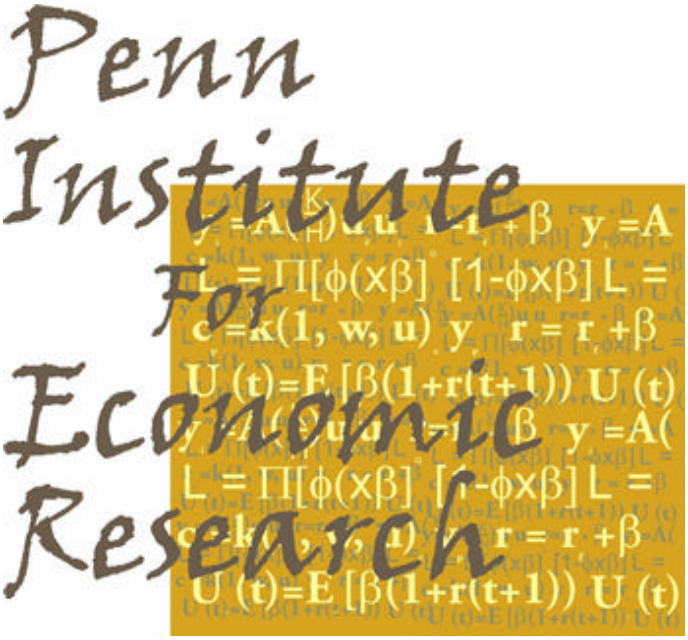


Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://www.econ.upenn.edu/pier

PIER Working Paper 04-015

"A Dynamic Model of Voting"

by

Arianna Degan

http://ssrn.com/abstract=534822

A Dynamic Model of Voting*

by Arianna Degan
University of Rochester and UQAM†

Abstract

We propose and estimate a dynamic model of voting with asymmetric information incorporating the three main factors affecting voting choices of individual citizens: party identification, policy preferences, and candidates' valence. Using individual-level data on voting decisions in two consecutive presidential elections, we identify and estimate (1) the distribution of voters' policy positions and (2) candidates' valence. In addition to providing an equilibrium interpretation of the observed voting profiles and electoral outcomes, we use the estimated model to conduct counterfactual experiments to assess the relative importance of candidates' policy positions, valence, and voters' information on the outcomes of elections and to evaluate the performance of the electoral process.

Keywords: party identification, policy preferences, consecutive elections, valence.

Correspondence Address: Arianna Degan. Département des sciences économiques, Université du Québec à Montréal, UQAM, CP 8888 Succursale Centre-Ville, Montréal , Québec, CANADA, H3C 3P8. E-mail: arianna@ssc.upenn.edu

*I thank Antonio Merlo for guidance, encouregment, and many helpful comments and suggestions. I also thank seminar participants at the University of Pennsylvania, in particular Bob Inman, for useful comments and suggestions. Finally, I thank Keith Poole for his assistance with the NOMINATE coordinates data.

†Visiting Assistant Professor at the Wallis Institute of Political Economy, University of Rochester. Assistant Professor at the Univérsité du Quebéc à Montréal.

1 Introduction

In representative democracies, where elected politicians make policy-relevant decisions on behalf of their constituents, voting is one of the main ways citizens participate in the political process. Citizens' voting choices determine who is elected, therefore may contain important information on citizens' political preferences.

The fact that most elections take place repeatedly over time suggests us that having observations on how the same individuals vote in consecutive elections provides a variation in voting data that, in addition to the cross-sectional variation, may be particularly useful to identify and estimate citizens' policy preferences as well as other unobservable objects of interest.

The following observations emerge from voting data from two consecutive U.S. presidential elections. First, we observe all possible voting profiles: individuals who vote for the democratic or republican candidate in both elections and individuals who vote for the democratic candidate in the first election and for the republican candidate in the second election or vice versa. Second, "voting persistence" and "switching behavior" are both quantitatively significant phenomena: while a large majority of voters, around 80 percent, vote for the same party's candidate in two consecutive elections, the remaining 20 percent vote for candidates of different parties in two consecutive elections. Third, voting patterns differ by party identification: voters who identify with a particular party are more likely to vote for that party's candidate, but the extent of "party loyalty" and switching behavior varies across parties.

Many researchers in political science have focused on characterizing the main determinants of voting.[1] The consensus view is that voting choices of individual citizens are typically affected by three factors: party identification (that is, a voter's attachment to a particular party), policy preferences, and candidates' valence (that is, candidates' personal characteristics about which everybody has the same pref-

[1] See, e.g., Campbell et al. (1960), RePass (1971), Jackson (1975), Jones and Page (1979), and Markus and Converse (1979).

erence such as honesty, charisma, integrity, trustworthiness, or leadership). While voters will in general differ with respect to their policy and party preferences, they will, by definition of valence, agree that candidates with relatively higher valence are preferable.[2]

In this paper we propose and estimate a dynamic model of voting with asymmetric information incorporating the three aspects of individual voting behavior. Using individual-level data on voting decisions in two consecutive presidential elections, we identify and estimate (1) the distribution of voters' policy positions and (2) candidates' valence. In addition to providing an equilibrium interpretation of the observed voting profiles and electoral outcomes, we use the estimated model to conduct counterfactual experiments to assess the relative importance of candidates' policy positions, valence, and voters' information on the outcomes of elections and to evaluate the performance of the electoral process.

