not voting be nonzero among voters who prefer candidate 1 and among those who prefer candidate 2. In typical parametric models, both conditions are implied by the presence of idiosyncratic choice-specific taste shocks with full support, as in multinomial probit or logit models.

For $y \in \mathcal{Y}$, let $\mathcal{A}(\gamma, y)$ denote the pre-image of $\mathcal{I}(\gamma, y)$ under the mapping $(B(\gamma_B, y, \cdot), C(\gamma_C, y, \cdot))$. Our injectivity result is given in Lemma 3 below. It relies on two preliminary results. Because Lemma 2 follows from argument analogous to that given for Lemma 1, we omit its proof.

Lemma 1. *Let Assumptions 1–5 hold. For all $y \in \mathcal{Y}$ and $\gamma \in \mathcal{G}_y^*$, $\sigma_1(\gamma, y)$ is strictly increasing in γ_B and strictly decreasing in γ_C.*

Proof. Take $y \in \mathcal{Y}$ and $\gamma \in \mathcal{G}_y^*$. From (A.1) and Assumption 3, it is immediate that $\sigma_1(\gamma, y)$ is weakly increasing in γ_B and weakly decreasing in γ_C. By Assumption 4, both monotonicity properties will be strict if there exists $(\epsilon_B, \epsilon_C) \in \mathcal{A}(\gamma, y)$ such that $B(\gamma_B, y, \epsilon_B) > 0$ and $B(\gamma_B, y, \epsilon_B) = C(\gamma_C, y, \epsilon_C)$. Proceeding by contradiction, suppose first that $B(\gamma_B, y, \epsilon_B) \leq 0$ for all $(\epsilon_B, \epsilon_C) \in \mathcal{A}(\gamma, y)$. Then we would have $\sigma_1(\gamma, y) = 0$, contradicting $\gamma \in \mathcal{G}_y^*$. So suppose instead that for all $(\epsilon_B, \epsilon_C) \in \mathcal{A}(\gamma, y)$, $B(\gamma_B, y, \epsilon_B) < C(\gamma_C, y, \epsilon_C)$ whenever $B(\gamma_B, y, \epsilon_B) > 0$. Then we would again have $\sigma_1(\gamma, y) = 0$. Finally, if for all $(\epsilon_B, \epsilon_C) \in \mathcal{A}(\gamma, y)$ we had $B(\gamma_B, y, \epsilon_B) > C(\gamma_C, y, \epsilon_C)$ whenever $B(\gamma_B, y, \epsilon_B) > 0$, Assumption 5 would be violated. □

Lemma 2. *Let Assumptions 1–5 hold. For all $y \in \mathcal{Y}$ and $\gamma \in \mathcal{G}_y^*$, $\sigma_2(\gamma, y)$ is strictly decreasing in γ_B and in γ_C.*

Lemma 3. *Let Assumptions 1–5 hold. For all $y \in \mathcal{Y}$, $\sigma(\cdot, y)$ is injective on $\mathcal{G}_y^*$.*

Proof. Take $y \in \mathcal{Y}$. Proceeding by contradiction, suppose that for distinct $\tilde{\gamma}$ and γ in $\mathcal{G}_y^*$ we had $\sigma(\tilde{\gamma}, y) = \sigma(\gamma, y)$. By Lemma 1, $\sigma_1(\tilde{\gamma}, y) = \sigma_1(\gamma, y)$ requires that the differences $\tilde{\gamma}_B - \gamma_B$ and $\tilde{\gamma}_C - \gamma_C$ have the same sign. But then by Lemma 2, $\sigma_2(\tilde{\gamma}, y) = \sigma_2(\gamma, y)$ could not hold. □

A.3 Micro Vote Shares

Let $s_{jdt}(\cdot)$ denote the vote share of option $j \in \{0, 1, 2\}$ in contest dt as a function of Z_{idt}. Although in every contest we observe $s_{0dt}(z)$ for all $z \in \mathcal{Z}_{dt}$, we do not observe $s_{1dt}(z)$ or $s_{2dt}(z)$ for any z. However, the panel structure of precincts within districts can allow identification of $s_{1dt}(z)$ and $s_{2dt}(z)$ on $\mathcal{Z}_{dt}$.

Assumption 6. *For each contest dt, the family of distributions $\{\zeta_{pt} : p \in \mathcal{P}_{dt}\}$ is boundedly complete with respect to ζ_{dt}.*

Assumption 6 is a standard bounded completeness condition, specifying what is meant by sufficiently rich cross-precinct variation in ζ_{pt}.[51] Without further restrictions, this is a demanding requirement, reflecting the need to discriminate between all (even arbitrarily similar) nonparmetric functions. But it is also standard in a range of other contexts. For example, it is the notion of instrument "relevance" needed for nonparametric identification in separable regression models with bounded regression functions (Newey and Powell (2003)). If any elements of Z_{idt} are continuous, this completeness condition requires a continuum of precincts and should, obviously, be thought of as an approximation. With discrete Z_{idt}, (bounded) completeness is a full rank assumption on a matrix of conditional probabilities $\rho_{pt}(z_k)$, where k indexes the points in $\mathcal{Z}_{dt}$ (Newey and Powell (2003)).

Lemma 4. *Under Assumption 6, the functions $s_{1dt}(\cdot)$ and $s_{2dt}(\cdot)$ are identified for all d.*

[51] See, e.g, Lehmann and Scheffe (1950), Newey and Powell (2003), Chernozhukov and Hansen (2005), and Andrews (2017). A necessary and sufficient condition for (bounded) completeness is that if for all (bounded) functions $\phi : \mathcal{Z}_{dt} \to \mathbb{R}$ we have $\mathbb{E}_z[\varphi(z)|p] = 0$ for almost all $p \in \mathcal{P}_{dt}$, then $\varphi(z) = 0$ a.s.–ζ_{dt}. Of course, bounded completeness is weaker than completeness.

