Voter Turnout and Preference Aggregation*

Kei Kawai†
University of California at Berkeley

Yuta Toyama‡
Northwestern University

Yasutora Watanabe§
Hong Kong University of Science and Technology

August 2016

Abstract

This paper studies how voter turnout affects aggregation of voter preferences in elections. Given that voting is costly, election outcomes disproportionately aggregate the preferences of voters with low voting cost or high preference intensity. We show that the correlation structure among preferences, costs, and perception of voting efficacy can be identified, and explore how the correlation affects preference aggregation. Using 2004 U.S. presidential election data, we find that minority, low-income, and less-educated voters are underrepresented. All of these groups tend to prefer Democrats except for the less-educated. Democrats would have won 8 more states if all eligible voters turned out.

keyword: *voter turnout, preference aggregation, election*

---

*We thank Heski Bar-Isaac, Karam Kang, Santiago Oliveros and Raul Sanchez de la Sierra for helpful comments.

†Department of Economics, University of California at Berkeley 530 Evans Hall #3880 Berkeley, CA, 94720-3880. Email: kei@berkeley.edu

‡Department of Economics, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208. Email: yutatoyama2012@u.northwestern.edu

§Department of Economics, HKUST Business School, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. Email: yasutorawatanabe@gmail.com

1 Introduction

Democracies rely on elections to aggregate the preferences of its citizens. Elections, however, aggregate the preferences of only those that participate. Studies of suffrage expansion document the importance of participation in various contexts such as the abolition of Apartheid in South Africa (Kroth et al., 2013), the passage of Voting Rights Act of 1965 (Husted and Kenny, 1997; Cascio and Washington, 2013), and the passage of womens' suffrage laws (Miller, 2008). Less dramatic measures that reduced the voting costs of certain groups of voters is found to affect policy in important ways (Fujiwara, 2015).

While most democracies now enjoy universal suffrage, participation in elections is far from perfect given the voluntary nature of voting. To the extent that preferences of those that turn out are systematically different from those that do not, election outcomes disproportionately aggregate the preferences of citizens with low voting cost or high preference intensity. Thus, how well elections aggregate the preferences of citizens, and whose preferences are underrepresented are open questions even in mature democracies.

The issue of preference aggregation and underrepresentation are also relevant from a policy perspective. The concern that the preferences of particular groups of voters are underrepresented has led some to argue for compulsory voting (See, e.g., Lijphart, 1997). More moderate policy proposals, such as introducing internet voting, relaxing registration requirements, and making an election day a holiday, are also motivated by similar concerns. Understanding how voter turnout affects preference aggregation can provide a basis for more informed discussions on these policy proposals.

In this paper, we explore the extent to which preferences are aggregated, which hinges on how preferences and determinants of turnout are correlated. We show that the joint distribution of preferences, voting cost, and perception of voting efficacy is identified, and estimate it using county-level voting data from the 2004 U.S. Presidential election. We then simulate the counterfactual election outcome under compulsory voting. The difference between the simulated and the actual outcomes allows us to quantify the degree to which preferences are aggregated.

We find that minority and low income citizens, who tend to prefer Democrats, are underrepresented. For example, a 1 percentage point increase in the fraction of Hispanics results in a 0.2 percentage point decrease in turnout and a 0.22 percentage point increase in the Democratic vote share relative to the Republican vote share. Less-educated and religious citizens, who tend to prefer Republicans, are also under-represented. In particular, a 1 year decrease in the years of schooling of all eligible voters results in a 6.21 percentage point decrease in turnout and a 4.68 percentage point increase in the Republican vote share relative to the Democratic vote share. In the counterfactual experiment in which we let all voters turn out, we find that the Democrats would have won the presidential election by gaining 82 more electors, overturning the results in 8 states.

The key challenge for studying the effect of turnout on preference aggregation is to identify the correlation between preferences and voting costs in the population. In particular, we need to identify how voter charactersitics, such as race and income, simultaneously determine preferences and costs. However, this is not a straightforward task because a high level of turnout among a particular set of voters may be due to low voting cost or high intensity of preference.

To illustrate, consider a plurality rule election in which voters choose either to vote for candidate A, B, or not turn out. Applying a discrete choice framework to the voter's decision, let $u_A(\mathbf{x})$ and $u_B(\mathbf{x})$ denote the utility of voting for candidates A and B, respectively, and $c(\mathbf{x})$ denote the cost of voting (relative to not voting), where $\mathbf{x}$ is a vector of voter characteristics. Then, the voter's mean utilities are as follows

$$\begin{aligned} V_A(\mathbf{x}) &= u_A(\mathbf{x}) - c(\mathbf{x}), \\ V_B(\mathbf{x}) &= u_B(\mathbf{x}) - c(\mathbf{x}), \text{ and} \\ V_0(\mathbf{x}) &= 0, \end{aligned}$$

where V_0 represents the mean utility of not voting. While one can identify $V_A(\mathbf{x}) = u_A(\mathbf{x}) - c(\mathbf{x})$ and $V_B(\mathbf{x}) = u_B(\mathbf{x}) - c(\mathbf{x})$ using vote share and turnout data (See Berry (1994) and Hotz and Miller (1993)), $u_A(\cdot)$, $u_B(\cdot)$, and $c(\cdot)$ are not separately identified without further restrictions. This is because making a voter care more about the election outcome (say, by adding $g(\mathbf{x})$ to $u_A(\mathbf{x})$ and $u_B(\mathbf{x})$) is observationally equivalent to lowering the cost of voting (by subtracting $g(\mathbf{x})$ from $c(\mathbf{x})$). Even if there are exogenous cost shifters, $\mathbf{z}$ (e.g., rainfall), they do not help separately identify $u_A(\cdot)$, $u_B(\cdot)$, and $c(\cdot)$.[1] Thus, most of the existing studies impose ad-hoc exclusion

[1] Suppose that the cost function is separated into two parts as $c = c_{\mathbf{x}}(\mathbf{x}) + c_{\mathbf{z}}(\mathbf{z})$, where $\mathbf{z}$ is a vector of cost shifters that is exluded from $u_A(\cdot)$ and $u_B(\cdot)$. Then, $u_A(\cdot) - c_{\mathbf{x}}(\cdot)$, $u_B(\cdot) - c_{\mathbf{x}}(\cdot)$ and

restrictions on the way $\mathbf{x}$ enters $u_A(\cdot)$, $u_B(\cdot)$, and $c(\cdot)$, assuming that $\mathbf{x}$ is excluded from either $u_k(\cdot)$ ($k = A, B$) or $c(\cdot)$. Imposing such exclusion restrictions assumes away the correlation structure among these terms and precludes the possibility that the preferences of those who vote are different from those who do not. Note that this identification challenge exists regardless of whether the data is at the individual level or at the aggregate level.

In this paper we uncover the correlation structure by identifying the joint distribution of preferences and costs in a setting in which $\mathbf{x}$ is allowed to enter both $u_k(\cdot)$ and $c(\cdot)$. Our identification is based on the simple observation that, unlike consumer choice problems where choosing not to buy results in the outcome of not obtaining the good, choosing not to turn out still results in either A or B winning the election. In the context of the canonical voting model of Downs (1957) and Riker and Ordeshook (1968), this observation implies that the voter's preference depends only on the utility difference between the two outcomes rather than the levels of utility associated with each outcome.[2] Barkume (1976) first used this observation to separately identify $u_k(\cdot)$ and $c(\cdot)$ in the context of property tax referenda for school districts.

To see how this observation leads to the identification of $u_k(\cdot)$ and $c(\cdot)$, consider the calculus of voting model of Downs (1957) and Riker and Ordeshook (1968). In these models, the utility of voting for candidate k can be expressed as $u_k = pb_k$ for $k = A, B$, where p is the voter's beliefs that she is pivotal.[3] Here, b_A is the utility difference between having candidate A in office and candidate B in office, and similarly for b_B. Hence, we have $b_A = -b_B$. The mean utilities can now be expressed as

$$\begin{aligned} V_A(\mathbf{x}) &= \quad pb_A(\mathbf{x}) - c(\mathbf{x}), \\ V_B(\mathbf{x}) &= -pb_A(\mathbf{x}) - c(\mathbf{x}), \text{ and} \\ V_0(\mathbf{x}) &= 0. \end{aligned}$$

The property that $b_A = -b_B$ allows us to separately identify preference and cost.

$c_{\mathbf{z}}(\cdot)$ are all seperately identified. However, $u_A(\cdot)$, $u_B(\cdot)$ and $c_{\mathbf{x}}(\cdot)$ are not separately identified. See the subsection titled Exogenous Cost Shifters towards the end of Section 4 for more details.

[2] This implication holds as long as the voters care about the ultimate outcome of the election. It may not hold for models in which voters gain utility from the act of voting for a candidate, e.g., models of expressive voting.

[3] More precisely, the utility of voting for candidate k *relative to not turning out* can be express as $u_k = pb_k$, by normalizing the utility of not turning out to be zero. See footnote 13 for details.

Adding the first two expressions above, we have $V_A(\mathbf{x})+ V_B(\mathbf{x}) = -2c(\mathbf{x})$ because $pb_A(\mathbf{x})$ cancels out. Given that $V_A(\mathbf{x})$ and $V_B(\mathbf{x})$ are both identified from the vote share and turnout data, $c(\cdot)$ is identified. Similarly, we can identify $pb_A(\cdot)$ because $V_A(\mathbf{x})- V_B(\mathbf{x}) = 2pb_A(\mathbf{x})$, and the left hand side is identified. Though this may appear mechnical, there is a straightforward intuition behind this result. $V_A(\mathbf{x})+ V_B(\mathbf{x})$ is primarily identified by voter turnout and $V_A(\mathbf{x})- V_B(\mathbf{x})$ is primarily identified by the vote share margin. Hence, voter turnout pins down $c(\cdot)$, while the difference in the two-party vote share pins down $pb_A(\cdot)$.

In this paper, we retain the basic structure of the calculus of voting model, but do not place additional restrictions on p such as rational expectations, in which p equals the actual pivot probability. In our model, we interpret p more broadly as the voter's perception of voting efficacy, which is allowed to differ across individuals and is allowed to be correlated with the true pivot probability in a general manner. In particular, we let p be a function of individual characteristics, $\mathbf{x}$, and the state in which the voter lives as $p = p_s \times \widetilde{p}(\mathbf{x})$, where p_s is a state fixed effect that we estimate. By letting p depend on each state, we can take into account the nature of the electoral college system.[4] We show that the ratio of state-specific component of efficacy, $p_s/p_{s'}$ $(\forall s, s')$, are identified in this model in addition to $\widetilde{p}(\cdot)$, $b_A(\cdot)$ and $c(\cdot)$ (up to a scalar normalization).[5] The ratio $p_s/p_{s'}$ is directly identified from the data without using equilibrium restrictions on p, such as rational expectations. Therefore, our identification and estimation results are agnostic about how voters formulate p.

Given the debate over how to model voter turnout, we briefly review the literature in order to relate our model to various models of voter turnout.[6] The model that we estimate in this paper is based on the decision theoretic model of voter turnout introduced by Downs (1957) and Riker and Ordeshook (1968). In their model, a voter turns out and votes for the preferred candidate if $pb-c+d > 0$, where p is the voter's beliefs over the pivot probability, b is the utility difference from having one candidate

[4] For example, electoral outcomes in battleground states such as Ohio were predicted to be much closer than outcomes in party strongholds such as Texas. Hence, we need to allow for the possibility that p is higher for voters in Ohio than for voters in Texas.

[5] More precisely, we can identify $p(\cdot)b_A(\cdot)$ state by state given that we have many counties within each state. Assuming that $\widetilde{p}(\cdot)$ and $b_A(\cdot)$ are common across states, we can identify $p_s/p_{s'}$. We also show that $\widetilde{p}(\cdot)$ and $b_A(\cdot)$ are separately identified up to a scalar multiple in our full specification with county level shocks to preferences and costs.

[6] For a survey of the literature, see, e.g., Dhillon and Peralta (2002), Feddersen (2004), and Merlo (2006).

in office relative to the other, c is the physical and psychological costs of voting, and d is the benefit from fullfilling civic duty of voting. While none of the terms in the calculus of voting are endogenized in the original papers, the decision theoretic model has provided a basic conceptual framework for much of the subsequent work on voting and turnout.

Subsequent papers to Riker and Ordeshook (1968) have endogenized or micro-founded each of the terms in the calculus of voting model in various ways. Ledyard (1984) and Palfrey and Rosenthal (1983, 1985) introduced the pivotal voter model in which pivot probabilities p are endogenized in a rational expectations equilibrium. They show that there exists an equilibrium with positive turnout in which voters have consistent beliefs about the pivot probabilities. Coate et al. (2008), however, points out that rational expectations pivotal voter model has difficulties matching the data on either the level of turnout or the winning margin.[7] Moreover, using laboratory experiments, Duffy and Tavits (2008) finds that voter's subjective pivot probabilities are much higher than actual pivot probabilities, which is at odds with the rational expectations assumption.

More recently, there are attempts at endogenizing p in ways other than rational expectations. For example, Minozzi (2013) proposes a model based on cognitive dissonance in the spirit of Akerlof and Dickens (1982) and Brunnermeier and Parker (2005). In his model, voters jointly choose p and whether or not to turn out in order to maximize subjective expected utility. Kanazawa (1998) introduces a model of reinforcement learning in which boundedly rational voters, who cannot compute the equilibrium pivot probabilities, form expectations about p from the correlation between their own past voting behavior and past election outcomes (See also Bendor et al. (2003) for a similar approach). While these models are based on the basic calculus of voting model, the p term in these models no longer carries the interpretation of the actual pivot probability.

Another strand of the literature endogenizes c and d terms. Harsanyi (1980) and Feddersen and Sandroni (2006) endogenize the d term by proposing a rule-utilitalian model in which voters receive a warm glow payoff from voting ethically. Based on their

[7]Note, however, that with aggregate uncertainty Myatt (2012) shows that the level of turnout can still be high with rational expectations. Levine and Palfrey (2007) also shows that combining the quantal response equilibrium with the pivotal voter model can generate high turnout, and finds that the results of laboratory experiments are consistent with the model prediction.

approach, Coate and Conlin (2004) estimates a group-utilitalian model of turnout. Shachar and Nalebuff (1999) also endogenizes the d term by considering the follow-the-leader model in which elites persuade voters to turn out. In a paper studying split-ticket voting and selective abstension in multiple elections, Degan and Merlo (2011) considers a model that endogenizes c to reflect the voter's mental cost of making mistakes.