We consider a two-period model of voting where in each period there are two candidates running for the presidency. Each candidate has an exogenous policy position and valence which are both constant over time. The incumbent (that is, the candidate who wins the election in the first period) runs for reelection in the second period and faces a new challenger. There is a continuum of voters who care about both the policy position and valence of the winning candidate in each election. While voters observe candidates' policy positions they do not observe their valence. Voters are heterogeneous with respect to their demographic characteristics, party identification, information status, and policy preferences.

Party identification enters our model in two different ways. First, we allow the distribution of voters' policy positions to differ by party identification. Second, we assume that party identification impacts access to information, in particular, voters

[2]For example, according to Hinich (1982b) valence includes characteristics beyond a candidate's immediate control that are unrelated to policy and affect voters' evaluations of candidates. See also Stokes (1963, 1992), Ansolabehere and Snyder (2000). Our definition of valence does not include a candidate's ability to implement specific policies, in which case voters may disagree that candidates with relatively higher valence are preferable.

with different party identifications may receive different signals on candidates' valence.

The model has a unique equilibrium with sincere voting. The equilibrium strategy prescribes a different voting behavior for each possible combination of candidates' valence in the two elections and implies a probability distribution over voting choices in the two periods. The equilibrium electoral outcome depends on candidates' policy positions and valence, voters' preferences, and the degree of asymmetric information, and it reveals information on the candidates' valence.

We structurally estimate our model using individual-level voting choices in the 1968 and 1972 U.S. presidential elections from the 1972 Center for Political Studies survey data. The estimates of the model allow us to quantify the effect of individual characteristics on voters' policy preferences and provide insight on the relation between demographics, party identification, and political views of American citizens. For example, we find that some characteristics have a similar effect on voters' policy preferences regardless of their party identification (e.g., blacks are more liberal than non-blacks), whereas other characteristics affect voters' policy preferences differently depending on their party identification (e.g., while more educated Democrats are relatively more liberal, more educated Republicans tend to be relatively more conservative). Our results suggest the importance of breaking down the aggregate relationship between demographics and policy preferences by considering their interaction with party identification.

Our estimates of the valence of the presidential candidates in 1968 and 1972 indicate that Humphrey (the democratic candidate in 1968) had high valence, McGovern (the democratic candidate in 1972) had low valence, and Nixon (the republican candidate in 1968 and again, as incumbent President, in 1972) also had low valence. This result is perhaps surprising given that Nixon won the 1968 election by a very small margin and then won again 1972 by a margin as large as that gained by Johnson in 1964 or Roosevelt in 1936. However, the results are consistent with anecdotal accounts of the events surrounding the 1968 and 1972 elections.

The two main counterfactual experiments allow us to assess the effect of primaries elections and information on electoral outcomes. The findings can be summarized as follows. First, while in 1968 none of the democratic candidates participating in the democratic presidential primaries could have defeated Nixon, virtually all the participants in the 1972 democratic primaries (other than McGovern) could have won the presidency. Second, had all voters been aware that Nixon was a low-valence candidate, Humphrey would have been elected President in 1968.

Before turning our attention to the description of the model, we will briefly discuss the relationship of our work to the existing literature. First, most of the existing empirical literature on voting estimates voters' policy preferences using data on individual self-reported attitudes towards policies and candidates (see, e.g., Cahoon, Hinich and Ordeshook 1978; Rabinowitz 1978; Enelow and Hinich 1984; Poole and Rosenthal 1984; Brady 1990; and Poole 1998). In contrast to these studies, we estimate the distribution of voters' policy preferences using their observed voting behavior, given their individual characteristics. Our approach relies on a revealed preference argument that identifies fundamental utility parameters from observed optimal choices. In this respect it is analogous to the approach used by Heckman and Snyder (1997), Poole and Rosenthal (1997), Londregan (2000), and Bailey (2001) to estimate legislators' policy preferences from observed roll call voting.[3] Since one may argue that self-reported measures of "proximity" to a particular policy or candidate are subject to the so called "projection" and "persuasion" bias (see, e.g., Markus and Converse 1979), and are not interpersonally comparable (see, e.g., Brady 1989), instead of using this information as an input in the estimation we use it to externally validate our empirical results.

Second, most of the previous empirical analyses of voting focus on single elections and abstract from the estimation of candidates' valence (see, e.g., Cahoon, Hinich, and Ordeshook 1978; Rabinowitz 1978; Enelow and Hinich 1984; Poole and Rosenthal

[3]Lewis (2001) estimates the distribution of voters' ideal points using individual-level voting returns on ballot propositions.

1984; Brady 1990; and Poole 1998).[4] An important innovation of our analysis is that by using the additional information contained in the sequence of voting choices by the same individuals in two consecutive elections, we can simultaneously estimate the distribution of voters' policy preferences and candidates' valence. The three main features of the data allowing us to separately identify these two objects are: the variation in the data generated by repeated voting; the fact that different individuals face the same candidates in each election; and the fact that the candidate who wins the first election also runs for office in the second election.

While in this paper we focus on the 1968-72 elections, our methodological approach is quite general and can be used to address a variety of issues emerging from voting data from multiple elections (both elections for the same public office over time and concurrent elections for different public offices).

In section 2 we describe the model, in section 3 we describe the equilibrium, in section 4 we discuss the modeling assumptions, in section 5 we describe the empirical analysis, in section 6 we conduct counterfactual experiments, and in section 7 we conclude.