Proof. Proof. Take arbitrary $j \in \{1, 2\}$. By definition,

$$s_{jpdt} = \int s_{jdt}(z)\, d\zeta_p(z) = \mathbb{E}_{Z_{idt}}\left[s_{jdt}(Z_{idt}) \,|p\right]. \tag{A.4}$$

This implies

$$\mathbb{E}_{Z_{idt}}\left[s_{jpdt} - s_{jt}(Z_{idt}) \,|p\right] = 0 \quad \forall p \in \mathcal{P}_{dt}. \tag{A.5}$$

Noting that shares are bounded, identification of $s_{jdt}(\cdot)$ follows from Assumption 6. □

Note that his result implies that the mapping $s_{0dt}(\cdot)$ is identified for each contest dt without using the fact that the conditional shares $s_{0dt}(z)$ are directly observed. This provides a strong overidentifying restriction and emphasizes a sense in which this result is stronger than necessary in our context. In practice, particularly with the addition of parametric structure, the observability of $s_{0dt}(\cdot)$ (and the identity A.3) will play a substantial role in pinning down $s_{1dt}(\cdot)$ and $s_{12dt}(\cdot)$. However, the formal result makes clear that the panel structure of precincts-within-districts is also powerful.

A.4 Voting Choice Functions

Lemmas 3 and 4 address the two fundamental distinctions (see p. 21) between the voting model and the class of demand models with micro data considered by Berry and Haile (2024). In particular, we can treat the micro-level choice probabilities $s_{jdt}(z)$ as observed, and the conditional choice probability mapping can be inverted. Identification of the conditional vote probability mappings $\sigma_1(\cdot)$ and $\sigma_2(\cdot)$ then follows from the results of Berry and Haile (2024) under the following additional assumptions.

Assumption 7. *$\mathcal{Z}_{dt} = \mathcal{Z}$.*

Assumption 8. *$g(\cdot)$ is injective on $\mathcal{Z}$.*

Assumption 9. *The sets $\mathcal{Z}$ and $\text{supp}\,\Xi_{dt}|Y_{dt}$ are open and connected.*

Assumption 10. *(i) $g(\cdot)$ is uniformly continuous and continuously differentiable; (ii) $\sigma(\cdot)$ is continuously differentiable; (iii) $\partial g(z)/\partial z$ and $\partial\sigma(\gamma)/\partial\gamma$ are nonsingular almost surely on $\mathcal{Z}$ and $\mathcal{G}$, respectively.*

Assumption 11. *(i) $\mathbb{E}\left[\Xi_{\ell dt}|X_t, W_t\right] = 0$ almost surely for $\ell = B, C$;*
(ii) In the class of functions $\Psi(X_t, Y_t)$ with finite expectation,
$\mathbb{E}\left[\Psi(X_t, Y_t)\,|X_t, W_t\right] = 0$ almost surely implies $\Psi(X_t, Y_t) = 0$ almost surely.

Recalling that we have conditioned on X_{dt}, Assumption 7 specifies that $\operatorname{supp} Z_{idt}$ is the same in all districts conditional on X_d. Because we do not restrict variation in the observed distributions ζ_{dt} across contests, this is not very restrictive. And, as discussed in Berry and Haile (2024), this condition can be relaxed at the cost of expositional clarity. Assumption 8 adds to the index structure a requirement that, given fixed values of (Ξ_{Bdt}, Ξ_{Cdt}), distinct values of Z_{idt} map to different values of the index. Assumptions 9 and 10 require continuously distributed Z_{idt} and technical conditions that together facilitate the use of calculus and continuity arguments in Berry and Haile (2024). We refer readers to Berry and Haile (2024) for additional discussion of these three assumptions, including how these and other assumptions exploited here may be relaxed. Theorem 2 in Berry and Haile (2024) then yields the following identification result.

Theorem 1. *Under Assumptions 1–11, the index mapping $g(\cdot)$, the voting choice functions σ_j for $j = 0, 1, 2$, and the values of (ξ_{Bdt}, ξ_{Cdt}) for all contests dt are identified.*

Extending this result to identification of the functions $\tilde{\sigma}_j$ follows the same argument. These functions take the same form as the functions σ_j with three alterations: (a) X_{dt} is now interpreted to include the instruments W_{dt}; (b) the endogenous Y_{dt} are dropped as arguments; and (c) the structural errors Ξ_{dt} are replaced by their reduced-form analogs Λ_{dt}. After making these substitutions above, the identification argument carries through without change. Of course, when Y_{dt} is dropped, the IV relevance condition (part (ii) of Assumption 11) holds trivially.

A.5 Joint Distribution of (B_{idt}, C_{idt})

Here we drop our conditioning on X_{dt} for clarity and provide additional conditions allowing us to obtain partial or full identification of the joint density of (B_{idt}, C_{idt}) conditional on either $(Z_{idt}, X_{dt}, Y_{dt}, \Xi_{dt})$ or $(Z_{idt}, X_{dt}, W_{dt}, \Lambda_{dt})$. To avoid additional notation, we will focus on the former conditioning set; however, the argument in either case is the same.

Suppose that

$$B_{idt} = \gamma_B(Z_{idt}, X_{dt}, Y_{dt}, \Xi_{Bdt}) + \mu_{Bid} \tag{A.6}$$

$$C_{idt} = \gamma_C(Z_{idt}, X_{dt}, Y_{dt}, \Xi_{Cdt}) + \mu_{Cid}, \tag{A.7}$$

where (μ_{Bid}, μ_{Cid}) are independent of (Z_{idt}, Ξ_{dt}) and have joint distribution $F_\mu(\mu_{Bid}, \mu_{Cid}|X_{dt}, Y_{dt})$ conditional on (X_{dt}, Y_{dt}). This is an additional nonparametric restriction on the functional forms of B_{idt} and C_{idt}. Note

that this specification permits random coefficients on X_{dt} and Y_{dt}.[52] Let $f_\mu(\mu_{Bid}, \mu_{Cid}|X_{dt}, Y_{dt})$ denote the conditional density associated with $F_\mu(\mu_{Bid}, \mu_{Cid}|X_{dt}, Y_{dt})$. With the functions σ_j known, Given Theorem 1 and the structure (8)–(10), we may treat the index functions and the realizations of the structural errors (ξ_{Bdt}, ξ_{Cdt}) as known.[53] For this paragraph, we will fix $(X_{dt}, Y_{dt}) = (x, y)$ and suppress these arguments in the notation. The voting choice function for candidate 1 takes the form

$$\sigma_1(\gamma) = \int_{-\gamma_B}^{\infty} \int_{-\infty}^{\gamma_B+\mu_{iB}-\gamma_C} f_\mu(\mu_{iB}, \mu_{iC})\, d\mu_{iC}\, d\mu_{iB} \tag{A.8}$$

and

$$\sigma_2(\gamma) = \int_{-\infty}^{-\gamma_B} \int_{-\infty}^{-\gamma_B-\mu_{iB}-\gamma_C} f_\mu(\mu_{iB}, \mu_{iC})\ d\mu_{iC}\, d\mu_{iB}. \tag{A.9}$$