In our paper we bring the calculus of voting model to the data without taking a particular stance on how p, c, or d terms are endogenized. Specifically, our identification and estimation do not use the restriction that p is equal to the actual pivot probability as in the rational expectations model. Instead, the p term that we recover can be broadly interpreted as the voter's perception of voting efficacy, and can be consistent with a wide class of models including Palfrey and Rosenthal (1983, 1985), Minozzi (2013), and Kanazawa (1998). We purposely aim to be agnostic about the different ways of modeling voter turnout so that the estimated terms can be interpreted as a reduced form of a diverse set of models that endogenize p, c, or d terms in specific ways. Instead of imposing equilibrium restrictions of a particular model a priori, we let the data directly identify p, b, and $c - d$ terms.

Relatedly, our paper does not impose a priori restrictions on how covariates should enter p, b, or $c-d$ terms, allowing, instead, the same set of covariates to affect all three terms. This is important because the way in which covariates enter p, b, and $c-d$ terms determines the correlation structure among them, which in turn, determines how well preferences are aggregated. In most of the existing studies, the set of covariates that enters p, b and $c-d$ terms are disjoint, precluding the possibility that preferences and costs are correlated. For example, Coate and Conlin (2004) and Coate et al. (2008) include demographic characteristics only in the b term,[8] while Shachar and Nalebuff (1999) include them in the $c - d$ term. In contrast, we can let each demographic characteristic enter all three terms p, b and $c - d$, allowing us to study the effects of turnout on preference aggregation.[9]

We use county-level data on voting outcomes from the 2004 U.S. Presidential Election to estimate the model.[10] Our data on turnout and vote share are constructed

[8] To be more precise, Coate and Conlin (2004) and Coate et al. (2008) use demographic characteristics as covariates for the fraction of population supporting one side.

[9] Degan and Merlo (2011) is a possible exception. They consider a model based on theories of regret in which the cost term is endogenized in a way that captures voters preferences over candidates. They include the same covariates in c and d terms.

[10] Although we use aggregate data, we account for the issue of ecological fallacy by computing the

taking into account the difference between the voting eligible population and the voting age population (McDonald and Popkin, 2001) using data on citizenship and the number of fellons. We also construct the joint distribution of demographic characteristics within each county from the 5% Public Use Microdata Sample of the Census. A benefit of using actual voting data over survey data is that we can avoid serious misreporting issues often associated with survey data such as overreporting of turnout and reporting bias in vote choice (see, e.g., Atkeson, 1999; DellaVigna et al., 2015).

We find that African Americans, Hispanics, and other minorities have high voting costs as do young, less-educated, low income, and religious voters. Moreover, young and less-educated voters have low perception of voting efficacy which further depresses turnout among these groups. Overall, Hispanic, young and less-educated voters are particularly underrepresented. We find that a one percentage point increase in the fraction of Hispanic voters decreases turnout by about 0.2 percentage points and a 1 year decrease in the years of schooling of all eligible voters results in a 6.2 percentage point decrease in turnout. In terms of preferences, minority, highly educated, low-income and non-religious voters are more likely to prefer Democrats.

Our results show that, overall, there is a positive correlation between voting cost and preference for Democrats that can be accounted for through observable characteristics. Except for two voter characteristics, namely years of schooling and being religious, we find that demographic characteristics that are associated with higher cost of voting are also associated with preferring Democrats. We also find that unobservable cost shocks are positively correlated with preference shocks for Democrats. Both of these correlations result in fewer Democratic votes relative to the preferences of the underlying population. We estimate that turnout is significantly lower among the electorate who prefer Democrats over Republicans, at 50.3%, than turnout among those who prefer Republicans over Democrats, at 63.5%.

Our findings using actual voting data are broadly consistent with findings based on survey data that document the differences in preferences between voters and non-voters (See, e.g., Citrin et al. (2003), Brunell and DiNardo (2004), Martinez and Gill (2005), and Leighley and Nagler (2013)).[11] Moreover, our paper provides an understanding of the mechanism that generates these differences between voters and

behavior of individual voters and aggregating them up to the county level.

[11]See also DeNardo (1980) and Tucker and DeNardo (1986) for early works that study the correlation between turnout and the Democratic vote share using aggregate data. For more recent work, see Hansford and Gomez (2010) that uses rainfall as an instrument for turnout.

non-voters through the correlation between preference, perception of efficacy, and voting costs. Our results also shed light on how preference *intensity* affects preference aggregation (See, e.g., Campbell (1999), Casella (2005) and Lalley and Weyl (2015)).

Regarding our results on the perception of voting efficacy, we find substantial across-state variation in our estimates of p_s, the state fixed effect in the voting efficacy. The estimates of p_s for some states are about twice as high as that of other states. Furthermore, the estimates are correlated with the ex-post closeness of the election: Battleground states such as Ohio and Wisconsin tend to have high estimates of p_s, while party strongholds such as New Jersey and California have low estimates, which is consistent with the comparative statics of the pivotal voter model. However, the magnitude of the estimated ratio of p_s is at most two for any pair of states. This is in contrast to a much larger variation in the ratio implied by the pivotal voter model with rational expectations.[12] Our results are more consistent with models of turnout in which voters' perception of efficacy are only weakly correlated with the actual pivot probabilities.

In our counterfactual we simulate the voting outcome under compulsory voting. We find that the vote share of Democrats increases in all states under compulsory voting. Overall, the increase in the Democrats' two-party vote share is about 5.8%. We also find that the increase in the Democratic vote share would overturn the election results in 8 states including key states such as Florida and Ohio resulting in the Democrats to win a plurality of the electors.

2 Model

2.1 Model Setup

Anticipating the empirical application of the paper, we tailor our model to the U.S. Presidential Election. Let $s \in \{1, ..., S\}$ denote a U.S. state, and $m \in \{1, ..., M_s\}$ denote a county in state s.

[12]The pivotal voter model with rational expectations predicts high variation in the ratio of pivot probabilities across states given the winner-take-all nature of the electoral college system. The distribution of voter preferences is such that voters in only a handful of swing states can have a reasonable probability of being pivotal (See, e.g., Shachar and Nalebuff, 1999).

Preference of Voters We consider a model of voting with two candidates, D and R. Each voter chooses to vote for one of the two candidates or not to vote. We let b_{nk} denote voter n's utility from having candidate $k \in \{D, R\}$ in office, p_n denote her perception of voting efficacy, and c_n denote her cost of voting. Given that there are only two possible outcomes (either D wins or R wins the election), the utility of voting for candidate k, U_{nk}, only depends on $b_{nD} - b_{nR}$ rather than b_{nD} and b_{nR} individually;

$$\begin{aligned} U_{nD} &= p_n(b_{nD} - b_{nR}) - c_n, \quad (1) \\ U_{nR} &= p_n(b_{nR} - b_{nD}) - c_n, \quad (2) \\ U_{n0} &= 0, \end{aligned}$$

where U_{n0} is the utility from not turning out, which we normalize to zero.[13] When p_n is the actual pivot probability, our model is the same as the pivotal voter model of Palfrey and Rosenthal (1983, 1985). However, we interpret p_n broadly as the voter's subjective perception of the voting efficacy as we discuss below. The cost of voting, c_n, includes both physical and psychological costs as well as possible benefits from fullfilling civic duty. Hence, c_n can be either positive or negative. When c_n is negative, the voter turns out regardless of the value of p_n and $b_{nD} - b_{nR}$.

We let the preferences of voter n in county m of state s depend on her demographic characteristics, $\mathbf{x}_n$, as

$$b_{nk} = b_k(\mathbf{x}_n) + \lambda_{sk} + \xi_{mk} + \varepsilon_{nk}, \text{ for } k \in \{D, R\},$$

where λ_{sk} is a state specific preference intercept that captures state level heterogeneity in voter preferences. ξ_{mk} and ε_{nk} are unobserved random preference shocks at the county level and at the individual level, respectively. ξ_{mk} captures unobserved factors that affect preferences at the county level, such as the benefits that the voters in county m receive due to particular policies supported by candidate k. Then, the expression for the utility difference is as follows;

$$b_{nD} - b_{nR} = b(\mathbf{x}_n) + \lambda_s + \xi_m + \varepsilon_n,$$

[13] Note that expressions (1) and (2) take the familiar form $pb - c$. This results from normalizing the utility of not turning out to be zero. See pages 29 and 30 of Riker and Ordeshook (1968) for derivation.

where $b(\mathbf{x}_n) = b_D(\mathbf{x}_n) - b_R(\mathbf{x}_n)$, $\lambda_s = \lambda_{sD} - \lambda_{sR}$, $\xi_m = \xi_{mD} - \xi_{mR}$, and $\varepsilon_n = \varepsilon_{nD} - \varepsilon_{nR}$. We assume that ε_n follows the standard normal distribution.

We also let the voting cost c_n be a function of voter n's characteristics as

$$c_n = c(\mathbf{x}_n) + \eta_m,$$

where η_m is a county-level shock on the cost of voting.[14] We assume that ξ_m and η_m are both independent of $\mathbf{x}_n$, but allow ξ_m and η_m to be correlated with each other.

We let the voting efficacy term, p_n, depend on both the demographic characteristics of voter n as well as the state in which she votes as,

$$p_n = p_s(\mathbf{x}_n) = p_s \times \tilde{p}(\mathbf{x}_n),$$

where p_s is a state specific coefficient that we estimate. It is important to let p_n depend on the state in which the voter votes because of the winner-take-all nature of the electoral votes in each state.[15] For example, in the 2004 Presidential Election, a vote in key states such as Ohio was predicted to matter a lot more than a vote elsewhere. Hence, our specification nests the rational expectations model as a special case in which p_s is equal to the actual pivot probability in state s and $\tilde{p}(\mathbf{x}_n) = 1$. However, rather than imposing the pivotal voter model (and hence placing equilibrium restrictions on p_n), we estimate p_s and $\tilde{p}(\cdot)$ directly from the data. This allows us to interpret p_n consistently with models of turnout that endogenize p_n in various ways. We also allow for the possibility that p_n depends on $\mathbf{x}_n$, the voter's social and economic status, which have been found to affect the voter's general sense of political efficacy (See e.g., Karp and Banducci, 2008).

Substituting the expressions for $b_{nD} - b_{nR}$, c_n, and p_n into equations (1) and (2),

[14] While we do not model the presence of other election such as gubernatorial and senetorial elections, previous studies (e.g., Smith, 2001) find that neither the presence nor the closeness of those elections affect turnout in presidential elections.

[15] In U.S. Presidential elections, the winner is determined by the Electoral College. Each U.S. State is allocated a number of electoral votes, roughly in proportion to the state's population. The electors of each state are awarded on a winner-take-all basis in all states except for Maine and Nebraska. The Presidential candidate who wins the plurality of electors' votes becomes the winner of the election.

the utility from choosing each of the alternatives can be expressed as follows;

$$\begin{aligned} U_{nD}(\mathbf{x}_n) &= p_s(\mathbf{x}_n)\left[b_s(\mathbf{x}_n) + \xi_m + \varepsilon_n\right] - c(\mathbf{x}_n) - \eta_m, \\ U_{nR}(\mathbf{x}_n) &= p_s(\mathbf{x}_n)\left[-b_s(\mathbf{x}_n) - \xi_m - \varepsilon_n\right] - c(\mathbf{x}_n) - \eta_m, \\ U_{n0}(\mathbf{x}_n) &= 0, \end{aligned}$$

where $b_s(\mathbf{x}_n)$ denotes $b(\mathbf{x}_n) + \lambda_s$.

Voter's Decision Voter n's problem is to choose alternative $k \in \{D, R, 0\}$ that gives her the highest utility,

$$k = \underset{\kappa \in \{D,R,0\}}{\arg\max} U_{n\kappa}(\mathbf{x}_n). \tag{3}$$

We can write the probability that voter n votes for candidate D as

$$\begin{aligned} &\Pr\left(D = \arg\max\nolimits_{\kappa \in \{D,R,0\}} U_{n\kappa}\right) \\ &\quad = \Pr\left(U_{nD} > U_{nR} \text{ and } U_{nD} > 0\right) \\ &\quad = \Pr\left(\varepsilon_n > -b_s(\mathbf{x}_n) - \xi_m \text{ and } \varepsilon_n > -b_s(\mathbf{x}_n) - \xi_m + \frac{c(\mathbf{x}_n) + \eta_m}{p_s(\mathbf{x}_n)}\right) \\ &\quad = 1 - \Phi\left(\max\left\{-b_s(\mathbf{x}_n) - \xi_m, -b_s(\mathbf{x}_n) - \xi_m + \frac{c(\mathbf{x}_n) + \eta_m}{p_s(\mathbf{x}_n)}\right\}\right), \end{aligned}$$

where Φ is the CDF of the standard normal. We can derive a similar expression for candidate R.

Figure 1 depicts the behavior of a voter as a function of ε_n. There are two cases to consider; one in which the cost of voting is positive (Case 1) and the other in which the cost of voting is negative (Case 2). In Case 1, a voter with a strong preference for one of the candidates (which corresponds to a large positive realization or a large nagative realization of ε_n) votes for her preferred candidate, while a voter who is relatively indifferent between the two candidates does not turn out. That is, a voter with high preference intensity relative to cost turns out, and a voter with low preference intensity does not. In Case 2, a voter always votes regardess of her preference intensity, as the cost of voting is negative.

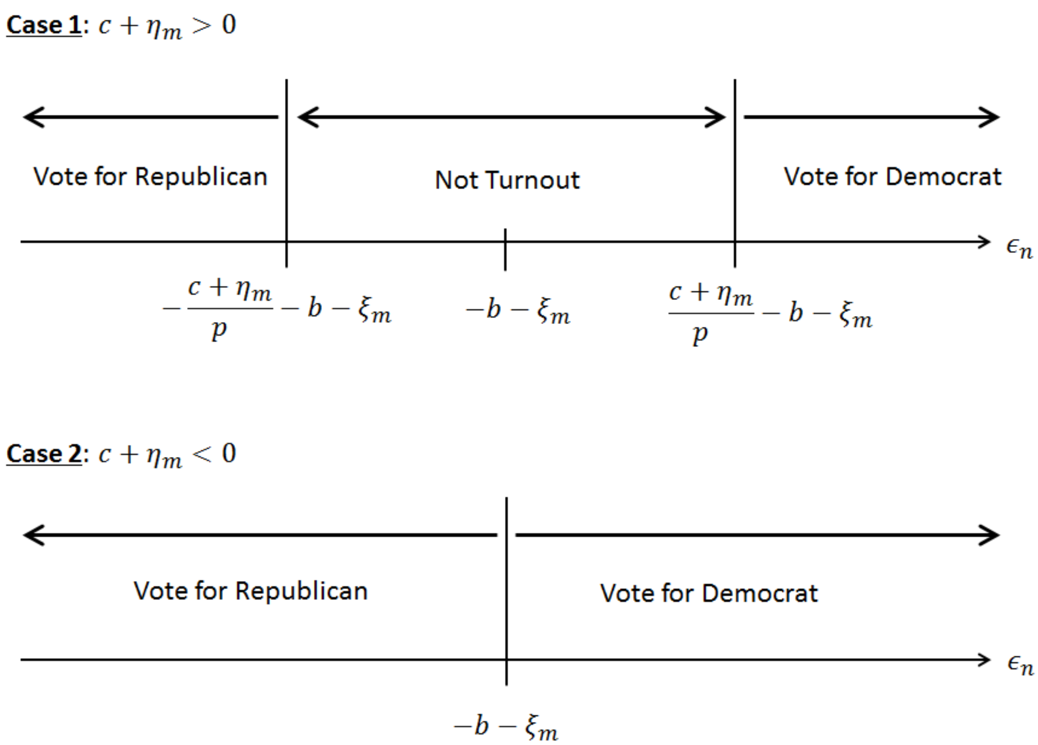

Figure 1: Voter's Decision as a Function of ε_n: The top panel correponds to the case in which a voter has positive costs of voting. The bottom panel corresponds to the case in which a voter has negative costs of voting.