2 The Model

There are two periods, 1 and 2. In each period there are two candidates running for President, D and R, where D denotes the democratic candidate and R the republican candidate. Each candidate $c \in \{D, R\}$ is characterized by a one-dimensional policy position $y_c \in [-1, 1]$, $y_D < y_R$, and valence $x_c \in \{L, H\}$, $L < H$, where the policy space $[-1, 1]$ is the traditional liberal-conservative space and L and H denote low and high valence respectively. Both y_c and x_c are exogenously given and fixed. In period

[4] Sachar (2000) uses data on repeated voting to estimate a model of habit persistence. Cahoon, Hinich, and Ordeshook (1978), and Enelow-Hinich (1984) correct their estimation procedure to account for valence, but they do not estimate valence simultaneously with voters' ideal points. Brady (1990) proposes a method to estimate a general ideal-points model that allows him to jointly estimate the distribution of voters' ideal points and that part of the average evaluation of candidates unrelated to the spatial dimension.

2 the incumbent President (that is, the candidate who wins the election in period 1) runs for reelection and faces a new challenger.

There is a continuum of voters distributed in the interval $[-1, 1]$. We index each voter by j. Voters observe the candidates' policy positions, but they do not observe their valence. However, voters know the distribution of valence in the population of potential candidates, and we let $q \in (0, 1)$ be the probability that a candidate has high valence.[5]

Each voter j has exogenous demographic characteristics $W_j \in W$, where W is the space of demographic characteristics. In addition to differing with respect to their demographic characteristics, voters are heterogeneous along three dimensions which we label as *party identification*, *information status,* and *policy preferences*.[6] Each voter j has an exogenous party identification, $k_j \in K = \{d, r, i\}$. Specifically, when a voter has a democratic party identification, $k_j = d$, it means he considers himself a Democrat; when a voter has a republican party identification, $k_j = r$, it means he considers himself a Republican; and when a voter has an independent party identification, $k_j = i$, it means he does not feel attached to any particular party, or, equivalently, he considers himself an Independent. We will alternatively say that voter j has party identification k or that voter j's group is k, where k denotes an element of K. The proportion of voters belonging to k is n_k ($n_k \in [0, 1], \sum_{k \in \{d,r,i\}} n_k = 1$).

Party identification affects access to information. Each voter j from group k has a probability m_k of becoming informed and a probability $(1 - m_k)$ of remaining uninformed. We let $I_j \in \{0, 1\}$ denote voter j's information status, where I_j takes the value 1 when the voter becomes informed and 0 when he remains uninformed. Information status is fixed during the two periods: that is, if a voter is informed

[5] In this paper we abstract from political competition and assume that voters condition their voting choices on the candidates' policy positions. Typically candidates for the presidency have served in either the House of Representatives or the Senate. Estimation of legislators' policy positions are available in the data. See section 5.

[6] Although party identification is exogenous in our model, because of its particular role in the model we do not classify it among the other individual characteristics.

in period 1 he will also be informed in period 2. Informed voters receive a signal about candidates' valence in each period. Party identification affects not only the probability of becoming informed but also the type of information received. Let S_k denote the signal space for a voter from group k in each period, and let s_0^t and s_k^t denote the signal received at time t by an uninformed and an informed voter from group k respectively. We assume that an uninformed voter, regardless of his party identification, does not receive any signal, $s_0^t = \{0, 0\}$; an informed voter with a democratic party identification receives a perfect signal about D's valence, $s_d^t = \{x_D^t, 0\}$; an informed voter with a republican party identification receives a perfect signal about R's valence, $s_r^t = \{0, x_R^t\}$; and an informed voter with an independent party identification receives a perfect signal about both candidates' valence, $s_i^t = \{x_D^t, x_R^t\}$.

The idea that party identification works as an information selection device is not completely new.[7] For example, Fiorina (1981, p.81) writes, "All individuals do not receive random samples of political information. One's party identification is no doubt associated with these kind of differences in receipt of information." We model the idea that voters with different party identifications have asymmetric political information by assuming that people who feel attached to a party are more likely to be informed about that party's candidate.[8] For tractability of the model we assume that party identification is exogenous, and in addition we assume that it is fixed.[9] These assumptions can be partially justified by the fact that we restrict our analysis to short-term dynamics.[10]

[7] See, e.g., Campbell et al. (1960), RePass (1971), Fiorina (1977, 1981), and Franklin and Jackson (1983).

[8] This assumption can be justified on the grounds that, typically, individuals with a partisan party identification go to their party's conventions, they read partisan newspapers, and they are more likely to have friends of the same party with whom they discuss their party's candidates.

[9] In our dataset party identification is measured only once. See section 5.1.

[10] Party identification is supposed to capture a long term attachment to a party. According to the view of the earliest "Michigan School" (Campbell et al., 1960), party identification is strictly related to individual socio-demographic characteristics. More recently, Jennings, Stoker, and Bowers (1999) find that party identification and, more generally, individual political preferences are mainly affected by political socialization within the family. We are aware that party identification is also subject to

Each voter j has a preferred policy position or ideal point y_j, which takes value in the interval $[-1, 1]$. We can characterize the distribution of policy positions of voters from group k by the function $Y_k(\cdot|W)$ with domain $[-1, 1]$. The assumption that the distribution of voters' ideal points has full support regardless of party identification is justified by evidence that voters' self-placement along a liberal-conservative scale is distributed on the full support of such a scale irrespective of voters' party identification.[11]

Voter j's utility in period t when candidate c is elected depends on both the distance between his ideal point y_j and the candidate's policy position y_c^t and on the candidate's valence x_c^t:

$$U_j^c = U(x_c^t, y_c^t, y_j) = \lambda x_c^t - (y_j - y_c^t)^2$$

where λ is the relative weight that all voters assign to valence.