Taking directional derivatives in the direction $v = (1, -1)$, we obtain

$$\nabla_v\, \sigma_1(\gamma) = \int_{-\infty}^{-\gamma_c} f_\mu(-\gamma_B, \mu_{iC})\, d\mu_{iC} + 2 \int_{-\gamma_B}^{\infty} f_\mu(\mu_{iB}, \mu_{iB} + \gamma_B - \gamma_C)\, d\mu_{iB}$$

and

$$\nabla_v\, \sigma_2(\gamma) = -\int_{-\infty}^{-\gamma_c} f_\mu(-\gamma_B, \mu_{iC})\, d\mu_{iC}.$$

Given the identification results above, the left side of each expression is known. Summing the right side of these expressions yields

$$2 \int_{-\gamma_B}^{\infty} f_\mu(\mu_{iB}, \mu_{iB} + \gamma_B - \gamma_C)\, d\mu_{iB}.$$

Differentiating this expression in the direction of $\tilde{v} = (1, 1)$ yields

$$2 f_\mu(-\gamma_B, -\gamma_C).$$

[52] It is possible to also let F_μ depend on Z_{idt}, although such specifications are not typical in practice. In this case the variation used in the final step of the argument here would be only that created by variation in Ξ_{dt}.

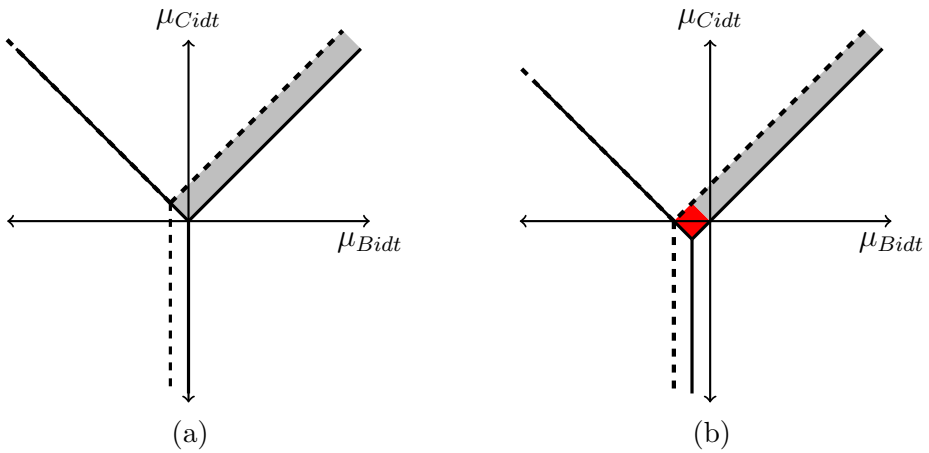
[53] The analysis in Berry and Haile (2024) employed a "rotation" normalization on the index vector that is without loss for their focus on demand, but which need not preserve the interpretation of indices appearing in a random utility specification that one might posit to generate this demand. Here, rotations of the index would be rotations of utilities; and when choice probabilities are defined by (8)–(10), rotations of the true model are no longer observationally equivalent; i.e., specifying the underlying structure as we do here imposes the "true rotation."

And although we conditioned on a particular value of (x, y), the same argument can be repeated at all values of (X_{dt}, Y_{dt}). This demonstrates the following result.

Theorem 2. *Let Assumptions 1–10 hold and suppose B_{idt} and C_{idt} take the forms (A.6) and (A.7), with $(\mu_{Bid}, \mu_{Cid}) \perp\!\!\!\perp (Z_{idt}, \Xi_{dt})$. Then the joint density $f_\mu(\cdot|X_{dt}, Y_{dt})$ is identified on the support of $(-\gamma_B(Z_{idt}, \Xi_{Bdt}), -\gamma_C(Z_{idt}, \Xi_{Cdt}))$ conditional $(Z_{idt}, X_{dt}, Y_{dt}, \Xi_{dt})$.*

Because the realizations of $\Xi_{dt} \equiv (\Xi_{Bdt}, \Xi_{Cdt})$ are already identified, variation in both Z_{idt} and Ξ_{dt} can provide the variation in the index vector that "traces out" the joint density $f(\cdot|X_{dt}, Y_{dt})$. Depending on the extent of this variation, Theorem 2 may yield identification of each conditional density on only some (rather than all) of its support. This is analogous to standard identification results for discrete choice models relying on special regressors with large support. Here we have no special regressors—there are no factors that alter only the utility of voting for candidate i; nonetheless, large support for $(-\gamma_B(Z_{idt}, \Xi_{Bdt}), -\gamma_C(Z_{idt}, \Xi_{Bdt}))$ conditional $(Z_{idt}, X_{dt}, Y_{dt}, \Xi_{dt})$ would deliver full rather than partial identification.

Figure 17



The origin is an initial value $(\hat{\gamma}_B, -\hat{\gamma}_C)$ of the index vector (γ_B, γ_C). Shaded areas represent changes in turnout resulting from a shifts in (γ_B, γ_C) in the direction $(1, -1)$ before (panel (a)) and after (panel (b)) a shift in the direction $(1, 1)$. Voters in red—those approximately at the origin—are responsible for the observed difference in turnout response across the two panels.