Advertising and Campaign Visits An important feature of Presidential Elections that we have not explicitly modeled up to now is the campaign activities of the candidates. Candidates target key states with advertisements and campaign visits during the election. These campaign activities are endogenous and they depend on the expected closeness of the race in each state (See, e.g., Strömberg, 2008; Gordon and Hartmann, 2013).

While we do not have a specific model of political campaigns, the model accounts for their effect on voters through the state fixed effect in the voter's utility, λ_s. Because we treat λ_s as parameters to be estimated, λ_s can be arbitrarily correlated with the closeness of the race in the state. Hence, our estimates of voter preferences are consistent even when campaign activities are endogenous. We note, however, that our couterfactual results take the level of campaigning as given.

Vote Share and Voter Turnout We can express the vote share for candidate k in county m, $v_{k,m}$, and the fraction of voters who do not turn out, $v_{0,m}$, as follows;

$$v_{D,m} \equiv \int 1 - \Phi\left(\max\left\{-b_s(\mathbf{x}_n) - \xi_m, -b_s(\mathbf{x}_n) - \xi_m + \frac{c(\mathbf{x}_n)+\eta_m}{p_s(\mathbf{x}_n)}\right\}\right) dF_{\mathbf{x},m}(\mathbf{x}_n), \quad (4)$$

$$v_{R,m} \equiv \int \Phi\left(\min\left\{-b_s(\mathbf{x}_n) - \xi_m, -b_s(\mathbf{x}_n) - \xi_m - \frac{c(\mathbf{x}_n)+\eta_m}{p_s(\mathbf{x}_n)}\right\}\right) dF_{\mathbf{x},m}(\mathbf{x}_n), \quad (5)$$

$$v_{0,m} \equiv 1 - v_{D,m} - v_{R,m}, \quad (6)$$

where $F_{\mathbf{x},m}$ denotes the distribution of $\mathbf{x}$ in county m. Denoting the number of eligible voters in county m by N_m and the number of counties in state s as M_s, the vote share for candidate k in state s can be expressed as $\sum_{m=1}^{M_s} N_m v_{k,m} \Big/ \sum_{m=1}^{M_s} N_m$. The candidate with the highest vote share in state s is allocated all of the electors of that state.[16] The candidate who wins the plurality of the electors becomes the overall winner of the presidential election.

Discussion on p The modeling in our paper is purposely agnostic about how p is endogenized. Similarly, our estimation approach avoids using restrictions that are specific to a particular way of modeling voter beliefs, such as rational expectations (Palfrey and Rosenthal (1983, 1985)), overconfidence (Duffy and Tavits, 2008), cognitive dissonance (Minozzi (2013)), etc. Regardless of the way p is endogenized, there exists an equilibrium p that corresponds to the data generating process. Our approach is to identify and estimate both the model primitives and equilibrium p directly from the data with as little structure as possible. This empirical strategy is similar in spirit to that in the estimation of incomplete models in which some primitives are estimated from the data without fully specifying the model. For example, Haile and Tamer (2003) recovers bidder values without fully specifying a model of an English auction, using only the restriction that the winning bid lies between the valuation of the losers and the valuation of the winner. Given that their estimation procedure avoids using restrictions that are specific to a particular model, the estimates are consistent under a variety of models.

In section 4, we show that the key primitives of the model are identified without fully specifying how voters form p. We also show that p is identified directly from the

[16] Maine and Nebraska use a different allocation method. Hence, we drop these two states from our sample. See also footnote 15.

data.[17] The strength of our approach is that we impose little restrictions on beliefs and hence our estimates are consistent under a variety of behavioral assumptions regarding how p is formed. On the other hand, this approach limits types of counterfactual experiments that we can conduct since we do not specify a particular model regarding p.

3 Data

In this section, we describe our data and provide summary statistics. We use county-level voting data obtained from David Leip's Atlas of U.S. Presidential Elections. This dataset is a compilation of election data from official sources such as state boards of elections. We merge this dataset with county-level demographics data from the 2000 U.S. Census and population data from the 2004 Annual Estimates of the Resident Population from the Census Bureau. We construct the data on eligible voters for each county by combining the population estimates from the 2004 Annual Estimates and age and citizenship information from the 2000 Census. We then adjust for the number of fellons at the state level. Hence, our data accounts for the difference between the voting age population and voting eligible population (see McDonald and Popkin, 2001).

We construct the joint distribution of voter demographic characteristics and citizenship at the county level by combining the county-level marginal distribution of each demographic variable with the 5% Public Use Microdata Sample (See Appendix A for details). We augment the Census data with county-level information on religion using Religious Congregations and Membership Study 2000. In particular, we define the variable "religious" using adherence to either "Evangelical Denominations" or "Church of Jesus Christ of Latter-day Saints."

Our data consist of a total of 2,909 counties from 40 states. Because we need a large number of counties within each state to identify state specific parameters such as p_s and λ_s, we drop states that have fewer than 15 counties. These states are Alaska, Connecticut, District of Columbia, Delaware, Hawaii, Massachusettes, New Hampshire, Rhode Island and Vermont. In addition, we drop Maine and Nebraska because these two states do not adopt the winner-take-all rule to allocate electors.

[17] More precisely, $p_s/p_{s'}$ is identified for any s and s', and $\tilde{p}(\cdot)$ is identified up to a scaler normalization. See Section 4 for details.

We also drop counties with a population less than 1,000 because the vote shares and turnout rate are very variable for these counties.[18] Table 1 presents summary statistics of the county-level vote share, turnout, and demographic characteristics. Note that a Hispanic person may be of any race according to the definition used in the Census.

Voting Data	Obs	Mean	Std. Dev.	Min	Max
Vote Share: Democrat	2,900	0.22	0.08	0.04	0.52
Vote Share: Republican	2,900	0.35	0.09	0.06	0.66
Turnout Rate	2,900	0.57	0.09	0.17	0.94

County Demographics	Obs	Mean	Std. Dev.	Min	Max
% Hispanic	2,900	0.06	0.12	0.00	0.98
% Black/African American	2,900	0.09	0.15	0.00	0.87
% Neither Black nor White	2,900	0.05	0.08	0.00	0.95
Mean Age	2,900	49.16	2.80	34.20	59.66
Mean Income (USD 1,000)	2,900	44.39	9.19	23.14	93.12
Mean Years of Schooling	2,900	12.43	0.62	10.23	15.34
% Agriculture	2,900	0.07	0.07	0.00	0.43
% Manufacturing	2,900	0.24	0.09	0.04	0.54
% Religious	2,900	0.26	0.17	0.00	1.00

Table 1: Summary Statistics – Voting Outcome and Demographic Characteristics of Eligible Voters. For Age, Income, and Years of Schooling, the table reports mean, standard deviation, minimum, and maximum of the county mean. "% Religious" is the share of population with adherence to either "Evangelical Denomination" or "Church of Jesus Christ of Latter-day Saints."

In order to provide a sense of how turnout and expected closeness are related, Figure 2 plots the relationship between the (ex-post) winning margin and voter turnout at the state level. The two variables are negatively correlated, although the fitted line is relatively flat. The slope of the fitted line implies that a decrease in the (ex-post) winning margin of 10 percentage points is associated with an increase in turnout of only about 1.3 percentage points. While the negative correlation may be capturing some of the forces of the rational-expectations pivotal voter model, the flatness of the

[18] In addition, we drop one county, Chattahoochee, GA, as the turnout rate is extremely low (18.8%) relative to all other counties. The turnout rate for the next lowest county is 33%.

slope suggests that turnout is unlikely to be fully accounted for by the pivotal voter model.

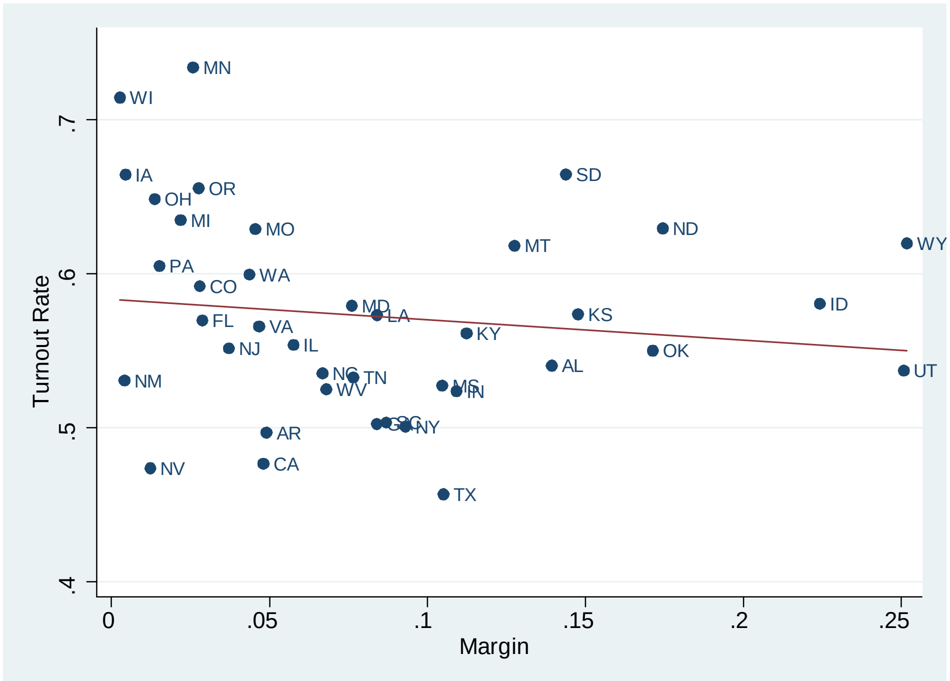

Figure 2: Relationship between Ex-Post Winning Margin and Voter Turnout. The slope coefficient is -0.13 and not statistically siginificant.

4 Identification

In this section, we discuss identification of our model as the number of counties *within each state* becomes large ($M_s \to \infty$, $\forall s$). Given that we have state specific parameters for $p_s(\cdot)$ and $b_s(\cdot)$, we require the number of observations per state to be large.

The identification of the model is based on the idea initially proposed by Barkume (1976). Recall that the observed vote shares are expressed as:

$$\begin{aligned}
v_{D,m} &\equiv \int 1 - \Phi\left(\max\left\{-b_s(\mathbf{x}_n) - \xi_m, -b_s(\mathbf{x}_n) - \xi_m + \frac{c(\mathbf{x}_n) + \eta_m}{p_s(\mathbf{x}_n)}\right\}\right) dF_{\mathbf{x},m}(\mathbf{x}_n), \\
v_{R,m} &\equiv \int \Phi\left(\min\left\{-b_s(\mathbf{x}_n) - \xi_m, -b_s(\mathbf{x}_n) - \xi_m - \frac{c(\mathbf{x}_n) + \eta_m}{p_s(\mathbf{x}_n)}\right\}\right) dF_{\mathbf{x},m}(\mathbf{x}_n), \\
v_{0,m} &\equiv 1 - v_{D,m} - v_{R,m}.
\end{aligned}$$

For exposition, consider the simple case in which there is no heterogeneity in voters' observable characteristics, so that $\mathbf{x}_n = \overline{\mathbf{x}}_m$ for all n in county m.[19] In this case, the above expressions simplify as follows:

$$\begin{aligned}
v_{D,m} &\equiv 1 - \Phi\left(\max\left\{-b_s(\overline{\mathbf{x}}_m) - \xi_m, -b_s(\overline{\mathbf{x}}_m) - \xi_m + \frac{c(\overline{\mathbf{x}}_m) + \eta_m}{p_s(\overline{\mathbf{x}}_m)}\right\}\right), \quad (7)\\
v_{R,m} &\equiv \Phi\left(\min\left\{-b_s(\overline{\mathbf{x}}_m) - \xi_m, -b_s(\overline{\mathbf{x}}_m) - \xi_m - \frac{c(\overline{\mathbf{x}}_m) + \eta_m}{p_s(\overline{\mathbf{x}}_m)}\right\}\right), \quad (8)\\
v_{0,m} &\equiv 1 - v_{D,m} - v_{R,m}. \quad (9)
\end{aligned}$$

We now show that the primitives of the model are identified from expressions (7), (8), and (9).

Using the fact that Φ is a strictly increasing function, we can rewrite expressions (7) and (8) as follows:

$$\begin{aligned}
\Phi^{-1}(1 - v_{D,m}) &= \max\left\{-b_s(\overline{\mathbf{x}}_m) - \xi_m,\ -b_s(\overline{\mathbf{x}}_m) - \xi_m + \frac{c(\overline{\mathbf{x}}_m) + \eta_m}{p_s(\overline{\mathbf{x}}_m)}\right\},\\
\Phi^{-1}(v_{R,m}) &= \min\left\{-b_s(\overline{\mathbf{x}}_m) - \xi_m,\ -b_s(\overline{\mathbf{x}}_m) - \xi_m - \frac{c(\overline{\mathbf{x}}_m) + \eta_m}{p_s(\overline{\mathbf{x}}_m)}\right\}.
\end{aligned}$$

Rearranging these two equations, we obtain the following expressions:

$$\begin{aligned}
\frac{\Phi^{-1}(1 - v_{D,m}) + \Phi^{-1}(v_{R,m})}{-2} &= b_s(\overline{\mathbf{x}}_m) + \xi_m, \text{ and} \quad (10)\\
\frac{\Phi^{-1}(1 - v_{D,m}) - \Phi^{-1}(v_{R,m})}{2} &= \max\left\{0, \frac{c(\overline{\mathbf{x}}_m)}{p_s(\overline{\mathbf{x}}_m)} + \frac{\eta_m}{p_s(\overline{\mathbf{x}}_m)}\right\}. \quad (11)
\end{aligned}$$

Note that the left hand side of (10) reflects the difference in the vote share, and the left hand side of (11) reflects the voter turnout rate. This is because, if we ignore the nonlinerlity of $\Phi^{-1}(\cdot)$ and the denominator, the former reduces to $1 - v_{D,m} + v_{R,m}$ and the latter to $1 - v_{D,m} - v_{R,m}$. The left hand side of expressions (10) and (11) can be directly computed using data on vote shares, $v_{D,m}$ and $v_{R,m}$.