Besides knowing the candidates' policy positions at the beginning of each period ($\{y_D^1, y_R^1\}$ and $\{y_D^2, y_R^2\}$) and candidates' distribution of valence in the population (q), voters know the signals' structure, the distribution of voters across party identifications, and the group specific distribution of ideal points (that is, they know m_k, $s_k^1(.)$, $s_k^2(.)$, n_k, $Y_k(.)$, $\forall k \in K$).

We abstract from abstention and we assume that all citizens vote sincerely in each period: given their party identification, information status, ideal point, signal, the two candidates' policy positions, and their beliefs about candidates' valence, citizens vote for the candidate who, if elected, gives them the highest expected utility. As a tie-breaking rule we assume that when a voter is indifferent between the two candidates he votes for each of them with equal probability.

---

short-term variations and that even its long-term component changes over time (especially in the years covered by our analysis). However, Green (1990) has shown that once measurement errors are taken into account, party identification appears to be very stable. He also suggests that "it may be a useful simplification to regard party identification as exogenous with respect to variables such as voting behavior, candidate evaluations, issues proximity, and retrospective performance evaluation."

[11] See footnote 50 for self-reported placement on a fixed-point liberal-conserative scale from the 1972 CPS survey and the summer 1969 Gallup opinion polls.

We can summarize the timing of the events as follows. Voters' party identification and information status are known before the beginning of period 1. At the beginning of period 1, all voters observe the identity of the two competing candidates (that is, who is R and who is D) and their policy positions; in addition, the informed voters receive a signal about the two candidates' valence. During period 1 they vote, and at the end of the same period they observe the outcome of the election.[12] Voters do not directly observe the winner's valence, but they update their beliefs using the information contained in period 1's electoral outcome and signal.[13] At the beginning of period 2 voters observe the identity of the new challenger and his policy position. In addition, the informed voters receive another signal on the candidates' valence.[14] During period 2 all voters vote, and at the end they observe the outcome of the second election.

3 Strategies and Equilibrium

A voting strategy for voter j with party identification $k_j = k$ and ideal point y_j is a pair of voting rules (v^1_{jk}, v^2_{jk}) which assign to voter j the candidate to vote for in each period. The voting rule in each period t is a mapping from the set of relevant information about candidates for voters from group k in period t, Ω^t_k, to the set of voting choices $\{D, R\}$. The voting rule in period 1 is $v^1_{jk} : \Omega^1_k \longrightarrow \{D, R\}$ and the voting rule in period 2 is $v^2_{jk} : \Omega^2_k \longrightarrow \{D, R\}$, where $\Omega^1_k = S_k \times \{0, 1\} \times [-1, 1]^2$, $\Omega^2_k = \{S_k\}^2 \times \{0, 1\} \times [-1, 1]^2 \times [-1, 1]^2 \times \Pi^1_D$, and $\Pi^1_D \subseteq [0, 1]$ is the set of period 1's vote shares for candidate D.[15] We denote by $\omega^t_{jk} \in \Omega^t_k$ the information about

[12] The vote share is a sufficient statistic for the electoral outcome.

[13] If instead we assume that voters do not know the distribution of policy preferences in the population but can observe the winning candidate's valence through his behavior while in office, as it will be clear after reading section 3, the equilibrium voting behavior for each state of the world wouldn't be affected.

[14] The signal is redundant for the informed voters of the same party as the incumbent.

[15] In period 1 a voter's information about the candidates is given by his signal in period 1, his information status, and the observed policy positions of the two candiates in period 1. In period 2 a voters' information about the candidates is given by his signals in both periods, his information status, the candidates' policy positions in both periods, and period 1's vote share.

candidates available to voter j from group k in period t, where $\omega^1_{jk} = (s^1_k, I_j, y^1_D, y^1_R)$ and $\omega^2_{jk} = (s^1_k, s^2_k, I_j, y^1_D, y^1_R, y^2_D, y^2_R, \pi^1_D)$.

Let $u_j(x^t_D, x^t_R, y^t_D, y^t_R, y_j) = U^R_j - U^D_j = \lambda(x^t_R - x^t_D) + [(y_j - y^t_D)^2 - (y_j - y^t_R)^2]$ denote the difference in voter j's utility when candidate R is elected rather than candidate D, and let $\{P_{k,I}\}$ denote the beliefs system about the distribution of candidates' valence of a voter from group k and information status I .

Proposition 1: *The unique equilibrium strategy with sincere voting* $\{(v^{1*}_{jk}, v^{2*}_{jk})\}_{k=\{d,r,i\}}$ *is characterized as follows. For any voter* j *with party identification* $k \in K$ *and with policy position* $y_j \in [-1, 1]$*:*

$$v^{1*}_{jk}(\omega^1_{jk}) = \begin{cases} R & \text{if } E[u_j(x^1_D, x^1_R, y^1_D, y^1_R, y_j)|\omega^1_{jk}, y_j] > 0 \\ D & \text{if } E[u_j(x^1_D, x^1_R, y^1_D, y^1_R, y_j)|\omega^1_{jk}, y_j] < 0 \\ \frac{1}{2}, \frac{1}{2} & \text{if } E[u_j(x^1_D, x^1_R, y^1_D, y^1_R, y_j)|\omega^1_{jk}, y_j] = 0 \end{cases}$$

and

$$v^{2*}_{jk}(\omega^2_{jk}) = \begin{cases} R & \text{if } E[u_j(x^2_D, x^2_R, y^2_D, y^2_R, y_j)|\omega^2_{jk}, y_j] > 0 \\ D & \text{if } E[u_j(x^2_D, x^2_R, y^2_D, y^2_R, y_j)|\omega^2_{jk}, y_j] < 0 \\ \frac{1}{2}, \frac{1}{2} & \text{if } E[u_j(x^2_D, x^2_R, y^2_D, y^2_R, y_j)|\omega^2_{jk}, y_j] = 0 \end{cases}$$

where the expectation is taken with respect to the system of beliefs $\{P_{k,I}\}$ *which is calculated using Bayes Rule.*

The proof is trivial and follows directly from sincere voting. Although voters are sincere and myopic, the two elections are linked by the effect of the aggregate outcome on beliefs and the fact that the incumbent's type and policy position as well as voter's information status are constant over time.