Figure 17 illustrates the argument. Here we fix a value of (X_{dt}, Y_{dt}) and suppress these in the notation. In panel (a) the solid "Y" shape partitions

voters into their choices at an arbitrary initial value $(\hat{\gamma}_B, \hat{\gamma}_C)$ of the index vector (γ_B, γ_C) (recall Figure 2). The dashed "Y" shows the partition after adding $(h, -h)$ to (γ_B, γ_C), with $h > 0$. The shaded area represents voters who turn out only after the change. Panel (b) shows the same comparative static starting from an index vector $(\hat{\gamma}_B+h, \hat{\gamma}_C+h)$. The shaded area again represents voters turning out only after the index vector shifts by $(h, -h)$. These voters include those shaded in panel (a) but also those in the red shaded area. As $h \to 0$, panel (a) represents

$$\nabla_v \, \sigma_1(\hat{\gamma}_B, \hat{\gamma}_C) + \nabla_v \, \sigma_2(\hat{\gamma}_B, \hat{\gamma}_C)$$

while the difference between the two shaded areas—i.e., the red region—represents

$$\nabla_v \, \nabla_{\tilde{v}} \, \left(\sigma_1(\hat{\gamma}_b, \hat{\gamma}_C) + \sigma_2(\hat{\gamma}_B, \hat{\gamma}_C)\right).$$

In this limit, the voters in red are those at the origin—i.e., those for whom $(\mu_{Bidt}, \mu_{Cidt}) = (-\hat{\gamma}_B, -\hat{\gamma}_C)$.

B Supplemental Appendix: Data

Table 13 summarizes the component measures used to construct our measures of state-level policies affecting voting costs.

Table 13: State Level Voting Policies

Variable	Mean	Std. Dev.	Min.	Max.
Voting inconvenience				
Absentee excuse req	0.378	0.488	0	1
No absentee in person	0.230	0.424	0	1
No state holiday (state empl.)	0.733	0.439	0	1
No early vote	0.514	0.503	0	1
No voting centers	0.784	0.414	0	1
No permanent absentee	0.811	0.394	0	1
No time off vote	0.419	0.497	0	1
No time off pay	0.595	0.494	0	1
ID requirements				
No voter id	0.351	0.481	0	1
Non strict id	0.270	0.447	0	1
Non strict photo	0.189	0.394	0	1
Strict id	0.068	0.253	0	1

The sample includes all state-years.

C Supplemental Appendix: Estimation

Here we provide additional details regarding our estimation procedure, including derivation of our quasi-likelihood. Recall key notation from the text:

- $\mathbb{I}_{pdt}$, the set of registered voters in precinct p
- $\mathbb{I}^A_{pdt}$, the subset of $\mathbb{I}_{pdt}$ who turn out, with $n^A_{pdt} = \left|\mathbb{I}^A_{pdt}\right|$
- $s_{ijdt} = 1\{i \text{ chooses option } j\}$, for $j \in \{0, 1, 2\}$
- $\sigma_j\left(z_{idt}; \theta_1, \delta\right) = E\left[s_{ijdt} | z_{idt}; \theta_1, \delta\right] = \Pr\left(s_{ijdt} = 1 | z_{idt}; \theta_1, \delta\right),$
- $\sigma^A_j\left(z_{idt}; \theta_1, \delta\right) = \sigma_j\left(z_{idt}; \theta_1, \delta\right) / \left(1 - \sigma_0\left(z_{idt}; \theta_1, \delta\right)\right)$, the probability of voting for j, conditional on turning out

- $\bar{s}_{1pdt} = \frac{1}{n^A_{pdt}} \sum_{i \in \mathbb{I}^A_{pdt}} s_{i1dt}$, the vote share (among actual votes) for candidate 1 in precinct p, with the associated random variable denoted $\bar{S}_{1pdt}$.

Let $\omega_{idt}(\theta_1, \delta)$ denote the binomial variance

$$var\left(s_{i1dt} | i \in \mathbb{I}^A_{pdt}, \theta_1, \delta\right) = \sigma^A_1\left(z_{idt}; \theta_1, \delta\right)\left(1 - \sigma^A_1\left(z_{idt}; \theta_1, \delta\right)\right).$$

Let

$$\mu_{1pdt}(\theta_1, \delta) = E\left[\bar{s}_{1pdt} | \mathbb{I}^A_{pdt}, \theta_1, \delta\right] = \frac{1}{n^A_{pdt}} \sum_{i \in \mathbb{I}^A_{pdt}} \sigma^A_1\left(z_{idt}; \theta_1, \delta\right)$$

denote expected vote share in precinct p, conditional on turnout. Let

$$\Omega_{pdt}(\theta_1, \delta) = \frac{1}{n^A_{pdt}} \sum_{i \in \mathbb{I}^A_{pdt}} \omega_{idt}(\theta_1, \delta),$$

which may be interpreted as the average variance of s_{i1pdt} across $i \in \mathbb{I}^A_{pdt}$. Let (θ^0_1, δ^0) denote the true values of the parameter θ_1 and the contest-level fixed effects.

As described in the text, the first-step likelihood takes the form

$$L(\theta_1, \delta) = \prod_t \prod_d \prod_p L^0\left(\mathbb{I}^A_{pdt}; \theta_1, \delta\right) \times L^1\left(\bar{s}_{1pdt} | \mathbb{I}^A_{pdt}; \theta_1, \delta\right)$$

where

$$L^0\left(\mathbb{I}^A_{pdt}; \theta_1, \delta\right) = \prod_{i \in \mathbb{I}_{pdt}} \sigma_0\left(z_{idt}; \theta_1, \delta\right)^{s_{i0dt}} \left(1 - \sigma_0\left(z_i; \theta_1, \delta\right)\right)^{1 - s_{i0dt}}$$

and

$$L^1_{pdt}\left(\bar{s}_{1pdt} | \mathbb{I}^A_{pdt}; \theta_1, \delta\right) = \sum_{\substack{\mathcal{I} \subset \mathbb{I}^A_{pdt}: \\ |\mathcal{I}| = \bar{s}_{1pdt} \times n^A_{pdt}}} \left(\prod_{i \in \mathcal{I}} \sigma^A_1\left(z_i; \theta_1, \delta\right) \prod_{i' \in \{\mathbb{I}^A_{pdt} - \mathcal{I}\}} \sigma^A_2\left(z_{i'dt}; \theta_1, \delta\right) \right).$$

To address the computational infeasibility of this likelihood (see the text), we follow Ainsworth (2020) in exploiting an approximation to each term L^1_{pdt}. This approximation is based on the fact that, by an appropriate central limit theorem (e.g., Lyapunov),

$$\sqrt{n^A_{pd}}\left(\bar{s}_{1pd} - \mu_{1pd}(\theta^0_1, \delta^0)\right) \to N\left(0, \Omega_{pd}(\theta^0_1, \delta^0)\right).$$

The normal approximation is known to be very good, even for moderate n^A_{pdt}, when μ_{1pdt} is not too close to 0 or 1.