We first consider identification of $b_s(\cdot)$ and $F_\xi(\cdot)$, the distribution of ξ. Taking the

[19] Note that we are well aware of the issues of ecological fallacy. In what follows, we consider the simplified setup with $\mathbf{x}_n = \overline{\mathbf{x}}_m$, $\forall n$ in county m, just for expositional purposes. In our empirical excercise, we fully deal with the fact that each county has a distribution of $\mathbf{x}$ by integrating the vote share for each $\mathbf{x}$ with respect to $F_{\mathbf{x},m}(\mathbf{x})$.

expectation of (10) conditional on $\overline{\mathbf{x}}_m$, we have

$$\mathbf{E}\left[\left.\frac{\Phi^{-1}\left(1-v_{D,m}\right)+\Phi^{-1}\left(v_{R,m}\right)}{-2}\right|\overline{\mathbf{x}}_m\right]=b_s(\overline{\mathbf{x}}_m), \tag{12}$$

because $\mathbf{E}\left[\xi_m|\overline{\mathbf{x}}_m\right]=0$. As the left-hand side of the above expression is identified, $b_s(\cdot)$ is (nonparametrically) identified for each s (Note that the asymptotics is with respect to the number of counties within each state). The fact that $b_s(\cdot)$ is identified by the left hand side of (12) is very intuitive. Because $b_s(\cdot)$ is the utility difference between Democrats and Republicans, $b_s(\cdot)$ should be identified by variation in the two-party vote share. The left hand side of (12) is simply the mean of the left hand side of (10), which reflects the two party vote share as we discussed above.

Now consider identification of $F_\xi(\cdot)$. Given that $b_s(\cdot)$ is identified and the left-hand side of (10) is observable, each realization of ξ_m can be recovered from (10). Hence, $F_\xi(\cdot)$ is also identified. Note that if $b_s(\cdot)$ is linear in $\overline{\mathbf{x}}_m$ (i.e., $b_s(\overline{\mathbf{x}}_m)=\boldsymbol{\beta}\overline{\mathbf{x}}_m$), one can simply regress the left hand side of expression (10) on $\overline{\mathbf{x}}_m$ by OLS to obtain $\boldsymbol{\beta}$ as coefficients and ξ_m as residuals.

We now discuss identification of $p_s(\cdot)$ and $c(\cdot)$. For simplicity, consider the case in which the second term inside the max operator of expression (11) is positive with probability 1, i.e.,

$$\frac{\Phi^{-1}\left(1-v_{D,m}\right)-\Phi^{-1}\left(v_{R,m}\right)}{2}=\frac{c(\overline{\mathbf{x}}_m)}{p_s(\overline{\mathbf{x}}_m)}+\frac{\eta_m}{p_s(\overline{\mathbf{x}}_m)}. \tag{13}$$

This corresponds to the case in which turnout rate is always less than 100%. In this case, we have

$$\mathbf{E}\left[\left.\frac{\Phi^{-1}\left(1-v_{D,m}\right)-\Phi^{-1}\left(v_{R,m}\right)}{2}\right|\overline{\mathbf{x}}_m\right] = \frac{c(\overline{\mathbf{x}}_m)}{p_s(\overline{\mathbf{x}}_m)}, \text{ and} \tag{14}$$

$$\mathbf{Var}\left[\left.\frac{\Phi^{-1}\left(1-v_{D,m}\right)-\Phi^{-1}\left(v_{R,m}\right)}{2}\right|\overline{\mathbf{x}}_m\right] = \frac{\mathbf{Var}(\eta_m)}{(p_s(\overline{\mathbf{x}}_m))^2}. \tag{15}$$

First, the ratio of $c(\cdot)/p_s(\cdot)$ is identified using (14) because the left hand side of (14) is identified. Given that the ratio $c(\cdot)/p_s(\cdot)$ is identified for all s and that $p_s(\cdot)=p_s\times\tilde{p}(\cdot)$, we can identify $p_s/p_{s'}$ from the ratio of $c(\overline{\mathbf{x}}_m)/p_{s'}(\overline{\mathbf{x}}_m)$ and $c(\overline{\mathbf{x}}_m)/p_s(\overline{\mathbf{x}}_m)$. Intuitively, the ratio $c(\cdot)/p_s(\cdot)$ is identified in each state by the mean turnout because the left hand side of (14) is simply the average of the left hand side of (11), which reflects the voter tunrout rate as discussed above. Moreover, $p_s/p_{s'}$ is identified by the ratio of mean turnout between two counties across states s and s' with the same demographics.

Second, $p_s(\cdot)$, $c(\cdot)$, and $F_\eta(\cdot)$ are identified up to a scale normalization (up to $\mathbf{Var}(\eta_m)$). Because the left hand side of (15) is identified, $p_s(\cdot)$ is identified up to $\mathbf{Var}(\eta_m)$. Then, $c(\cdot)$ is identified (also up to $\mathbf{Var}(\eta_m)$) from (14). Given that $p_s(\cdot)$ and $c(\cdot)$ are identified, we can recover the distribution of η_m, $F_\eta(\cdot)$, from (13). Intuitively, variance of turnout identifies $p_s(\cdot)$ as the left hand side of (15) reflects the variance of turnout. Then, $c(\cdot)$ is identified by the mean divided by the standard deviation of turnout.

Note that when the second term in (11) is not always positive, expression (11) can be seen as a censored regression of the form $y_m = \max\{0, g(\overline{\mathbf{x}}_m) + \epsilon_m\}$ with heteroskedasticity. We show in Appendix B that $p_s(\cdot)$, $c(\cdot)$, and $F_\eta(\cdot)$ are also identified up to $\mathbf{Var}(\eta_m)$ in this general case.

The discussion up to now is based on the simplified case in which the vote shares for each demographic characteristics are observed, or equivalently, all voters in county m have the same demographic characteristics, i.e., $\mathbf{x}_n = \overline{\mathbf{x}}_m$ for all n. As long as there is enough variation in $F_{\mathbf{x},m}(\mathbf{x})$, we can recover the vote shares conditional on each $\mathbf{x}$ and apply the identification discussion above.

Correlation between Unobserved Cost and Preference Shocks Our identification makes no assumptions regarding the correlation between the unobservables ξ_m and η_m. As ξ_m and η_m enter separately in (10) and (11), $\xi_m \perp \overline{\mathbf{x}}_m$ and $\eta_m \perp \overline{\mathbf{x}}_m$ are sufficient to identify the unknown primitives on the right hand side in each equation. Hence, we do not require any restrictions on the joint distribution of ξ_m and η_m. In fact, we can nonparametrically identify the joint distribution of ξ_m and η_m from the joint distribution of the residuals in each equation. In our estimation we specify the distribution of (ξ_m, η_m) as a bivariate Normal with correlation coefficient ρ. We find that ρ is positive, implying that the cost of voting tends to be higher for counties in which voters are more Democratic.

Exogenous Cost Shifters Lastly, we discuss identification when there exist instruments (e.g., rainfall) that shift the cost of voting but not the preference of the voters. The point we wish to make is that the existence of exogenous cost shifters are neither necessary nor sufficient for identification.

To illustrate this point, consider the following discrete choice setup with instruments $\mathbf{z}_n$,

$$\begin{aligned} V_A &= u_A(\mathbf{x}_n) - c_{\mathbf{x}}(\mathbf{x}_n) - c_{\mathbf{z}}(\mathbf{z}_n) \\ V_B &= u_B(\mathbf{x}_n) - c_{\mathbf{x}}(\mathbf{x}_n) - c_{\mathbf{z}}(\mathbf{z}_n) \\ V_0 &= 0. \end{aligned}$$

where V_k denotes mean utility of choosing $k \in \{A, B, 0\}$. Here, $u_A(\mathbf{x}_n)$ is not necessarily equal to $-u_B(\mathbf{x}_n)$ and the cost function is separated into two components, $c_{\mathbf{x}}(\mathbf{x}_n)$ and $c_{\mathbf{z}}(\mathbf{z}_n)$, where $\mathbf{z}_n$ is a vector of cost shifters that is excluded from $u_k(\mathbf{x}_n)$. For any arbitrary function $g(\mathbf{x}_n)$, consider an alternative model with $\tilde{u}_k(\mathbf{x}_n) = u_k(\mathbf{x}_n) + g(\mathbf{x}_n)$ $(k \in \{A, B\})$ and $\tilde{c}_{\mathbf{x}}(\mathbf{x}_n) = c_{\mathbf{x}}(\mathbf{x}_n) + g(\mathbf{x}_n)$ as follows:

$$\begin{aligned} V_A &= \tilde{u}_A(\mathbf{x}_n) - \tilde{c}_{\mathbf{x}}(\mathbf{x}_n) - c_{\mathbf{z}}(\mathbf{z}_n) \\ V_B &= \tilde{u}_B(\mathbf{x}_n) - \tilde{c}_{\mathbf{x}}(\mathbf{x}_n) - c_{\mathbf{z}}(\mathbf{z}_n) \\ V_0 &= 0. \end{aligned}$$

Because $\tilde{u}_k(\mathbf{x}_n) - \tilde{c}_{\mathbf{x}}(\mathbf{x}_n) = u_k(\mathbf{x}_n) - c_{\mathbf{x}}(\mathbf{x}_n)$, the two models are observationally equivalent, and thus, $u_k(\mathbf{x}_n)$ and $c_{\mathbf{x}}(\mathbf{x}_n)$ are not separately identified. In particular, correlation between preference and cost cannot be identified because this model cannot rule out $c_{\mathbf{x}}(\mathbf{x}_n) = 0$, $\forall \mathbf{x}_n$. This is true even if $\mathbf{z}_n$ has a rich support. Hence, it is *not* the the availability of instruments, but rather the obseravation that we can express $u_A(\mathbf{x}_n) = -u_B(\mathbf{x}_n)$ that idetifies the primitives of the model.

5 Specification and Estimation

5.1 Specification

We now specify $b_s(\cdot)$, $c(\cdot)$, $p_s(\cdot)$ and the distribution of (ξ, η) for our estimation. The function $b_s(\cdot)$, which is the utility difference from having candidates D and R in office, is specified as a function of a state-level preference shock, λ_s, and demographic characteristics, $\mathbf{x}_n$, consisting of age, race, income, religion, occupation, and years of schooling, i.e.,

$$b_s(\mathbf{x}_n) = \lambda_s + \boldsymbol{\beta}_b'\mathbf{x}_n.$$

The intercept, λ_s, is a parameter that we estimate for each state and it captures the state-level preference shock for the Democrats that is unaccounted for by demographic chracteristics. Voting cost is also specified as a linear function of $\mathbf{x}_n$ as

$$c(\mathbf{x}_n) = \boldsymbol{\beta}_c[1, \mathbf{x}_n']'.$$

We do not specifically model the presence of other elections such as gubernatorial and senetorial elections because previous studies (e.g., Smith, 2001) find that neither the presence nor the closeness of other elections affect turnout in presidential elections.

We specify the voter's perception of efficacy as $p_s \times \tilde{p}(\mathbf{x}_n)$, where $\tilde{p}(\cdot)$ is a function of the voter's age, income, years of schooling as follows,[20]

$$\tilde{p}(\mathbf{x}_n) = \exp(\boldsymbol{\beta}_p'\mathbf{x}_n).$$

We normalize $p_s = 1$ for Alabama and normalize $\tilde{p}(\cdot)$ such that $\tilde{p}(\bar{\mathbf{x}}) = 1$ where $\bar{\mathbf{x}}$ is the national average of $\mathbf{x}_n$.[21]

We specify the joint distribution of county-level preference shock ξ and cost shock η as a bivariate normal, $N(0, \Sigma)$, where Σ is the variance covariance matrix with diagonal elements equal to σ_ξ^2, σ_η^2 and off-diagonal elements $\rho\sigma_\xi\sigma_\eta$.

Finally, we considered specifications that include weather related variables in $c(\cdot)$ in addition to the demographic characteristics. However, there is not enough variation in precipitation or temperature on the day of the 2004 Presidential election to affect turnout in a significant way.[22] Thus, we do not include any weather related variables.

[20] The set of variables we include in $p_s(\mathbf{x}_n)$ is a subset of $\mathbf{x}_n$ that takes continuous values. We do not include dummy variables such as race, occupation and religion. Variation in $c(\cdot)$ changes the utility level additively, while variation in $p_s(\cdot)$ changes it multiplicatively. As dummy variables takes only 0 and 1, it is difficult in practice to distinguish whether the effect of those variables are additive or multiplicative. Thus, estimating the model with dummy variables in both cost and efficacy is difficult, and we include only continuous variables in $p_s(\cdot)$.

[21] Note that we need two normalizations. Because we express $p_s(\mathbf{x}_n)$ as $p_s \times \tilde{p}(\mathbf{x}_n)$, we need a scalar normalization on either p_s or $\tilde{p}(\mathbf{x}_n)$. We need an additional normalization because $p_s(\cdot)$, $c(\cdot)$ and $F_\eta(\cdot)$ are identified only up to the variance of η. Assuming that $\tilde{p}(\bar{\mathbf{x}}) = 1$ eliminates this degree of freedom.

[22] We included weather variables in the simple model that assumes $\mathbf{x}_n = \mathbf{x}_m$ (i.e., the demographic characteristics of voters in each county are assumed to be the same within county) and found the coefficients on the weather variables small and insignificant.

5.2 Estimation

We use method of moments to estimate the model parameters. Recall that the vote shares in county m are given by expressions (4), (5) and (6), where $F_{\mathbf{x},m}$ is the distribution of $\mathbf{x}_n$ in county m. For a fixed vector of model parameters, $\boldsymbol{\theta} = (\boldsymbol{\beta}_b, \{\lambda_s\}, \boldsymbol{\beta}_c, \{p_s\}, \boldsymbol{\beta}_p, \sigma_\xi, \sigma_\eta, \rho)$, we can compute moments of expressions (4), (5) and (6) by integrating over ξ and η. Our estimation is based on matching the moments generated by the model with the corresponding sample moments.