Notice that even though the equilibrium strategy is unique, an individual's actual voting choice depends on the realized information status and on which of the finite and discrete combinations of candidates' valence is realized. Let $X = \{HHHH, HHHL, HHLH, HLHH, HLHL, HLLL, LHLL, LHLH, LHHH, LLLL, LLLH, LLHL\}$ denote the set of states of the world, with $x = \{x^1_D x^1_R x^2_D x^2_R\}$ as its generic element. The result of proposition 1, together with the assumptions that the utility is quadratic with respect to the difference in policy positions and that candidates'

valence follows a Bernoulli distribution, can be used to derive a simpler and more useful characterization of the equilibrium strategy.

Proposition 2: (i) *The equilibrium strategy prescribes a cut-off voting rule in each period t. For any voter j, party identification k, ideal point y_j, information on candidates $\omega_{jk}^t \in \Omega_k^t$, there is a unique cut-off point in each period t, $\overline{y}^t(\omega_{jk}^t)$ such that: if $y_j < \overline{y}^t(.)$ it is optimal to vote for D, if $y_j > \overline{y}^t(.)$ it is optimal to vote for R, if $y_j = \overline{y}^t(.)$ the voter is indifferent and will vote for $D(R)$ with probability $\frac{1}{2}$.*
(ii) *Each cut-off point has the form: $\overline{y}^t(\omega_{jk}^t) = m_t - \frac{\lambda\Delta_j^t}{g_t}$, where*
$\Delta_j^t = E[(x_R^t - x_D^t)|\omega_{jk}^t]$ *is the expected difference in valence between candidate R and D in period t perceived by individual j;*
$m_t = \frac{y_R^t + y_D^t}{2}$ *is the "midpoint" of the two candidates' policy positions; and*
$g_t = 2(y_R^t - y_D^t)$ *is the "gap" between the two candidates' policy positions in period t.*

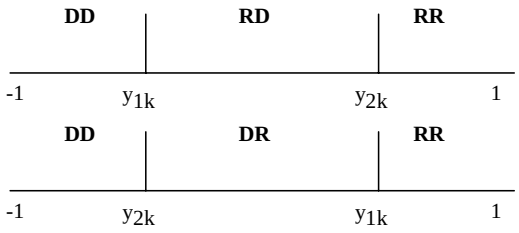
Proof: Voter j's optimal strategy is to vote for $R(D)$ if $Eu_j > 0 (< 0)$. Voter j will vote for $R(D)$ if $\lambda\Delta_j^t + [(y_j - y_D^t)^2 - (y_j - y_R^t)^2] > 0$ (< 0). Solving for j's ideal point, voter j will vote for R (D) in period t if $y_j > \frac{y_R^t + y_D^t}{2} - \frac{\lambda\Delta_j^t}{2(y_R^t - y_D^t)}$ $(<)$.

We use proposition 2 to analyze the possible voting profiles that emerge in equilibrium. Let $V = \{RR, RD, DR, DD\}$ be the set of dynamic optimal voting profiles, where RR denotes the profile of a voter who votes for the republican candidate in two consecutive elections, RD denotes the profile of a voter who votes for the republican candidate in period 1 and for the democratic candidate in period 2, and so on. Proposition 2 tells us that if a voter's ideal point in any period is to the left of his cut-off point, he will vote for D, and if it is to the right of his cut-off point, he will vote for R. It follows that a voter's dynamic voting choice is determined by the locations of his cut-off points in the two periods and their relationship with respect to his ideal point. Let y_{tk} and y_{t0} be the cut-off points at time t for an informed voter from group k and for an uninformed voter respectively.[16] Take an informed voter j

[16]The cut-off points of informed voters are independent of their party identification.

with party identification k and ideal point y_j. When $y_{2k} < y_j < y_{1k}$ his voting profile will be DR, when $y_{1k} < y_j < y_{2k}$ his voting profile will be RD (figure 1). Analogously, the uninformed voter will generate the profile DR when $y_{20} < y_j < y_{10}$ and RD when $y_{10} < y_j < y_{20}$.

Figure 1. Cut-off points and voting profiles: informed voters from group k



A first observation is that only the elements which enter in the expression for the cut-off points are relevant to the equilibrium strategy. Voter j's cut-off points depend on the candidates' policy positions and on his information (both the one contained in his signals and in period 1′s electoral outcome). In each period t, any cut-off point can be expressed simply as a combination of the midpoint, m_t; the gap, g_t; and the weighted expected difference in valence, $\lambda\Delta_j^t$. For any information status and signal it is easy to show that $\lambda\Delta_j^t$ is a linear function of $\lambda(H-L)$ only. This implies that the weight on valence λ and the perceived maximum difference in valence $(H-L)$ never enter separately in the equilibrium characterization.