Using this fact, we approximate the log-likelihood with the log-quasi-likelihood

$$\tilde{\mathcal{L}}(\theta_1, \delta) = \sum_t \sum_d \sum_p \left\{ \mathcal{L}^0_{pdt}\left(\mathbb{I}^A_{pdt}; \theta_1, \delta\right) + \tilde{\mathcal{L}}^1_{pdt}\left(\bar{s}_{1pdt} | \mathbb{I}^A_{pdt}; \theta_1, \delta\right) \right\},$$

where $\mathcal{L}^0_{pdt}\left(\mathbb{I}^A_{pdt}; \theta_1, \delta\right) = \ln L^0_{pdt}\left(\mathbb{I}^A_{pdt}; \theta_1, \delta\right)$ and (letting ϕ denote the standard normal pdf)

$$\begin{aligned}
\tilde{\mathcal{L}}^1_{pdt}\left(\bar{s}_{1pdt} | \mathbb{I}^A_{pdt}; \theta_1, \delta\right) &= \ln \phi \left(\frac{\bar{s}_{1pdt} - \mu_{1pdt}(\theta_1, \delta)}{\sqrt{\Omega_{pdt}(\theta_1, \delta)/n^A_{pdt}}} \right) \\
&= \ln \left[\frac{1}{\sqrt{2\pi \Omega_{pdt}(\theta_1, \delta)/n^A_{pdt}}} \exp\left(-\frac{\left(\bar{s}_{1pdt} - \mu_{1pdt}(\theta_1, \delta)\right)^2}{2\Omega_{pdt}(\theta_1, \delta)/n^A_{pdt}} \right) \right] \\
&= \left[-\ln\left(\sqrt{2\pi/n^A_{pdt}}\right) - \ln \sqrt{\Omega_{pdt}(\theta_1, \delta)} - \frac{n^A_{pdt}}{2} \frac{\left(\bar{s}_{1pdt} - \mu_{1pdt}(\theta_1, \delta)\right)^2}{\Omega_{pdt}(\theta_1, \delta)} \right].
\end{aligned}$$

Using this quasi-likelihood, we estimate θ_1 using interior point minimization with a known gradient. As noted in the text, we compute the contest-level fixed effects $\delta(\theta_1)$ in a nested fixed point algorithm, matching contest-level turnout and vote shares exactly, following typical practice in the analogous demand estimation setting (Berry, Levinsohn, and Pakes (1995, 2004)). Let $\bar{s}_{jdt}$ denote the observed choice share for option j in contest dt, and let $\mu_j(\theta_1, \delta)$ denote the share predicted by the model. At each candiate value of θ_1, we solve the system of equations $\bar{s}_{jdt} - \mu_j(\theta_1, \delta) = 0 \quad \forall (j, d, t)$ using Newton's method. Our estimator of θ_1 is thus the solution to

$$\min_{\theta_1} \; -\tilde{\mathcal{L}}(\theta_1, \delta) \;\; s.t. \;\; \delta = \underset{\delta'}{\text{argsolv}}[\bar{s}_{jdt} - \mu_{jdt}(\theta_1, \delta') = 0 \; \forall j, d, t]$$

For inference, we use standard errors derived from the Generalized Method of Moments (GMM) equivalent of our approach. Let ι_{dt} denote the exogenous variables "instruments" (x_{dt}, w_{dt}) entering the reduced forms. Define the moment vector for each contest dt as

$$g_{dt}(\theta_1, \theta_2) = \begin{bmatrix} \frac{\partial \mathcal{L}_{dt}(\theta_1, \delta(\theta_1))}{\partial \theta_1} \\ \iota^\top_{dt}(\delta_{Bdt} - \alpha_B - x_{dt}\gamma_{xB} - w_{dt}\gamma_{wB}) \\ \iota^\top_{dt}(\delta_{Cdt} - \alpha_C - x_{dt}\gamma_{xC} - w_{dt}\gamma_{wC}) \end{bmatrix},$$

where the first moment is the score vector of the contest dt log-likelihood with respect to the parameters θ_1, and other two moments are the least squares normal equations at the contest level.[54] We compute the optimal weighting matrix using the inverse of the asymptotic variance-covariance matrix for the moments. As described in the text, this covariance matrix treats the first-step and second-step parameters as independent, consistent with the much larger first-step sample size and our imposition of a perfect fit to district-level vote shares in the first step. The covariance structure for the second-step estimates is clustered at the district level, allowing arbitrary cross-district heteroskedasticity and dependence (over time and between the shocks to costs and benefits) within district.

[54]Whereas the full set of moments $g_{dt}(\theta_1, \theta_2)$ at the parameter estimates $(\hat{\theta}_1, \hat{\theta}_2)$ is used for inference, the optimally weighted stacked normal equations are also used for estimation of θ_2, as described in the text.