To be more precise, we define the first and second order moments implied by the model as follows,

$$
\begin{aligned}
\hat{v}_{k,m}(\boldsymbol{\theta}) &= \mathbf{E}_{\xi,\eta}[v_{k,m}(\xi,\eta;\boldsymbol{\theta})],\ \forall k \in \{D,R\}, \\
\hat{v}_{k,m}^{squared}(\boldsymbol{\theta}) &= \mathbf{E}_{\xi,\eta}[v_{k,m}(\xi,\eta;\boldsymbol{\theta})^2],\ \forall k \in \{D,R\}, \\
\hat{v}_{m}^{cross}(\boldsymbol{\theta}) &= \mathbf{E}_{\xi,\eta}[v_{D,m}(\xi,\eta;\boldsymbol{\theta}) v_{R,m}(\xi,\eta;\boldsymbol{\theta})],
\end{aligned}
$$

where $v_{k,m}(\xi,\eta;\boldsymbol{\theta})$ is the vote share of candidate k given a realization of (ξ,η) and parameter $\boldsymbol{\theta}$.[23] Denoting the observed vote share of candidate k in county m as $\mathsf{v}_{k,m}$, our objective function, $J(\theta)$, is given by,

$$
J(\boldsymbol{\theta}) = \sum_{k=\{D,R\}} \left(\frac{J_{1,k}(\boldsymbol{\theta})}{\widehat{\mathbf{Var}}(\mathsf{v}_{k,m})} + \frac{J_{2,k}(\boldsymbol{\theta})}{\widehat{\mathbf{Var}}(\mathsf{v}_{k,m}^2)} \right) + \frac{J_3(\boldsymbol{\theta})}{\widehat{\mathbf{Var}}(\mathsf{v}_{D,m}\mathsf{v}_{R,m})},
$$

where

$$
\begin{aligned}
J_{1,k}(\boldsymbol{\theta}) &= \frac{1}{M}\sum_{m=1}^{M}(\hat{v}_{k,m}(\boldsymbol{\theta}) - \mathsf{v}_{k,m})^2,\ \forall k \in \{D,R\}, \\
J_{2,k}(\boldsymbol{\theta}) &= \frac{1}{M}\sum_{m=1}^{M}(\hat{v}_{k,m}^{squared}(\boldsymbol{\theta}) - \mathsf{v}_{k,m}^2)^2,\ \forall k \in \{D,R\}, \\
J_3(\boldsymbol{\theta}) &= \frac{1}{M}\sum_{m=1}^{M}\left(\hat{v}_{m}^{cross}(\boldsymbol{\theta}) - \mathsf{v}_{D,m}\mathsf{v}_{R,m}\right)^2.
\end{aligned}
$$

$J_{1,k}$ is the sum of squared differences between the expectation of the predicted vote

[23] Computing $\hat{v}_{k,m}(\boldsymbol{\theta})$, $\hat{v}_{k,m}^{squared}(\boldsymbol{\theta})$, and $\hat{v}_{m}^{cross}(\boldsymbol{\theta})$ requires integration over (ξ,η). For integration, we use a quadrature with $[5 \times 5]$ nodes with pruning (See Jäckel, 2005) with a total of 21 nodes.

share ($\hat{v}_{k,m}(\boldsymbol{\theta})$) and the actual vote share ($\mathsf{v}_{k,m}$).[24] $J_{2,k}$ is the sum of squared differences between $\hat{v}_{k,m}^{squared}(\boldsymbol{\theta})$ and the squared vote share, $\mathsf{v}_{k,m}^2$. J_3 is the sum of squared differences between the predicted and the actual cross terms. M is the total number of counties, $\sum_{s=1}^{S} M_s$, and $\widehat{\mathbf{Var}}(z)$ denotes the sample variance of z. Note that, at the true parameter, each of the moments ($J_{1,k}$, $J_{2,k}$, J_3) should be zero in expectation.

The construction of our objective function closely follows our identification argument. The first moment, $J_{1,k}$, matches the conditional expectation of the vote shares from the model with that from the data. Intuitively, $J_{1,k}$ corresponds to (12) and (14), and pins down $\boldsymbol{\beta}_b$, $\{\lambda_s\}$, $\boldsymbol{\beta}_c/p_s$, $\boldsymbol{\beta}_c/\boldsymbol{\beta}_p$ and $p_s/p_{s'}$, following our identification discussion. The second and third moments, $J_{2,k}$ and J_3, correspond to (15). These moments pin down $\boldsymbol{\beta}_p$, σ_ξ,σ_η and ρ.

6 Results

The set of parameters that we estimate include those that are specific to each state, ($\{\lambda_s\}$,$\{p_s\}$) and those that are common across all states ($\boldsymbol{\beta}_b$,σ_ξ,$\boldsymbol{\beta}_c$,σ_η,$\boldsymbol{\beta}_p$,ρ). Table 2 reports the estimates of the latter set, while Tables 4 and 5 collect parameter estimates of the former set.

Estimates of $\boldsymbol{\beta}_b$, σ_ξ, $\boldsymbol{\beta}_c$, σ_η, $\boldsymbol{\beta}_p$, and ρ The first column of Table 2 reports the preference parameters. The estimate of the constant term in the first column corresponds to the preference of the voter who has $\mathbf{x}_n$ equal to the national average and has λ_s equal to Alabama. We find that age and income enter the utility difference, $b_D - b_R$, negatively, implying that young and lower income voters are more likely to prefer Democrats. Years of schooling enters the utility difference positively, thus more educated voters are more likely to prefer Democrats. Hispanics, African American, and other non-Whites also prefer Democrats relative to non-Hispanics and Whites (excluded categories). In terms of occupation, voters in manufacturing and agriculture are more likely to prefer Republicans compared to voters in the service sector (excluded category). The Religion variable carries a negative coefficient implying that Religious voters prefer Republicans more than non-Religious voters.

In the second column of Table 2, we report the cost parameter estimates. The

[24] We only use moments based on (4) and (5). The moment based on (6) is redundant because (4), (5) and (6) sum up to one.

	Preference		Cost		Efficacy	
	Estimate	SE	Estimate	SE	Estimate	SE
Age	-0.0038	(0.0033)	-0.0325	(0.0165)	0.0226	(0.0100)
Years of Schooling	0.0829	(0.0211)	-0.2802	(0.1109)	0.0170	(0.0613)
log(income)	-0.4685	(0.0537)	-1.7382	(0.5317)	-1.0221	(0.2295)
Hispanic	0.8652	(0.1479)	1.5157	(0.4080)		
African American	1.4706	(0.0771)	0.7715	(0.2023)		
Other Races	1.2665	(0.1469)	1.5375	(0.4321)		
Agriculture	-1.7834	(0.2142)	-5.7448	(3.1592)		
Manufacturing	-0.0228	(0.0788)	-0.3592	(0.1372)		
Religious	-0.5913	(0.0585)	0.2112	(0.0794)		
Constant	-0.0623	(0.0437)	1.0129	(0.1334)		
Sigma	0.1895	(0.0086)	0.1293	(0.1167)		
Rho		0.7291 (0.5427)				

Table 2: Parameter Estimates: The table reports the parameter estimates of the voters' preferences, costs, and perception of voting efficacy. The estimate of the constant term in the first column corresponds to the preference of the voter who has $\mathbf{x_n}$ equal to the national average and has λ_s equal to Alabama. The estimate of the constant term in the second column corresponds to the cost of a voter whose characteristics are set to the national mean. The variable log(income) is log of income divided by 1000. Standard errors are reported in parentheses.

estimate of the constant term in the second column corresponds to the cost of a voter whose characteristics are set to the national mean.We find that age, years of schooling, and income enter voting cost negatively. This implies that older, more educated, and higher income voters have lower cost of voting. Hispanic, African American, and Other Races have higher cost of voting relative to non-Hispanics and Whites. Voters in the service sector and Religious voters also have relatively high cost of voting.

The third column of Table 2 reports the efficacy parameters. We find that Age and Years of Schooling enter the perception of efficacy positively, while Income enters negatively. This implies that older, more-educated and lower-income citizens tend to have higher perception of efficacy. Given that older and more-educated voters have lower voting cost as well, they are more likely to be over-represented compared with young and less-educated voters. Regarding income, the overall effect on participation depends on the relative magnitudes of the cost and efficacy coefficients.

Finally, the last row of the table reports the estimate of ρ, which is the correlation between unobservable shocks ξ and η. The estimate is positive (0.729), implying that the correlation in the unobservable shocks tends to result in the election outcome to underweight the preference for Democrats.

To sum, estimates reported in Table 2 show that preference, voting costs, and perception of efficacy are correlated through both the observable characteristics and the unobservable shocks. Unobservable shocks in preference and cost have positive correlation. Regarding observable characteristics, younger voters, for example, tend to prefer Democrats, have higher cost of voting, and have lower perception of efficacy, making the election outcome underweight the support for Democrats from these voters. We find a similar correlation between preference and cost of voting for lower-income voters, minority voters, and voters in the service sector. In contrast, less-educated voters and Religious voters have higher cost of voting, while preferring Republicans.

Representation and Preference Aggregation We now examine i) which demographic groups are under/overrepresented, and ii) how preference aggregation is affected by the overall correlation among preference, voting cost, and perception of efficacy induced through demographics. Table 3 reports how marginal changes in the demographic variables affect the election outcome. In particular, the table reports the effect of a one-year increase in Age and Years of Schooling among the electorate, a one-percentage increase in Income among the electorate, and a one-percentage point increase in the population share of each demographic characteristic (Hispanic, African American, etc.).[25] The first two columns report the effect on vote shares. The third and fourth columns report the effect on the vote share margin and on turnout. The extent to which a particular demographic group is underrepresented or overrepresented is reflected in the marginal effect on turnout.

Table 3 shows that younger, less-educated, and lower-income citizens are likely to

[25] For computing the effect of a one-percentage point increase in the fraction of Hispanics (and similarly for other dummy variables such as Agriculture and Religious), we inject Hispanic voters to each county equaling 1% of the county's voter population. The demographic characteristics of the injected Hispanic voters such as age and income are chosen to reflect the demographics of the existing Hispanic voters in each county. We then simulate the vote share and turnout with the new demographic composition. An alternative way to simulate the outcome is to use the unconditional distribution of the demographics of each county (as opposed to the conditional distribution of Hispanic voters in each county). We find the results using the unconditional distribution to be very similar to the ones using the conditional distribution.

	Democrats		Republicans		Margin (D - R)		Turnout	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Baseline	26.92%		30.07%		-3.15%		56.99%	
Age	0.57%	(0.17%)	0.53%	(0.14%)	0.05%	(0.11%)	1.10%	(0.28%)
Years of Schooling	5.44%	(1.25%)	0.77%	(1.24%)	4.68%	(0.97%)	6.21%	(2.30%)
Income	-0.08%	(0.04%)	0.17%	(0.05%)	-0.25%	(0.03%)	0.10%	(0.08%)
Hispanic	0.01%	(0.07%)	-0.21%	(0.03%)	0.22%	(0.07%)	-0.20%	(0.09%)
Non-Hispanic	0.00%	(0.01%)	0.02%	(0.00%)	-0.02%	(0.01%)	0.03%	(0.01%)
African American	0.24%	(0.03%)	-0.23%	(0.02%)	0.47%	(0.05%)	0.01%	(0.02%)
Other Races	0.06%	(0.07%)	-0.24%	(0.03%)	0.30%	(0.07%)	-0.17%	(0.08%)
White	-0.05%	(0.01%)	0.06%	(0.01%)	-0.11%	(0.02%)	0.02%	(0.01%)
Agriculture	-0.16%	(0.08%)	0.58%	(0.22%)	-0.74%	(0.17%)	0.42%	(0.28%)
Manufacturing	0.03%	(0.02%)	0.02%	(0.02%)	0.01%	(0.03%)	0.05%	(0.03%)
Service	0.00%	(0.01%)	-0.02%	(0.01%)	0.02%	(0.01%)	-0.02%	(0.01%)
Religious	-0.12%	(0.02%)	0.08%	(0.02%)	-0.20%	(0.03%)	-0.04%	(0.02%)
Non-Religious	0.02%	(0.00%)	-0.02%	(0.00%)	0.04%	(0.01%)	0.01%	(0.00%)

Table 3: Marginal Effects of Demographic Characteristics on Election Outcomes. The table reports changes in vote share, vote share difference, and turnout from marginal changes in demographic characteristics. The reported changes correspond to a one-year increase in Age and Years of Schooling among the electorate, a one-percent increase in income among the electorate, and a one-percentage point increase in the population share of each demographic characteristic listed in rows 5 to 14. Standard errors are reported in parentheses.

be underrepresented. In particular, a one-year increase in years of schooling increases voter turnout by 6.2 percentage points, indicating that less-educated citizens are heavily underrepresented. Also, Hispanic, Other Races, People in Service Industry, and Religious voters tend to be underrepresented. Among the dummy characteristics, people in agriculture sector is the most overrepresented group with 0.42 percentage point increase in turnout resulting from a one-percentage point increase in population.

Table 3 also reports the relationship between representation and preference aggregation. In the table, Age, Years of Schooling, African-American, Manufacturing, and Religious have the same sign in both turnout and the vote share difference between Democrat and Republican. This implies that older, more-educated, African-American, Manufacturing, and non-Religious individual tend to be overrepresented *and* prefer Democrats. In contrast, Income, Hispanic, White, Other Races, and Agriculture and Service Industry have opposite sign in turnout and vote share difference. This implies that lower-income, Hispanic, Other Races, and Service Industry individuals tend to be underrepresented *and* prefer Democrats, while White and Agirculture industry individuals tend to be overrepresented *and* prefer Republicans. These results shows that there is a systematic selection in preferences of those who turn out.

	Estimate	SE		Estimate	SE
Alabama	0 (Normalized)		New Jersey	0.335	(0.050)
Arkansas	0.453	(0.046)	New Mexico	0.004	(0.088)
California	0.307	(0.070)	New York	0.318	(0.042)
Colorado	0.245	(0.058)	North Carolina	0.136	(0.036)
Florida	0.082	(0.041)	North Dakota	0.240	(0.048)
Georgia	-0.024	(0.030)	Ohio	0.239	(0.037)
Idaho	0.026	(0.047)	Oklahoma	0.123	(0.038)
Illinois	0.333	(0.035)	Oregon	0.249	(0.053)
Indiana	0.195	(0.038)	Pennsylvania	0.255	(0.041)
Iowa	0.434	(0.039)	South Carolina	0.011	(0.032)
Kansas	0.002	(0.038)	South Dakota	0.223	(0.043)
Kentucky	0.328	(0.036)	Tennessee	0.376	(0.039)
Louisiana	-0.062	(0.034)	Texas	-0.038	(0.035)
Maryland	0.210	(0.055)	Utah	0.044	(0.061)
Michigan	0.313	(0.036)	Virginia	0.166	(0.038)
Minnesota	0.427	(0.038)	Washington	0.386	(0.061)
Mississippi	-0.068	(0.033)	West Virginia	0.292	(0.045)
Missouri	0.295	(0.034)	Wisconsin	0.457	(0.039)
Montana	0.162	(0.047)	Wyoming	-0.042	(0.055)
Nevada	0.194	(0.066)			

Table 4: Estimates of State Preference Fixed Effects Relateve to $\lambda_{Alabama}$. Standard errors are reported in parentheses. Higher values imply stronger preference for Democrats.