A second observation is that the model puts restrictions on which state is potentially compatible with the observed voting patterns. For example, if the cut-off points related to a particular state and candidates' positions are such that, for some k, $y_{1k} > y_{2k}$ and $y_{10} > y_{20}$, then we can conclude that such a state is incompatible with observing all the voting profiles within group k. This is because voters in group k would only generate the profiles DD, DR, and RR.[17] In general, to generate both

[17]This situation occurs for example when $y_D^1 > y_D^2$, and candidate R is the winner of the first election (which corresponds to having $m_1 > m_2$ and $g_1 < g_2$). In this case the observed voting profiles would be incompatible with a state where all candidates have a high valence (state $HHHH$), and

switching patterns (DR and RD) within each group, we need the cut-off points of informed and uninformed voters to be such to generate different switching behaviors (that is, it must be that either $y_{2k} < y_{1k}$ and $y_{20} > y_{10}$ or $y_{2k} > y_{1k}$ and $y_{20} < y_{10}$, $\forall k$).[18] Proposition 2 also allows us to derive the following result about electoral outcomes.

Proposition 3: *In equilibrium* π_D^1 *completely reveals the valence of period* 1's candidates.

The intuition is that for any possible state, the equilibrium vote share is uniquely determined by the fraction of voters with different party identifications, their probability of being informed, and the distribution of policy positions. Voters know n_d, n_r, n_i, m_d, m_r, m_i, Y_d, Y_r, Y_i and the signal structure. Since by assumption in any group k at each policy point with positive mass there is a fraction of voters m_k that becomes informed, voters can perfectly calculate the vote share corresponding to any realization of the two candidates' valence in period 1. Such shares will be different for any valence pair, so by observing π_D^1 voters know both candidates' valence.[19]

A corollary to proposition 3 is that π_D^1 completely reveals the valence of the incumbent. That the outcome of the first period's election reveals the valence of the incumbent is what makes the model an equilibrium model instead of just a simple individual decision making problem. As a consequence, some voters - the ones that in the first period do not receive a signal on the valence of the candidate that will win the election - in equilibrium will use the information contained in the electoral

this incompatibility would arise from the behavior of Independents. Specifically, the cut-off points of an informed Independent would be $y_{1i} = m_1$ and $y_{2i} = m_2$, while the cut-off points of an uninformed Independent would be $y_{10} = m_1$ and $y_{20} = m_2 - \frac{z}{2g_2}$. Since $y_{1i} > y_{2i}$ and $y_{10} > y_{20}$, Independents in equilibrium will not generate RD.

[18]The reader can find in Degan (2003) the details about which states are compatible with different configurations (that is, pair of candidates' policy positions in the two periods and identity of the incumbent), observing all four voting profiles, and the corresponding cut-off points.

[19]For the proof see Appendix A. There is a particular combination of parameters where uninformed voters cannot distinguish among states where in period 1 both candidates have either a low or a high valence.

outcome and, more in general, all voters in period 2 will condition their voting choices on the incumbent's valence.

4 Discussion of Modeling Choices

As mentioned in the introduction, there are two key features emerging from voting data from two consecutive elections.[20] First and most importantly, we observe all four voting profiles (DD, DR, RD, and RR) both at the aggregate level and within each group k and each profile constitutes a statistically significant phenomenon. Second, the four voting profiles are heterogeneous across groups, that is, the frequency of each profile changes with party identification. Before turning to the empirical analysis, we would like to discuss the role of our modeling assumptions in the production of equilibrium predictions coherent with such features.

We show that, within our framework, each element of the model (the policy dimension, the valence dimension, asymmetric information, and party identification) is necessary for this purpose. Indeed, whenever we drop any of these elements the model becomes inconsistent with the data.

When voters do not care about policies, their heterogeneity is due only to differences in information. In this case the model generates too little variation in dynamic voting behavior because everyone with the same information votes the same way. For any state of the world there is at least one group of voters - those belonging to the party whose candidate will become incumbent in period 2 - who can generate at most two voting profiles (DD and RD or DR and RR).[21] The informed voters of such a group can only generate one voting profile, either DD or RR, while the uninformed voters can only generate two voting profiles, either DD and RD (when the informed

[20] Notice that although our analysis focuses on two particular consecutive elections such features emerge in all consecutive presidential elections.

[21] A secondary consequence of this case is that there is complete double switching after the period 1's valence LL. In this case all informed voters from group d vote R, and the informed voters from group r vote D.

generate DD) or DR and RR (when the informed generate RR).[22]

When voters do not care about valence ($\lambda = 0$), we go back to the standard one-dimensional spatial model of voting, where voting behavior is driven only by policy concerns. In particular, when $-(y_j - y_D^t)^2 > -(y_j - y_R^t)^2$ the voter votes for D; when $-(y_j - y_D^t)^2 < -(y_j - y_R^t)^2$ he votes for R; he randomizes with probability $(\frac{1}{2}, \frac{1}{2})$ otherwise. The preferences over the policy space are single-peaked, and the winner in each election is the candidate preferred by the voter with the median policy position.[23] With no utility for valence, party identification loses its informational role because there is no relevant information to be conveyed. In terms of voting patterns, the model generates different voting profiles within each party identification only because policy preferences are heterogeneous within each group. However, any kind of switching behavior is unidirectional. When candidate D is the incumbent and candidate R in period 2 has a policy position to the left (right) of candidate R's in period 1, the model will only generate DD, RR, DR (DD, RR, RD) because period 2's cut-off point is smaller (greater) than period 1's.