D Supplemental Appendix: Additional Tables Referenced in the Text

Table 14: Vote Shares Among 5% Marginal Voters

State	Republican Share	State	Republican Share
AK	0.331	MT	0.277
	[0.271, 0.387]		[0.217, 0.338]
AL	0.397	NC	0.291
	[0.344, 0.445]		[0.236, 0.344]
AR	0.414	ND	0.399
	[0.353, 0.471]		[0.328, 0.465]
AZ	0.276	NE	0.395
	[0.223, 0.327]		[0.334, 0.452]
CA	0.196	NH	0.254
	[0.154, 0.240]		[0.200, 0.308]
CO	0.238	NJ	0.223
	[0.183, 0.293]		[0.176, 0.270]
CT	0.201	NM	0.248
	[0.155, 0.246]		[0.200, 0.296]
DE	0.224	NV	0.258
	[0.179, 0.268]		[0.208, 0.308]
FL	0.288	NY	0.205
	[0.235, 0.340]		[0.165, 0.245]
GA	0.302	OH	0.312
	[0.251, 0.352]		[0.255, 0.366]
HI	0.137	OK	0.434
	[0.103, 0.172]		[0.371, 0.491]
IA	0.289	OR	0.236
	[0.231, 0.346]		[0.188, 0.284]
ID	0.349	PA	0.295
	[0.280, 0.416]		[0.243, 0.346]
IL	0.230	RI	0.196
	[0.185, 0.275]		[0.153, 0.240]
IN	0.342	SC	0.335
	[0.285, 0.396]		[0.282, 0.385]
KS	0.356	SD	0.422
	[0.297, 0.411]		[0.359, 0.479]
KY	0.428	TN	0.366
	[0.371, 0.481]		[0.308, 0.420]
LA	0.413	TX	0.310
	[0.355, 0.466]		[0.260, 0.358]
MA	0.160	UT	0.363
	[0.120, 0.202]		[0.298, 0.424]
MD	0.197	VA	0.262
	[0.157, 0.237]		[0.212, 0.311]
ME	0.250	VT	0.147
	[0.198, 0.302]		[0.111, 0.185]
MI	0.272	WA	0.236
	[0.222, 0.321]		[0.189, 0.285]
MN	0.225	WI	0.295
	[0.173, 0.280]		[0.243, 0.344]
MO	0.342	WV	0.418
	[0.286, 0.394]		[0.355, 0.474]
MS	0.356	WY	0.382
	[0.300, 0.407]		[0.308, 0.454]

Table 15: Voting Cost Counterfactual: Change in Shares

State	Baseline	High Cost	Low Cost	State	Baseline	High Cost	Low Cost
AK	0.554	0.586	0.544	MT	0.556	0.579	0.549
		[0.570, 0.603]	[0.538, 0.549]			[0.567, 0.591]	[0.543, 0.554]
AL	0.627	0.643	0.608	NC	0.520	0.551	0.518
		[0.636, 0.650]	[0.596, 0.620]			[0.537, 0.567]	[0.509, 0.528]
AR	0.678	0.697	0.660	ND	0.696	0.703	0.676
		[0.688, 0.707]	[0.651, 0.668]			[0.697, 0.710]	[0.666, 0.686]
AZ	0.491	0.501	0.469	NE	0.643	0.675	0.642
		[0.489, 0.512]	[0.458, 0.480]			[0.661, 0.691]	[0.639, 0.645]
CA	0.365	0.394	0.365	NH	0.463	0.486	0.454
		[0.378, 0.412]	[0.365, 0.365]			[0.474, 0.499]	[0.440, 0.470]
CO	0.470	0.492	0.464	NJ	0.402	0.434	0.402
		[0.474, 0.510]	[0.460, 0.468]			[0.418, 0.451]	[0.398, 0.406]
CT	0.379	0.402	0.370	NM	0.436	0.470	0.436
		[0.390, 0.415]	[0.356, 0.386]			[0.452, 0.489]	[0.435, 0.437]
DE	0.397	0.421	0.388	NV	0.467	0.503	0.467
		[0.409, 0.434]	[0.377, 0.399]			[0.485, 0.522]	[0.465, 0.468]
FL	0.508	0.525	0.492	NY	0.355	0.385	0.354
		[0.517, 0.533]	[0.483, 0.501]			[0.371, 0.401]	[0.348, 0.361]
GA	0.537	0.538	0.505	OH	0.552	0.561	0.528
		[0.531, 0.545]	[0.491, 0.519]			[0.555, 0.567]	[0.516, 0.539]
HI	0.251	0.267	0.241	OK	0.682	0.707	0.673
		[0.251, 0.283]	[0.236, 0.246]			[0.697, 0.719]	[0.667, 0.679]
IA	0.519	0.550	0.516	OR	0.424	0.457	0.424
		[0.535, 0.566]	[0.511, 0.522]			[0.437, 0.479]	[0.424, 0.424]
ID	0.669	0.682	0.655	PA	0.500	0.528	0.499
		[0.676, 0.689]	[0.646, 0.665]			[0.514, 0.544]	[0.487, 0.512]
IL	0.412	0.444	0.412	RI	0.343	0.360	0.327
		[0.427, 0.463]	[0.412, 0.412]			[0.351, 0.368]	[0.315, 0.339]
IN	0.577	0.578	0.541	SC	0.568	0.584	0.548
		[0.572, 0.583]	[0.525, 0.557]			[0.577, 0.593]	[0.533, 0.563]
KS	0.603	0.605	0.571	SD	0.672	0.691	0.656
		[0.593, 0.617]	[0.556, 0.586]			[0.680, 0.702]	[0.648, 0.663]
KY	0.642	0.667	0.631	TN	0.613	0.615	0.582
		[0.656, 0.680]	[0.620, 0.643]			[0.604, 0.625]	[0.568, 0.596]
LA	0.651	0.668	0.633	TX	0.528	0.552	0.516
		[0.661, 0.676]	[0.623, 0.642]			[0.540, 0.565]	[0.511, 0.521]
MA	0.292	0.319	0.291	UT	0.642	0.669	0.635
		[0.304, 0.336]	[0.284, 0.300]			[0.652, 0.686]	[0.631, 0.639]
MD	0.346	0.375	0.346	VA	0.473	0.482	0.453
		[0.360, 0.392]	[0.346, 0.347]			[0.478, 0.487]	[0.441, 0.464]
ME	0.444	0.477	0.443	VT	0.270	0.293	0.269
		[0.461, 0.494]	[0.435, 0.453]			[0.280, 0.306]	[0.262, 0.277]
MI	0.480	0.496	0.463	WA	0.429	0.453	0.422
		[0.489, 0.504]	[0.452, 0.473]			[0.434, 0.471]	[0.419, 0.426]
MN	0.470	0.495	0.469	WI	0.498	0.500	0.466
		[0.481, 0.510]	[0.465, 0.474]			[0.493, 0.506]	[0.451, 0.481]
MO	0.584	0.608	0.575	WV	0.646	0.683	0.643
		[0.598, 0.619]	[0.566, 0.584]			[0.665, 0.701]	[0.641, 0.644]
MS	0.589	0.589	0.551	WY	0.697	0.718	0.697
		[0.589, 0.589]	[0.531, 0.570]			[0.705, 0.730]	[0.695, 0.700]