Estimates of State Specific Effects, λ_s and p_s Table 4 presents the estimates of the state fixed effects in the voter's utility relative to $\lambda_{Alabama}$. Larger values imply that the voters in the corresponding state prefer Democrats. These state fixed effects capture the state-specific preferences for candidates after controlling for demographics, which may include inherent preferences of the voters and/or the effect of campaign activities by the candidates. The estimates show that states such as Lousiana, Mississippi, and Wyoming have strong preferences for Republicans, while states such as Arkansas and Wisconsin have strong preference for Democrats. Note also that Democratic strongholds such as New York and California tend to have high estimated values of λ_s, while "red" states such as Georgia and Texas tend to have low estimated values of λ_s.

	Estimate	SE		Estimate	SE
Alabama	1 (Normalized)		New Jersey	0.693	(0.063)
Arkansas	0.733	(0.039)	New Mexico	1.191	(0.092)
California	0.827	(0.061)	New York	0.823	(0.048)
Colorado	0.954	(0.063)	North Carolina	0.826	(0.036)
Florida	1.044	(0.053)	North Dakota	0.983	(0.051)
Georgia	0.832	(0.034)	Ohio	1.204	(0.051)
Idaho	1.012	(0.062)	Oklahoma	0.980	(0.040)
Illinois	0.995	(0.043)	Oregon	1.197	(0.063)
Indiana	0.751	(0.043)	Pennsylvania	0.875	(0.051)
Iowa	1.097	(0.051)	South Carolina	0.832	(0.037)
Kansas	0.819	(0.040)	South Dakota	1.375	(0.076)
Kentucky	1.046	(0.041)	Tennessee	0.887	(0.039)
Louisiana	1.185	(0.054)	Texas	0.887	(0.038)
Maryland	0.758	(0.049)	Utah	1.086	(0.060)
Michigan	1.061	(0.048)	Virginia	0.826	(0.036)
Minnesota	1.416	(0.072)	Washington	0.901	(0.051)
Mississippi	1.114	(0.048)	West Virginia	0.879	(0.042)
Missouri	1.090	(0.042)	Wisconsin	1.379	(0.071)
Montana	0.938	(0.052)	Wyoming	1.000	(0.058)
Nevada	0.737	(0.109)			

Table 5: Estimates of Statel Voting Efficacy Fixed Effects. Standard errors are reported in parentheses. Alabama is set at 1 for normalization.

Table 5 reports the state specific component of the perception of voting efficacy, p_s, with normalization $p_{Alabama} = 1$. High values of p_s correspond to high perception of voting efficacy after controlling for demographics. Because electors are determined at the state level, the perception of voting efficacy may vary across states, partly reflecting the pivot probability of the election in each state. We find that Minnesota, South Dakota, Wisconsin, and Ohio have the highest estimated values of voting efficacy, while New Jersey, Arkansas and Nevada have the lowest estimated values. Except for South Dakota, states with the highest efficacy correspond to those that were considered as battleground states. We also find that states considered as party strongholds such as California and New York have low estimated values. These results suggest a weak positive relationship between pivot probability and perception of voting efficacy. To illustrate this relationship, Figure 3 plots the estimates of p_s and the winning margin, which proxies for the pivot probability. The estimated perception of voting efficacy and the margin have a negative relationship with a slope of -0.16.

While some of the forces of the rational-expectations pivotal voter model seem to be at play, the estimated values of p_s suggest that the pivotal voter model is

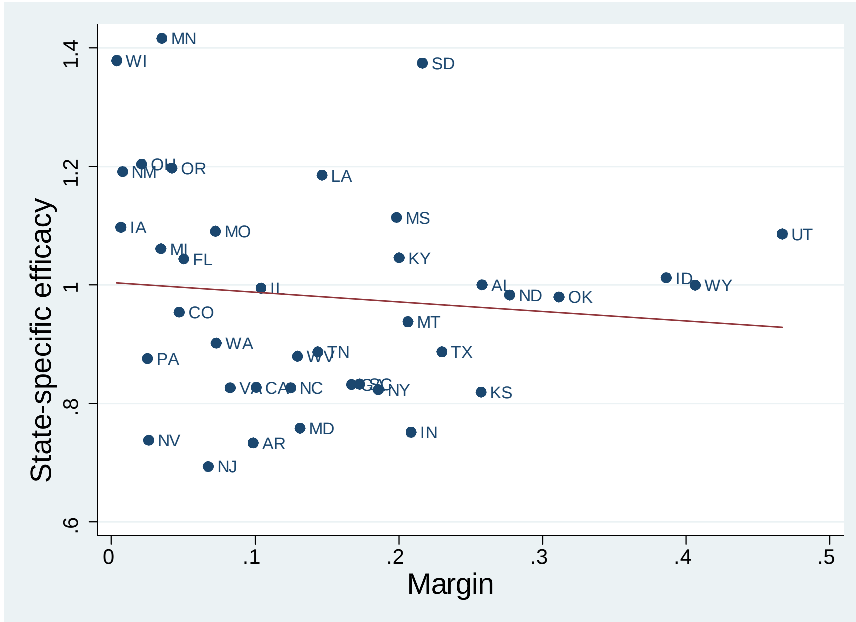

Figure 3: Margin and State-specific Efficacy. The horizontal axis is the winning margin and the vertical axis is the estiamte of the state-specific component in voting efficacy. The fitted line has a slope of –0.16.

unlikely to explain voting behavior very well. Models of voting based on rational expectations would require p_s in battleground states to be of orders of magnitude greater than those in party strongholds. However, our estimates of p_s fall within a narrow range; the ratio of the estimated state-level efficacy parameters, $p_s/p_{s'}$, is at most two. This is unlikely under the rational-expectations pivotal voter model. Our results highlight the importance of relaxing the assumption of rational expectations on pivot probabiliy.

Fit In order to assess the fit of our model, Figure 4 plots the county-level vote shares, voter turnout, and vote share margin predicted from the model against the data. The predicted vote shares, turnout, and vote share margin are computed by integrating out the draws of ξ and η. The plots line up around the 45 degree line, which suggests that the model fits the data well.

In previous work, Coate et al. (2008) discusses the difficulty of fitting the winning margin using the rational expectation pivotal voter model. In our paper, the model can fit the winning margin in the data well because we do not impose the rational expectations assumption.

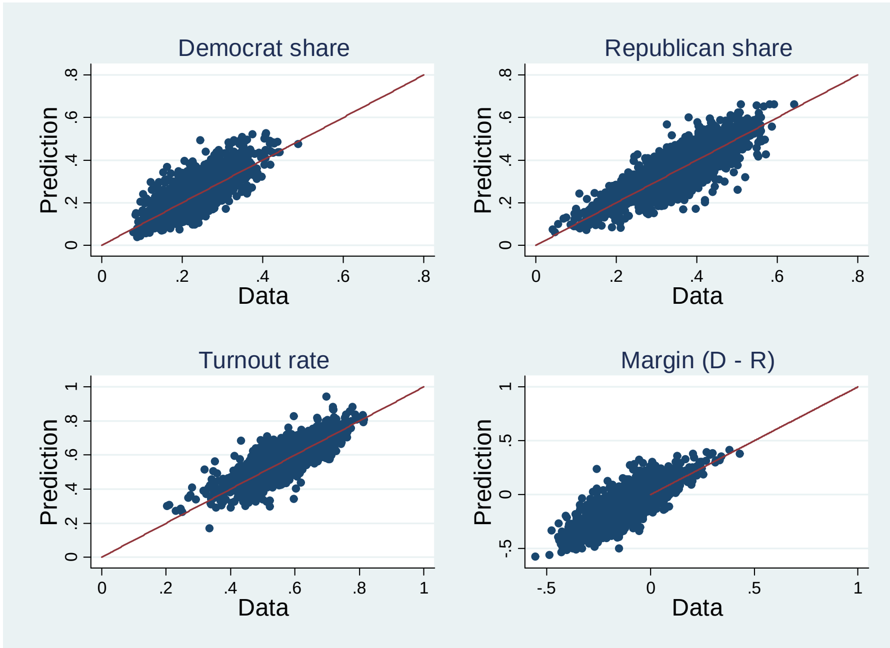

Figure 4: Fit. The figure plots county level predicted vote share, turnout, and vote share margin against data.

Turnout and Intensity of Preferences We now discuss how intensity of preferences is distributed and how it affects preference aggregation. Our discussion is related to the recent literature that studies how preference intensity afffects preference aggregation (See, e.g., Campbell (1999), Casella (2005) and Lalley and Weyl (2015)), In our discussion, we interpret the absolute value of the utility difference, $|b(\mathbf{x}_n)| = |b_s(\mathbf{x}_n) + \varepsilon_n|$, as intensity of preference. Note that our discussion on the intensity in this subsection may depend on the distributional assumption of idiosyncratic preference error ε_n.

First, Figure 5 plots $b(\mathbf{x}_n)$ and $c(\mathbf{x}_n)$ for each vector of demographic characteristics, $\mathbf{x}_n$, when the state specific preference parameter is set to the mean of all states, i.e., $\Sigma\lambda_s/S$. Each circle represents a pair $(b(\mathbf{x}_n),\ c(\mathbf{x}_n))$ with the diameter of the circle proportional to the mass of voters with characteristics $\mathbf{x}_n$ and ε_n in the population. Circles located to the right of the figure correspond to voters with high preference intensity for Democrats, circles to the left of the figure are those with high preference intensity for Republicans, and circles in the middle of the figure are those who are relatively indifferent. The figure shows a positive correlation between the voting cost and the preference for Democrats, implying that voters with high preference intensity

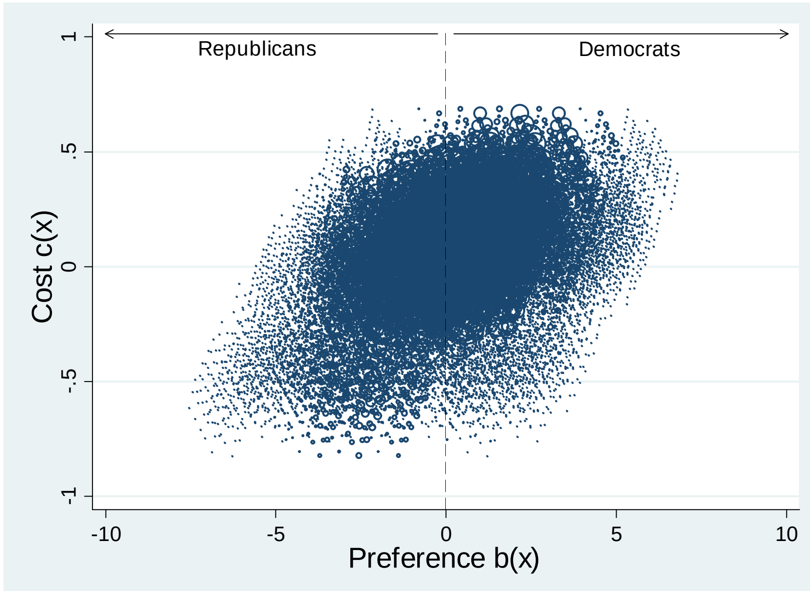

Figure 5: Scatter plot of $b(\mathbf{x}_n)$ and $c(\mathbf{x}_n)$. The figure plots $b(\mathbf{x}_n)$ and $c(\mathbf{x}_n)$ for each $\mathbf{x}_n$ when the state-specific preference dummy is set to the mean, $\Sigma\lambda_s/S$. The diameter of the circles is proportional to the mass of voters with characteristics $\mathbf{x}_n$ and ε_n.

for Democrats tend to have high cost of voting, while voters with high preference intensity for Republicans tend to have low cost of voting. The figure also shows that voters with negative cost of voting are largely Republican supporters.

Second, Figure 6 plots the histogram of the utility difference, $b(\mathbf{x}_n)$ (top panel), and the proportion of voters who turn out for given levels of $b(\mathbf{x}_n)$ (bottom panel). Note that the top panel of the figure is constructed by weighting $b(\mathbf{x}_n)$ by the distribution of $\mathbf{x}_n$ and ε_n. The top panel shows that the distribution of the utility difference is roughly centered around zero, and has a slightly fatter tail on the Democrat's side. The bottom panel shows that there is high turnout among voters with high preference intensity for either party. This results is consistent with the theoretical result of Campbell (1999) in which he shows that an alternative preferred by a minority with strong preference intensity is likely to win. The panel also shows that the turnout is higher among Republican supporters than Democratic supporters at the same level of preference intensity.

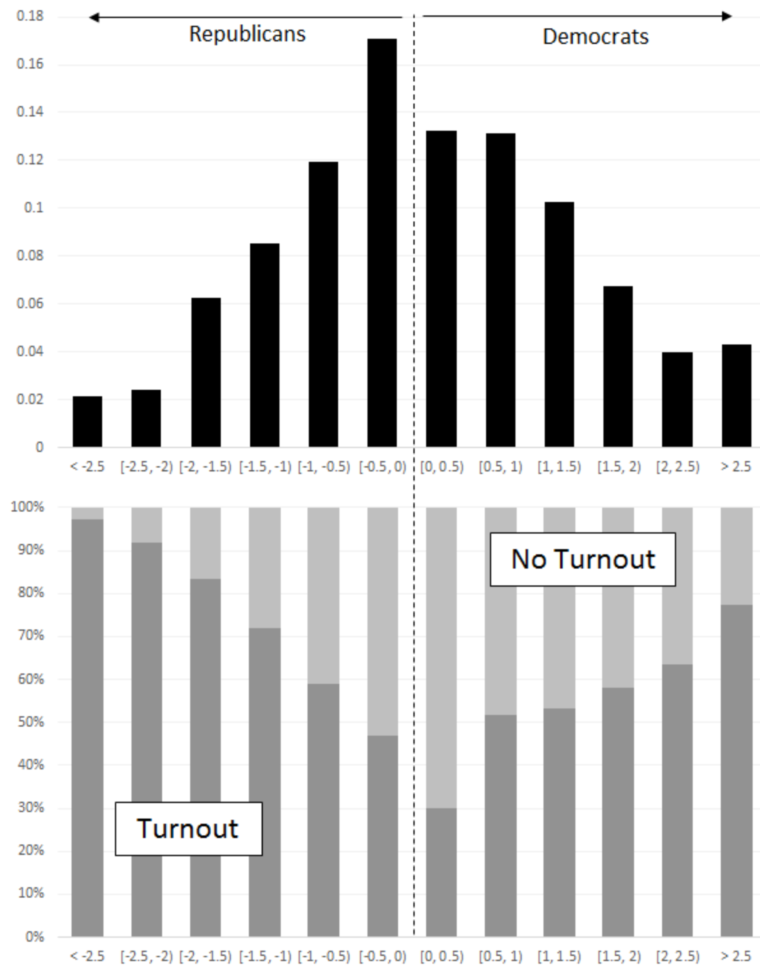

Figure 6: Histogram of Preference Intensity and Franction of Tunrout by Preference Intesity. The top panel plots the histogram of the utility difference, $b(\mathbf{x}_n)$. The bottom panel plots the proportion of those who turn out for given levels of preference intensity.