A distinctive feature of our model is that different switching patterns within k can only be generated by voters with different information. If we eliminate asymmetric information within a group - that is, if all the voters in group k are either informed or uninformed - the model can only generate three out of four voting profiles (DD, RR, RD or DD, RR, DR) within each group.

When we drop party identification - that is, when we assume that there is a common probability of becoming informed across groups ($m_k = m$, $\forall k$), that all the informed receive the same signal ($s_k^t = s^t$, $\forall k, t$), and that all voters are drawn from the same distribution of policy positions ($Y_k = Y$, $\forall k$) - we cannot explain the differences in voting patterns across groups.[24]

[22] See Degan (2003) for the details.

[23] The median is taken with respect to the aggregate distribution of policy positions.

[24] In the context of our model we could explain the difference in voting patterns across groups only if voters sort into party identification groups based on their individual characteristics and if these have have different effects on voters' policy posisitions. However, although in the data there is some sorting among party identifications based on individual characteristics (e.g., race), it is not extreme

5 Empirical Analysis

5.1 Data

We focus our attention on the 1968 and 1972 U.S. presidential elections.[25] In 1968, consistent with our model, there are two new candidates running for office: Hubert Humphrey and Richard Nixon. Humphrey, a Senator from Minnesota since 1949 and Vice-President since 1964, is the candidate for the democratic party. Nixon, a Senator from California from 1951 to 1953, Vice-President from 1953 to 1960, and an unsuccessful presidential candidate in 1960, is the candidate for the republican party.[26] In 1968 Nixon wins the election with 43.42% of the popular vote to Humphrey's 42.72%.[27] In 1972 Nixon runs as an incumbent against the democratic candidate George McGovern, a Senator from South Dakota since 1962 and chairman of the Reform Commission of the Democratic Party. In 1972 Nixon wins the election by a great margin, gaining 60.69% of the popular vote to McGovern's 37.53%, and then resigns in 1974 after Watergate.[28]

For the purpose of our empirical analysis we need variables for voters' party identifications, their voting choices, their demographic characteristics, and candidates' positions. We take these variables from two sources of data. The first is the 1972 Center for Political Studies (later National Election Studies) data. The second is the DW-NOMINATE legislators' coordinates data.[29]

The 1972 CPS dataset is particularly appropriate for the estimation of our model

enough to justify differences in policy positions across party identifications.

[25]There are two main reasons why we concentrate on two periods. First, we have data on individuals' repeated voting, of a reasonable sample size, for only two consecutive elections. Second, the assumption that both party identification and voters' preferences are constant over time can be justified only over a short period of time.

[26]Although Nixon run for President in 1960, he disappeared from the political scene from 1962 until his candidancy in 1968.

[27]In 1968 George Wallace, a third candidate, receives 13.53% of the popular vote, and other minor candidates receive the remaining 0.33%. We focus only on the two major parties' candidates. We will discuss the robustness of our results to this restriction in section 5.5.

[28]In 1972 there is not a major third candidate running for the presidency, and minor candidates receive the remaining 1.78% of the popular vote.

[29]See Poole and Rosenthal (1997, 1999) and McCarty, Poole, and Rosenthal 1997.

for several reasons. It is an individual-level dataset containing observations on voting choices in 1968 and 1972.[30] It has two different half samples, each representative of the cross-section voting age population in 1972, which can be used to test the performance of the model out of sample. It contains data on individuals' socio-demographic characteristics and party identification.[31]

In all CPS/NES studies respondents are asked the following question: "Generally speaking, do you usually think of yourself as a Republican, Independent, Democrat, or what?" We use such a 3-point categorization as a measure of party identification.[32] We use a dummy variable in each election, R_{68} and R_{72}, which takes the value one if the respondent voted for the republican candidate and zero if the respondent voted for the democratic candidate. Analogously, we use dummy variables, RR, RD, DR, DD, for the respondents' two-period voting profiles. We use a dummy variable, BLACK, which takes the value one if the respondent is black and a dummy variable, FEMALE, which takes the value one if the respondent is a female. To capture the effect of different regions we use a dummy variable, SOUTH, for the solid south.[33] We use two dummies for education level: EDUH for education levels strictly lower than a high school degree and EDUC for education levels greater than or equal to a college degree.[34] We use AGE as a continuous scaled variable for the respondent's age and a

[30]Wright (1993) analyzes the problem of measurement errors in vote choices in NES survey data. Our data on voting choices in 1968 come from retrospective voting questions. We are aware of measurement errors related to recall questions. To have an idea of the magnitude of misreporting in our context, we looked at the 1972-76 panel data and compared the data about the 1972 presidential vote from the 1972 post-election wave to the data from the 1976 pre-election wave. In our selected sample only 7% of respondents reported unconsistent voting choices in the two waves. Himmelweit, Biberman, and Stockdale (1978) analyze the vote bias in recalls.

[31]Each survey contains a pre-election and a post-election wave. Both questions on party identification and retrospective voting are asked in the pre-election wave. This reduces the problem of ex-post rationalization.

[32]There is also a party identification variable which uses a 7-point categorization and takes into consideration the strength of party identification. We choose the 3-point categorization first of all because our model doesn't incorporate the strength of party affiliation, and second because our measure is more stable over time. Our qualitatively results do not change if we treat Independent liners as partisans.

[33]The states included in the solid south are: Alabama, Virginia, Arkansas, Florida, Georgia, Louisiana, Mississipi, North Carolina, South Carolina, and Texas.