Table 16: High Cost Counterfactual: Abstention Share

State	Baseline	High Cost	Low Cost
AK	0.374	0.426	0.357
		[0.410, 0.443]	[0.351, 0.362]
AL	0.359	0.384	0.321
		[0.371, 0.398]	[0.305, 0.337]
AR	0.339	0.369	0.311
		[0.353, 0.385]	[0.297, 0.325]
AZ	0.297	0.320	0.259
		[0.307, 0.334]	[0.248, 0.270]
CA	0.287	0.353	0.287
		[0.331, 0.375]	[0.283, 0.292]
CO	0.164	0.207	0.155
		[0.190, 0.226]	[0.152, 0.158]
CT	0.301	0.342	0.274
		[0.328, 0.356]	[0.257, 0.289]
DE	0.354	0.399	0.328
		[0.385, 0.413]	[0.315, 0.341]
FL	0.293	0.323	0.261
		[0.313, 0.333]	[0.250, 0.271]
GA	0.325	0.333	0.271
		[0.322, 0.344]	[0.255, 0.287]
HI	0.342	0.397	0.317
		[0.374, 0.422]	[0.309, 0.325]
IA	0.254	0.306	0.246
		[0.291, 0.322]	[0.241, 0.251]
ID	0.140	0.157	0.120
		[0.152, 0.163]	[0.113, 0.127]
IL	0.341	0.409	0.341
		[0.389, 0.432]	[0.339, 0.343]
IN	0.359	0.364	0.301
		[0.357, 0.370]	[0.284, 0.317]
KS	0.294	0.305	0.247
		[0.292, 0.317]	[0.232, 0.262]
KY	0.379	0.418	0.355
		[0.406, 0.432]	[0.342, 0.368]
LA	0.361	0.389	0.330
		[0.377, 0.401]	[0.318, 0.341]
MA	0.290	0.351	0.282
		[0.327, 0.378]	[0.267, 0.299]
MD	0.320	0.391	0.319
		[0.369, 0.413]	[0.318, 0.320]
ME	0.282	0.339	0.274
		[0.322, 0.357]	[0.266, 0.283]
MI	0.332	0.362	0.295
		[0.354, 0.371]	[0.281, 0.308]
MN	0.136	0.177	0.133
		[0.165, 0.191]	[0.130, 0.136]
MO	0.295	0.335	0.274
		[0.322, 0.347]	[0.262, 0.284]
MS	0.399	0.399	0.331
		[0.392, 0.405]	[0.309, 0.353]
MT	0.161	0.195	0.149
		[0.185, 0.206]	[0.144, 0.153]
NC	0.269	0.320	0.261
		[0.305, 0.337]	[0.251, 0.270]
ND	0.156	0.170	0.129
		[0.164, 0.176]	[0.122, 0.137]
NE	0.268	0.322	0.265
		[0.306, 0.338]	[0.260, 0.270]
NH	0.241	0.277	0.217
		[0.265, 0.289]	[0.202, 0.231]
NJ	0.320	0.383	0.316
		[0.365, 0.403]	[0.312, 0.320]
NM	0.327	0.394	0.326
		[0.374, 0.416]	[0.325, 0.327]
NV	0.324	0.389	0.322
		[0.369, 0.410]	[0.320, 0.324]
NY	0.390	0.455	0.383
		[0.435, 0.476]	[0.374, 0.392]
OH	0.290	0.307	0.249
		[0.301, 0.313]	[0.237, 0.260]
OK	0.303	0.344	0.286
		[0.331, 0.358]	[0.280, 0.293]
OR	0.288	0.359	0.288
		[0.334, 0.384]	[0.286, 0.291]
PA	0.261	0.309	0.249
		[0.293, 0.326]	[0.236, 0.262]
RI	0.372	0.405	0.333
		[0.393, 0.417]	[0.318, 0.347]
SC	0.350	0.375	0.308
		[0.367, 0.384]	[0.289, 0.327]
SD	0.295	0.327	0.269
		[0.313, 0.342]	[0.257, 0.281]
TN	0.318	0.328	0.270
		[0.316, 0.340]	[0.255, 0.285]
TX	0.362	0.408	0.342
		[0.392, 0.424]	[0.335, 0.349]
UT	0.223	0.269	0.213
		[0.251, 0.287]	[0.210, 0.217]
VA	0.273	0.291	0.230
		[0.284, 0.297]	[0.216, 0.244]
VT	0.258	0.316	0.250
		[0.297, 0.336]	[0.236, 0.265]
WA	0.250	0.304	0.239
		[0.283, 0.327]	[0.233, 0.244]
WI	0.358	0.364	0.297
		[0.355, 0.372]	[0.281, 0.313]
WV	0.380	0.439	0.375
		[0.420, 0.458]	[0.374, 0.377]
WY	0.065	0.089	0.063
		[0.081, 0.098]	[0.062, 0.065]