7 Counterfactual Experiments

In our first counterfactual, we consider what the outcome of the election would be if the preferences of all the eligible voters are aggregated. In the results section above, we show that preference, voting cost, and perception of efficacy are correlated, and that preferences of those who turn out do not necessarily reflect the preferences of the general population. In this counterfactual, we quantify the degree to which this discrepancy affects preference aggregation.

In this counterfactual, we compute the election outcome by setting the voting cost to zero. Then, individuals vote for Democrats or Republicans depending on the sign of $\widehat{b}_s(\mathbf{x}_n)+\widehat{\xi}_m+\varepsilon_n$, where $\widehat{b}_s(\cdot)$ and $\widehat{\xi}_m$ are the estimates of the net utility difference and county-level preference shock.[26] Hence, we calculate the counterfactual county-level vote shares, $\tilde{v}_{D,m}$ and $\tilde{v}_{R,m}$, as

$$\begin{aligned} \tilde{v}_{D,m} &\equiv \int \Phi\left(\widehat{b}_s(\mathbf{x}_n)+\widehat{\xi}_m\right) dF_{\mathbf{x},m}(\mathbf{x}_n), \\ \tilde{v}_{R,m} &\equiv 1-\tilde{v}_{D,m}. \end{aligned}$$

Note that our counterfactual results are robust to equlibrium adjustments to voters' perception of efficacy because a voter's decision depends only on the sign of the utility difference.

	Two-Party Vote Share			# of Electors	
	Democrats	Republicans	Turnout Rate	Democrats	Republicans
Actual	48.3%	51.7%	55.1%	208	268
Counterfactual	54.1%	45.9%	100.0%	300	176
	(3.2%)	(3.2%)	n.a.	(68.6)	(68.6)

Table 6: Counterfactual Outcome Under Full Turnout. The table compares the actual outcome with the counterfactual outcome in which all voters turn out. The reported outcomes do not include the results for the 11 states that we drop from the sample. Standard errors are reported in prenthesis.

In Tables 6 and 7 we compare the actual outcomes to the counterfactual outcomes

[26] Note that there is a unique value of $(\hat{\xi}_m, \hat{\eta}_m)$ that rationalizes the actual vote outcome given our estimates of voter preferences, costs and perception of efficacy. We use these values of $(\hat{\xi}_m, \hat{\eta}_m)$ to compute our counterfactual.

for the 39 states in our sample. The first row of Table 6 reports the actual vote share, turnout rate, and the number of electors for the two parties. We report our counterfactual results in the second row. We find that the Democratic vote share in the counterfactual increases from 48.3% to 54.1%, reflecting our earlier finding that the utility difference and the voting costs are positively correlated.

In terms of the number of electors, we find that the results are overturned in 8 states at the estimated parameters, and the number of electors for the Democrats increases by 92 from 208 to 300. Overall, the Democrats would have won the 2004 Presidential election if the preferences of all voters had been aggregated. Although there are 11 states that are not in our sample, 300 electors are larger than the number of electors needed to win the election (270).[27] The standard errors in our parameter estimates translate to about a 82% probability that the number of electors for Democrats exceeds 270. Note that this probability is the lower bound which assumes that all states not included in our sample vote for the Republicans electors.

Table 7 presents the state-level breakdown of the counterfactual results. The two-party vote share of the Democrats increases under compulsory voting in all states, but there is a considerable heterogeneity in the change across states. For example, in Texas, we find that the change in the two-party vote share for the Democrat is more than 10 percentage points (from 38.5% to 49.8%), while, in Minnesota, the change is only 1.7 percentage points. In general, the increase tends to be large for states with low turnout. The shaded rows in the table correspond to the states in which the winning party under the counterfactual switches from the actual result. There are 8 such states including key states such as Florida and Ohio.

In our second counterfactual, we examine how the outcomes change as we vary the level of turnout. We compute the outcomes corresponding to different levels of turnout (from 10% to 100% in increments of 10%) by adding (or subtracting) a constant to the cost of voting for all voters. Table 8 reports the results. We find a monotonic increase for both the vote share and the number of electors in favor of Democrats as the turnout increases, except for the change in the number of electors between 10% and 20% turnout. We also find that the election results would be overturned at a voter turnout level between 70% and 80%, at least for the subset of the states in our sample. Unlike the counterfactual of the compulsory voting discussed above, however,

[27] There are a total of 538 electors including states that are excluded from our sample. The candidate needs 270 electors to win.

	Two-Party Vote Share of Democrats			Turnout Rate	# of Electors for Democrats		
	Actual	Counterfactual		Actual	Actual	Counterfactual	
Alabama	37.1%	42.9%	(2.7%)	54.0%	0	0	(0.0)
Arkansas	45.1%	51.2%	(3.4%)	49.7%	0	6	(3.0)
California	55.0%	64.1%	(5.0%)	47.7%	55	55	(12.7)
Colorado	47.6%	53.8%	(3.2%)	59.2%	0	9	(2.6)
Florida	47.5%	52.3%	(2.8%)	57.0%	0	27	(9.7)
Georgia	41.6%	48.8%	(3.3%)	50.2%	0	0	(4.6)
Idaho	30.7%	38.1%	(2.8%)	58.1%	0	0	(0.0)
Illinois	55.2%	59.9%	(3.1%)	55.4%	21	21	(4.3)
Indiana	39.6%	46.1%	(3.1%)	52.4%	0	0	(1.1)
Iowa	49.7%	52.3%	(2.0%)	66.4%	0	7	(2.3)
Kansas	37.1%	44.2%	(3.0%)	57.4%	0	0	(0.0)
Kentucky	40.0%	44.7%	(2.5%)	56.1%	0	0	(0.0)
Louisiana	42.7%	47.7%	(2.6%)	57.3%	0	0	(1.0)
Maryland	56.6%	61.5%	(2.8%)	57.9%	10	10	(1.5)
Michigan	51.7%	54.8%	(2.2%)	63.5%	17	17	(4.2)
Minnesota	51.8%	53.5%	(1.5%)	73.4%	10	10	(2.3)
Mississippi	40.1%	45.6%	(2.7%)	52.7%	0	0	(0.0)
Missouri	46.4%	49.2%	(2.0%)	62.9%	0	0	(3.6)
Montana	39.7%	45.1%	(2.6%)	61.8%	0	0	(0.0)
Nevada	48.7%	57.5%	(4.8%)	47.4%	0	5	(1.4)
New Jersey	53.4%	60.3%	(3.7%)	55.2%	15	15	(3.5)
New Mexico	49.6%	56.3%	(3.9%)	53.1%	0	5	(1.4)
New York	59.3%	65.2%	(3.6%)	50.1%	31	31	(4.0)
North Carolina	43.8%	49.2%	(3.0%)	53.5%	0	0	(5.4)
North Dakota	36.1%	42.0%	(2.5%)	62.9%	0	0	(0.0)
Ohio	48.9%	51.4%	(1.9%)	64.9%	0	20	(7.7)
Oklahoma	34.4%	42.1%	(3.1%)	55.0%	0	0	(0.0)
Oregon	52.1%	54.9%	(2.2%)	65.6%	7	7	(1.7)
Pennsylvania	51.3%	54.3%	(2.4%)	60.5%	21	21	(5.4)
South Carolina	41.4%	47.3%	(3.1%)	50.3%	0	0	(0.9)
South Dakota	39.2%	43.5%	(2.0%)	66.5%	0	0	(0.0)
Tennessee	42.8%	47.7%	(2.7%)	53.3%	0	0	(1.9)
Texas	38.5%	49.8%	(4.5%)	45.7%	0	0	(14.7)
Utah	26.7%	35.0%	(2.9%)	53.7%	0	0	(0.0)
Virginia	45.9%	51.7%	(2.9%)	56.6%	0	13	(5.5)
Washington	53.6%	57.9%	(3.0%)	59.9%	11	11	(2.5)
West Virginia	43.5%	48.9%	(3.0%)	52.5%	0	0	(1.6)
Wisconsin	50.2%	52.1%	(1.5%)	71.4%	10	10	(2.9)
Wyoming	29.7%	37.3%	(2.8%)	62.0%	0	0	(0.0)
Total	48.3%	54.1%	(3.2%)	55.1%	208	300	(68.6)

Table 7: State-level simulation results with compulsory voting. Shaded rows correspond to the states in which the winning party under the counterfactual differs from the actual data. Total number of electors is 538 (the total number of electors for the states included in our data is 476), and 270 electors are needed to win the election..

	Vote Share		Two-Party	Electors	
Turnout	Democrats	Republicans	Democrat Share	Democrats	Republicans
100%	54.1%	45.9%	54.1%	300	176
90%	47.7%	42.4%	52.9%	300	176
80%	41.2%	38.8%	51.5%	281	195
70%	35.1%	34.9%	50.2%	225	251
60%	29.3%	30.7%	48.9%	220	256
(Actual) 55.1%	26.6%	28.5%	48.3%	208	268
50%	23.8%	26.2%	47.6%	198	278
40%	18.5%	21.5%	46.2%	198	278
30%	13.4%	16.6%	44.8%	72	404
20%	8.7%	11.3%	43.4%	62	414
10%	4.3%	5.7%	43.0%	120	356

Table 8: Election Outcomes at Various Levels of Turnout. We change the constant term in the voter's cost function to compute the election outcome under various levels of turnout.

this counterfactual exercise (other than the 100% turnout case) has a limitation that we cannot account for possible endogenous changes in the perception of efficacy.

References

Akerlof, George A and William T Dickens, "The economic consequences of cognitive dissonance," *The American Economic Review*, 1982, pp. 307–319.

Atkeson, Lonna Rae, ""Sure, I voted for the winner!" Overreport of the primary vote for the party nominee in the national election studies," *Political Behavior*, 1999, *21* (3), 197–215.

Barkume, Anthony J, "Identification of preference for election outcomes from aggregate voting data," *Public Choice*, 1976, *27* (1), 41–58.

Bendor, Jonathan, Daniel Diermeier, and Michael Ting, "A behavioral model of turnout," *American Political Science Review*, 2003, *97* (02), 261–280.

Berry, Steven T, "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, 1994, *25* (2), 242–262.

Brunell, Thomas L and John DiNardo, "A propensity score reweighting approach to estimating the partisan effects of full turnout in American presidential elections," *Political Analysis*, 2004, *12* (1), 28–45.

Brunnermeier, Markus K and Jonathan A Parker, "Optimal Expectations," *American Economic Review*, 2005, *95*, 4.

Campbell, Colin M, "Large electorates and decisive minorities," *Journal of Political Economy*, 1999, *107* (6), 1199–1217.

Cascio, Elizabeth U and Ebonya Washington, "Valuing the Vote: The Redistribution of Voting Rights and State Funds Following the Voting Rights Act of 1965," *The Quarterly Journal of Economics*, 2013.

Casella, Alessandra, "Storable votes," *Games and Economic Behavior*, 2005, *51* (2), 391–419.

Citrin, Jack, Eric Schickler, and John Sides, "What if everyone voted? Simulating the impact of increased turnout in senate elections," *American Journal of Political Science*, 2003, *47* (1), 75–90.

Coate, Stephen and Michael Conlin, "A group rule: Utilitarian approach to voter turnout: Theory and evidence," *American Economic Review*, 2004, pp. 1476–1504.

_ , _ , and Andrea Moro, "The performance of pivotal-voter models in small-scale elections: Evidence from Texas liquor referenda," *Journal of Public Economics*, 2008, *92* (3), 582–596.

Degan, Arianna and Antonio Merlo, "A structural model of turnout and voting in multiple elections," *Journal of the European Economic Association*, 2011, *9* (2), 209–245.

DellaVigna, Stefano, John A List, Ulrike Malmendier, and Gautam Rao, "Voting to tell others," Technical Report 2015.

DeNardo, James, "Turnout and the Vote: The Joke's on the Democrats," *American Political Science Review*, 1980, *74* (02), 406–420.

Dhillon, Amrita and Susana Peralta, "Economic theories of voter turnout," *The Economic Journal*, 2002, *112* (480), F332–F352.

Downs, Anthony, *An economic theory of democracy*, New York, 1957.

Duffy, John and Margit Tavits, "Beliefs and voting decisions: A test of the pivotal voter model," *American Journal of Political Science*, 2008, *52* (3), 603–618.

Feddersen, Timothy and Alvaro Sandroni, "A theory of participation in elections," *The American Economic Review*, 2006, pp. 1271–1282.

Feddersen, Timothy J, "Rational choice theory and the paradox of not voting," *Journal of Economic Perspectives*, 2004, pp. 99–112.

Fujiwara, Thomas, "Voting Technology, Political Responsiveness, and Infant Health: Evidence from Brazil," *Econometrica*, 2015, *83* (2), 423–464.

Gordon, Brett R and Wesley R Hartmann, "Advertising effects in presidential elections," *Marketing Science*, 2013, *32* (1), 19–35.

Haile, Philip A and Elie Tamer, "Inference with an incomplete model of English auctions," *Journal of Political Economy*, 2003, *111* (1), 1–51.

Hansford, Thomas G and Brad T Gomez, "Estimating the electoral effects of voter turnout," *American Political Science Review*, 2010, *104* (02), 268–288.

Harsanyi, John C, "Rule utilitarianism, rights, obligations and the theory of rational behavior," *Theory and Decision*, 1980, *12* (2), 115–133.

Hotz, V Joseph and Robert A Miller, "Conditional choice probabilities and the estimation of dynamic models," *The Review of Economic Studies*, 1993, *60* (3), 497–529.

Husted, Thomas A and Lawrence W Kenny, "The Effect of the Expansion of the Voting Franchise on the Size of Government," *Journal of Political Economy*, 1997, pp. 54–82.

Jäckel, Peter, "A note on multivariate Gauss-Hermite quadrature," Technical Report 2005. Sourced from: www.pjaeckel.webspace. virginmedia.com/ANoteOnMultivariateGaussHermiteQuadrature.pdf.

Kanazawa, Satoshi, "A possible solution to the paradox of voter turnout," *The Journal of Politics*, 1998, *60* (04), 974–995.

Karp, Jeffrey A and Susan A Banducci, "Political efficacy and participation in twenty-seven democracies: How electoral systems shape political behaviour," *British Journal of Political Science*, 2008, *38* (02), 311–334.

Kroth, Verena, Valentino Larcinese, and Joachim Wehner, "A Better Life For All? Democratization and Electrification in Post-Apartheid South Africa," 2013.

Lalley, P Steven and E Glen Weyl, "Quadratic Voting," 2015.

Ledyard, John O, "The pure theory of large two-candidate elections," *Public Choice*, 1984, *44* (1), 7–41.

Leighley, Jan E and Jonathan Nagler, *Who Votes Now?: Demographics, Issues, Inequality, and Turnout in the United States*, Princeton University Press, 2013.

Levine, David K and Thomas R Palfrey, "The paradox of voter participation? A laboratory study," *American political science Review*, 2007, *101* (01), 143–158.

Lijphart, Arend, "Unequal Participation: Democracy's Unresolved Dilemma Presidential Address," *American political science review*, 1997, *91* (01), 1–14.

Martinez, Michael D and Jeff Gill, "The effects of turnout on partisan outcomes in US presidential elections 1960–2000," *Journal of Politics*, 2005, *67* (4), 1248–1274.

McDonald, Michael P and Samuel L Popkin, "The myth of the vanishing voter," *American Political Science Review*, 2001, *95* (4), 963–974.

Merlo, Antonio, "Whither political economy? Theories, facts and issues," in Whitney Newey Richard Blundell and Torsten Persson, eds., *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society*, Vol. I, Cambridge: Cambridge University Press, 2006, pp. 381–421.

Miller, Grant, "Women's suffrage, political responsiveness, and child survival in American history," *The Quarterly Journal of Economics*, 2008, *123* (3), 1287.

Minozzi, William, "Endogenous Beliefs in Models of Politics," *American Journal of Political Science*, 2013, *57* (3), 566–581.

Myatt, David, "A rational choice theory of voter turnout," *London Business School*, 2012.

Palfrey, Thomas R and Howard Rosenthal, "A strategic calculus of voting," *Public Choice*, 1983, *41* (1), 7–53.

_ **and** _ , "Voter participation and strategic uncertainty," *American Political Science Review*, 1985, *79* (01), 62–78.

Riker, William H and Peter C Ordeshook, "A Theory of the Calculus of Voting," *American political science review*, 1968, *62* (01), 25–42.

Shachar, Ron and Barry Nalebuff, "Follow the leader: Theory and evidence on political participation," *American Economic Review*, 1999, pp. 525–547.

Smith, Mark A, "The contingent effects of ballot initiatives and candidate races on turnout," *American Journal of Political Science*, 2001, pp. 700–706.

Strömberg, David, "How the Electoral College influences campaigns and policy: the probability of being Florida," *The American Economic Review*, 2008, *98* (3), 769–807.

Tucker, Harvey J and James DeNardo, "Does heavy turnout help Democrats in presidential elections?," *American Political Science Review*, 1986, *80* (04), 1291–1304.

8 Appendix

8.1 Appendix A: Data Construction

In this Appendix, we explain how we construct the joint distribution of demographic characteristics and citizenship status at the county level. We first use the 5% Public Use Microdata Sample of the 2000 U.S. Census (hereafter PUMS), which is an individual-level dataset, to estimate the covariance matrix between the demographic variables and citizenship information within each public use microdata area (PUMA). In particular, we estimate the joint distribution of the discrete demographic characteristics (Race, Hispanic, Citizenship) by counting the frequency of occurence, and then estimate a covariance matrix for the continuous demographic variables (Age, Income, years of schooling) for each bin. Because the PUMA and counties do not necessarily coincide, we estimate covariance matrices for each PUMA and then use the correspondence chart provided in the PUMS website to obtain estimates at the county level.

In the second step, we construct the joint distribution of demographic characteristics by combining the covariance matrix estimated in the first step and the marginal distributions of each of the demographic variables at the county level obtained from Census Summary File 1 through File 3. We discretize continuous variables into coarse bins. We discretize age into 3 bins, income into 6 bins: (1) \$0-\$25,000, (2) \$25,000-\$50,000, (3)\$50,000-\$75,000, (4) \$75,000-\$100,000, (5) \$100,000-\$150,000, (6) \$150,000-) and years of schooling into 5 bins: (1) Less than 9th grade, (2) 9th-12th grade with no diploma, (3) highschool graduate, (4) some college with no degree or associate degree, (5) bachelor degree or higher. so that there are 1,620 bins in total. The joint distribution of demographic characteristics that we create gives us a probability mass over each of the1,620 bins for each county.

Finally, we augment the census data with religion data obtained from Religious Congregations and Membership Study 2000. This data has information on the share of the population with adherence to either “Evangelical Denominations” or “Church of Jesus Christ of Latter-day Saints” at the county level. Because the Census does not collect information on religion, we do not know the correlation between the religion variable and the demographic characteristics in the Census. Thus, we assume independence of the religion variable and other demographic variables. As a result, there are 3,240 bins in our demographics distribution.

8.2 Appendix B: Identification of $c(\cdot)$, $p(\cdot)$, and F_η in the general case

In this Appendix, we show that $c(\cdot)$, $p(\cdot)$, and F_η are identified even when the max operator in equation (11) binds with positive probability. Note that our argument in the main text considered only the case in which the max operator never binds. Recall that

$$\underbrace{\frac{\Phi^{-1}\left(1-v_{D,m}\right)-\Phi^{-1}\left(v_{R,m}\right)}{2}}_{\equiv Y_m} = \max\left\{0, \frac{c(\bar{\mathbf{x}}_m)}{p(\bar{\mathbf{x}}_m)} + \frac{\eta_m}{p(\bar{\mathbf{x}}_m)}\right\}, \ \eta_m \perp \bar{\mathbf{x}}_m. \tag{16}$$

In this appendix, we work with the normalization that the value of $p(\cdot)$ at some $\bar{\mathbf{x}}_m = \mathbf{x}_0$ as $p(\mathbf{x}_0) = 1$. This amounts to a particular normalization of variance of η. Note that the distribution of Y_m (the left hand side of equation (16)) conditional on $\bar{\mathbf{x}}_m = \mathbf{x}_0$ is a truncated distribution with mass at zero. Figure 7 illustrates this when the mass at zero is less than 50% and F_η is symmetric and single peaked at zero.

First, we present our identification discussion for the case that F_η is symmetric and single peaked at zero. As Figure 7 illustrates, the median of Y_m conditional on $\bar{\mathbf{x}}_m = \mathbf{x}_0$ directly identifies $c(\mathbf{x}_0)$ under these assumptions. Also, the density of η, f_η, is identified above the point of truncation. Formally, $f_\eta(F_\eta^{-1}(t))$ is identified for any $t > t(\mathbf{x}_0)$, where

$$t(\mathbf{x}_0) = \Pr\left(Y_m = 0 | \, \mathbf{x}_0\right).$$

Hence, $f_\eta(0)$ is identified from the height of the density of Y_m at the median.

Now consider $\mathbf{x}_1 \neq \mathbf{x}_0$. Assume again that $t(\mathbf{x}_1) < 0.5$. Then, $c(\mathbf{x}_1)/p(\mathbf{x}_1)$ is identified from the conditional median of Y_m and $p(\mathbf{x}_1)f_\eta(0)$ is identified by the height of the conditional density of Y_m at the median. Given that $f_\eta(0)$ is identified, $c(\mathbf{x}_1)$ and $p(\mathbf{x}_1)$ are both identified. Moreover, F_η is identified over its full support if there exists sufficient variation in $\mathbf{x}$, i.e., $\inf_{\mathbf{x}} t(\mathbf{x}) = 0$

We now consider the case in which F_η is not restricted to be symmetric and single peaked and $t(\mathbf{x}_0)$ may be less than 0.5. The distribution of Y_m is identified above $t(\mathbf{x}_0)$ as before. Now consider $\mathbf{x}_1 \neq \mathbf{x}_0$. Similar as before, we identify $p(\mathbf{x}_1)f_\eta(F_\eta^{-1}(\tau))$ for τ above $t(\mathbf{x}_1)$.[28] If we let τ be any number larger than $\max\{t(\mathbf{x}_0), t(\mathbf{x}_1)\}$, both $f_\eta(F_\eta^{-1}(\tau))$ and $p(\mathbf{x}_1)f_\eta(F_\eta^{-1}(\tau))$ are identified. Hence $p(\mathbf{x}_1)$ is identified. Similarly, $p(\cdot)$ is identified for all $\mathbf{x}$.

We now consider identification of $c(\cdot)$. We present two alternative assumptions on F_η and show that $c(\cdot)$ can be identified under either assumption. First, assume that the median of η is zero, $Med(\eta) = 0$, and that there exists $\mathbf{x} = \mathbf{x}_2$ such that

[28] Note that we identify $f_{\eta/p(\mathbf{x}_1)}\left(F^{-1}_{\eta/p(\mathbf{x}_1)}(t)\right)$, where $f_{\eta/p(\mathbf{x}_1)}(\cdot)$ and $F^{-1}_{\eta/p(\mathbf{x}_1)}(\cdot)$ are the density of $\eta/p(\mathbf{x}_1)$ and the inverse distribution of $\eta/p(\mathbf{x}_1)$, respectively. Note that $f_{\eta/p(\mathbf{x}_1)}\left(F^{-1}_{\eta/p(\mathbf{x}_1)}(t)\right) = p(\mathbf{x}_1)f_\eta(F_\eta^{-1}(t))$.

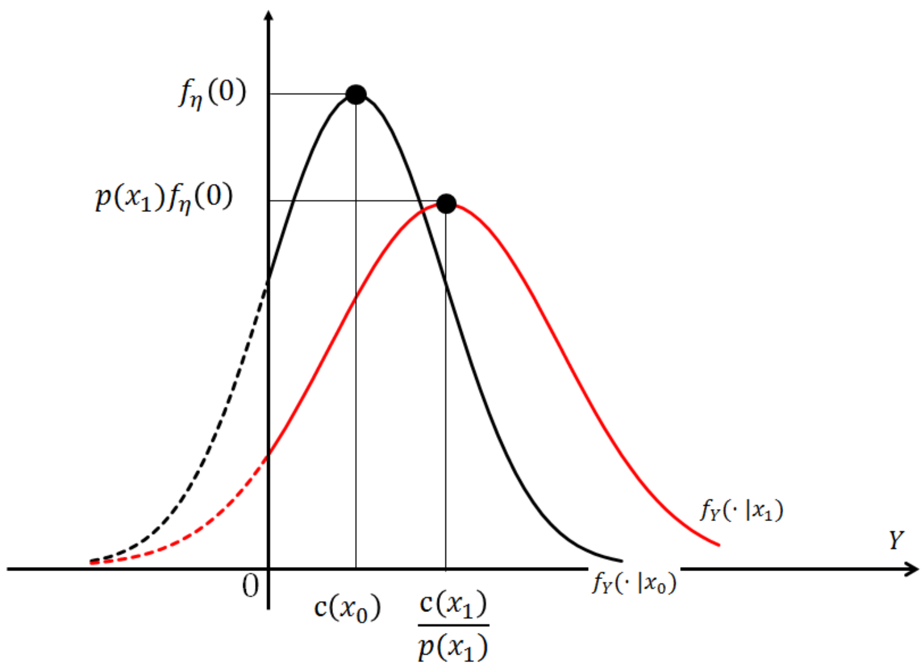

Figure 7: The distribution of Y_m conditional on $\mathbf{x} = \mathbf{x}_0$ and $\mathbf{x} = \mathbf{x}_1$ when the distribution of η is symmetric and single peaked, and $t(\mathbf{x}_0)$, $t(\mathbf{x}_1) < 0.5$, where $t(\mathbf{x})$ is the probability that Y_m is equal to zero conditioanl on $\mathbf{x}$. Note that the distribution of Y_m is truncated at zero. The conditional median of Y_m identifies $c(\mathbf{x}_0)$ and $c(\mathbf{x}_1)/p(\mathbf{x}_1)$, and the height of the density at the conditional median identifies $f_\eta(0)$ and $p(\mathbf{x}_1)f_\eta(0)$.

$t(\mathbf{x}_2) < 1/2$. The latter assumption means that more than half of the counties have turnout less than 100% when $\mathbf{x} = \mathbf{x}_2$. Then, the median of Y_m conditional on $\mathbf{x}_2$ identifies $c(\mathbf{x}_2)/p(\mathbf{x}_2)$. Now consider any $\mathbf{x}_1 \neq \mathbf{x}_2$ and let τ be any number larger than $\max\{t(\mathbf{x}_2), t(\mathbf{x}_1)\}$. Let z_1 and z_2 be the τ quantile of Y_m conditional on $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively. z_1 and z_2 are clearly identified. Then,

$$
\begin{aligned}
& p(\mathbf{x}_1)\left[F^{-1}_{\eta/p(\mathbf{x}_1)}(\tau) - F^{-1}_{\eta/p(\mathbf{x}_1)}(1/2)\right] = p(\mathbf{x}_2)\left[F^{-1}_{\eta/p(\mathbf{x}_2)}(\tau) - F^{-1}_{\eta/p(\mathbf{x}_2)}(1/2)\right] \\
& \Leftrightarrow p(\mathbf{x}_1)\left[z_1 - \frac{c(\mathbf{x}_1)}{p(\mathbf{x}_1)}\right] = p(\mathbf{x}_2)\left[z_2 - \frac{c(\mathbf{x}_2)}{p(\mathbf{x}_2)}\right] \\
& \Leftrightarrow \frac{c(\mathbf{x}_1)}{p(\mathbf{x}_1)} = z_1 - \frac{p(\mathbf{x}_1)}{p(\mathbf{x}_2)}\left(z_2 - \frac{c(\mathbf{x}_2)}{p(\mathbf{x}_2)}\right).
\end{aligned}
\tag{17}
$$

Given that all of the terms on the right hand side of (17) are identified, $c(\mathbf{x}_1)/p(\mathbf{x}_1)$ is identified.

Alternatively, assume that $\mathbf{E}(\eta) = 0$ and $\inf_{\mathbf{x}} t(\mathbf{x}) = 0$. We now show that $c(\cdot)$ is identified under these alternative assumptions. Intuitively, this latter assumption means that there exist values of $\mathbf{x}$ for which the max operator is never binding. In this case, we can fully recover the distribution of $F_\eta(\cdot)$. Then we can identify the distribution of $c(\mathbf{x})/p(\mathbf{x}) + \eta_m/p(\mathbf{x})$ for any $\mathbf{x}$. Hence we identify $c(\cdot)/p(\cdot)$.