Table 17: Partisan Bias State Estimates

State	50-50	Smoothed 50-50
AK	0	0
AL	0.333	0.191
	[0.333, 0.333]	[0.175, 0.206]
AR	0.500	0.056
	[0.333, 0.500]	[0.000, 0.083]
AZ	-0.037	-0.086
	[-0.037, -0.037]	[-0.099, -0.074]
CA	-0.107	-0.071
	[-0.107, -0.082]	[-0.078, -0.059]
CO	0.048	0.048
	[0.048, 0.048]	[0.048, 0.048]
CT	-0.067	0.022
	[-0.200, -0.067]	[-0.022, 0.022]
DE	0	0
FL	0.086	0.004
	[0.086, 0.086]	[0.000, 0.012]
GA	0.238	0.175
	[0.238, 0.238]	[0.175, 0.183]
HI	0.000	0.000
	[0.000, 0.000]	[0.000, 0.000]
IA	-0.167	-0.028
	[-0.167, -0.167]	[-0.028, -0.028]
ID	0.000	0.000
	[0.000, 0.000]	[0.000, 0.000]
IL	-0.111	-0.080
	[-0.111, -0.111]	[-0.080, -0.080]
IN	0.333	0.148
	[0.333, 0.333]	[0.148, 0.148]
KS	0.000	0.083
	[0.000, 0.000]	[0.056, 0.083]
KY	0.333	0.185
	[0.333, 0.333]	[0.185, 0.204]
LA	0.667	0.333
	[0.667, 0.667]	[0.296, 0.333]
MA	-0.185	-0.049
	[-0.185, -0.185]	[-0.074, -0.049]
MD	-0.500	-0.222
	[-0.500, -0.500]	[-0.222, -0.222]
ME	0.000	0.056
	[0.000, 0.000]	[0.056, 0.056]
MI	0.143	0.103
	[0.143, 0.143]	[0.103, 0.103]
MN	0.333	0.111
	[0.333, 0.333]	[0.111, 0.111]
MO	0.333	0.208
	[0.250, 0.333]	[0.181, 0.208]
MS	0.500	0.278
	[0.500, 0.500]	[0.278, 0.306]

State	50-50	Smoothed 50-50
MT	0	0
NC	0.436	0.188
	[0.385, 0.436]	[0.162, 0.188]
ND	0	0
NE	-0.111	-0.074
	[-0.111, -0.111]	[-0.074, -0.037]
NH	0.000	0.000
	[0.000, 0.000]	[0.000, 0.000]
NJ	0.000	-0.009
	[0.000, 0.000]	[-0.009, -0.009]
NM	-0.333	-0.074
	[-0.333, -0.333]	[-0.074, -0.074]
NV	-0.333	-0.139
	[-0.333, -0.333]	[-0.139, -0.139]
NY	-0.012	-0.025
	[-0.012, -0.012]	[-0.025, -0.021]
OH	0.333	0.201
	[0.333, 0.333]	[0.201, 0.201]
OK	0.333	0.089
	[0.200, 0.333]	[0.044, 0.089]
OR	0.067	-0.022
	[0.067, 0.067]	[-0.044, -0.022]
PA	0.148	0.148
	[0.148, 0.185]	[0.142, 0.161]
RI	0.000	0.000
	[0.000, 0.000]	[0.000, 0.000]
SC	0.429	0.159
	[0.429, 0.429]	[0.159, 0.159]
SD	0	0
TN	0.482	0.259
	[0.482, 0.482]	[0.247, 0.259]
TX	0.037	-0.006
	[0.037, 0.056]	[-0.006, 0.003]
UT	0.000	-0.028
	[0.000, 0.000]	[-0.028, -0.028]
VA	0.152	0.071
	[0.152, 0.152]	[0.071, 0.071]
VT	0	0
WA	0.000	0.044
	[0.000, 0.000]	[0.033, 0.044]
WI	0.250	0.208
	[0.250, 0.250]	[0.208, 0.208]
WV	0.111	0.074
	[0.111, 0.111]	[0.074, 0.074]
WY	0	0

Table 18: Smoothed 50-50 Excess Seats Estimates

State	Estimate	State	Estimate
AK	0	MT	0
AL	1.333 [1.222, 1.444]	NC	2.444 [2.111, 2.444]
AR	0.222 [0.000, 0.333]	ND	0
AZ	-0.778 [-0.889, -0.667]	NE	-0.222 [-0.222, -0.111]
CA	-3.778 [-4.111, -3.111]	NH	0.000 [0.000, 0.000]
CO	0.333 [0.333, 0.333]	NJ	-0.111 [-0.111, -0.111]
CT	0.111 [-0.111, 0.111]	NM	-0.222 [-0.222, -0.222]
DE	0	NV	-0.556 [-0.556, -0.556]
FL	0.111 [0.000, 0.333]	NY	-0.667 [-0.667, -0.556]
GA	2.444 [2.444, 2.556]	OH	3.222 [3.222, 3.222]
HI	0.000 [0.000, 0.000]	OK	0.444 [0.222, 0.444]
IA	-0.111 [-0.111, -0.111]	OR	-0.111 [-0.222, -0.111]
ID	0.000 [0.000, 0.000]	PA	2.667 [2.556, 2.889]
IL	-1.444 [-1.444, -1.444]	RI	0.000 [0.000, 0.000]
IN	1.333 [1.333, 1.333]	SC	1.111 [1.111, 1.111]
KS	0.333 [0.222, 0.333]	SD	0
KY	1.111 [1.111, 1.222]	TN	2.333 [2.222, 2.333]
LA	2.000 [1.778, 2.000]	TX	-0.222 [-0.222, 0.111]
MA	-0.444 [-0.667, -0.444]	UT	-0.111 [-0.111, -0.111]
MD	-1.778 [-1.778, -1.778]	VA	0.778 [0.778, 0.778]
ME	0.111 [0.111, 0.111]	VT	0
MI	1.444 [1.444, 1.444]	WA	0.444 [0.333, 0.444]
MN	0.889 [0.889, 0.889]	WI	1.667 [1.667, 1.667]
MO	1.667 [1.444, 1.667]	WV	0.222 [0.222, 0.222]
MS	1.111 [1.111, 1.222]	WY	0

Table 19: Partisan Bias Republican Seat Shares by State Vote Share

State Vote Share	Seat Share	State Vote Share	Seat Share
0.25	0.014	0.55	0.670
	[0.013, 0.016]		[0.669, 0.672]
0.30	0.038	0.60	0.759
	[0.035, 0.040]		[0.758, 0.762]
0.35	0.116	0.65	0.831
	[0.115, 0.118]		[0.829, 0.834]
0.40	0.219	0.70	0.883
	[0.217, 0.221]		[0.882, 0.884]
0.45	0.394	0.75	0.933
	[0.390, 0.397]		[0.930, 0.936]
0.50	0.545		
	[0.543, 0.548]		