[34]The excluded category includes high school degree and some college.

dummy variable, MINCOME, for income levels greater than the median.[35]

From the original data set, containing 2705 observations, we select respondents who voted either for a republican or democratic candidate in both 1968 and 1972 and for whom we have data on demographic characteristics and party identification.[36] The resulting sample contains 1083 observations.[37] Table 1 contains descriptive statistics of the selected sample from the first source of data.

Table 1. Descriptive statistics

Variables	All	r	d	i
R68	655	344	111	200
R72	716	339	117	200
RR	584	330	91	163
RD	71	14	20	37
DR	132	9	86	37
DD	296	7	231	58
BLACK	97	5	76	16
EDUH	346	104	181	61
EDUC	213	90	54	69
SOUTH	195	43	113	39
FEMALE	585	201	245	139
AGE*	240	104	92	44
MINCOME	615	222	214	179
TOT	1083	360	428	295

*We include in the cell Age the number of respondent 62 or older

The DW-NOMINATE is a dynamic model that estimates, separately for the House of Representatives and the Senate, legislators' coordinates on a two-dimensional policy space using legislators' roll call voting choices throughout their career.[38] Similarly, the coordinates of Presidents are estimated using their support roll calls.[39] The first

[35]In different versions of the model we have excluded age and/or income. The main results of our analysis were uneffected.

[36]To evaluate the robustness of our estimates with keeping only individuals who voted in both elections, we performed out-of-sample predictions on those that voted in only one election (see section 5.5).

[37]See Appendix B for further details on the dataset.

[38]One empirical reason we do not include Wallace in our analysis is that he has not been a member of Congress, hence we do not have a measure of his policy position.

[39]Notice that even if the President does not vote, "Presumably, if the President were able to vote, he would vote in the direction indicated in the support calls." Poole and Rosenthal (1999, p.9).

dimension of the DW-NOMINATE coordinates has been interpreted as the traditional liberal-conservative dimension. Therefore, we use the first dimension of the NOMINATE estimates as a measure of candidates' positions on a liberal-conservative scale. In particular, we use the "constant model" version of DW-NOMINATE, where candidates' coordinates are constrained to remain constant over the candidate's whole career.[40]

There are two main reasons why such measures of candidates' coordinates are particularly appealing. First, they are restricted to lie within the interval $[-1, 1]$. Second, the fact that legislators are constrained to have a constant position allows us to compare coordinates of the elected Presidents with those of their challengers who typically served in Congress in different years. Table 2 reports the coordinates, on the liberal-conservative dimension, of the presidential candidates considered in our study.[41]

Table 2. Presidential candidates' coordinates

Candidates	Humphrey H	McGovern G	Nixon R
coordinate:	-.34	-.467	.451

5.2 Estimation Procedure

In the model voters know their own ideal points but the econometrician doesn't. We are interested in the link between demographic characteristics and policy preferences for each group of voters. Here we choose a particular functional form for the distribution of ideal points Y_k. Specifically, we assume that y_j is drawn from a beta distribution with support $[-1, 1]$ and parameters (p_{kj}, h).[42] We parametrize the first parameter of such distribution to j's party identification and demographic characteristics (BLACK, EDUH, EDUC, SOUTH, FEMALE, AGE, MINCOME), and we

[40]The DW-NOMINATE "constant model" coordinates were available on-line at *htpp://voteview.uh.edu.*

[41]To provide a term of comparison, we can, for example, give the coordinates of Presidents Carter and Reagan, which are - .364 and .608 respectively.

[42]We choose the beta distribution because it is the most flexible distribution, and it is defined on a finite support.

restrict the second parameter to be the same both within and across groups. Since the coefficients of a beta distribution must be strictly positive, $p_{kj} > 0$ and $h > 0$, our parametrization becomes

$$p_{kj} = \exp(\beta_k W_j)$$
$$h = \exp(\alpha)$$

where β_k and α are preference parameters to be estimated.

It is useful to remember that candidates' positions are exogenous parameters available in the data and that for any candidates' positions pair and state of the world, the unique equilibrium with sincere voting induces a different voting behavior. Notice that we cannot estimate valence directly because different candidates' valence lead to different equilibrium voting behaviors, and consequently, to different conditions on the parameters we want to estimate. However, since for any state of the world there is a unique equilibrium voting behavior and we observe an ex-post voting behavior, we can estimate the parameters of the model by maximum likelihood conditional on each state. We then pick as an estimate of valence the state whose equilibrium maximum likelihood is the highest, indicating that the observed voting profile was more likely under that state.

For reasons that will be explained in the next paragraph, we let $z = \lambda(H - L)$. For given z, m_d, m_i, m_r, q, candidates' positions $y = \{y_D^1, y_R^1, y_D^2, y_R^2\}$, and distributions of ideal points $B_{kj}(\cdot)$, we can derive the likelihood of the observed voting profile V_j of individual j from group k conditional on state $x = (x_D^1 x_R^1 x_D^2 x_R^2)$ as:[43]

$$L(V_j|\beta_k, \alpha, m_k, z, q, y, W_j, k, x) = \int_{-1}^{1} L(V_j|\beta_k, \alpha, m_k, z, q, y, W_j, k, x, y_u) b_{kj}(y_u) dy_u$$

We fix q to 0.5, which is equivalent to an uninformative prior, and we estimate β_k, α, m_d, m_r, m_i, and z conditional on the state x.[44] The parameter z is a composite

[43] We use b_{kj} to indicate the density function of the policy position of voter j with characteristics W_j belonging to group k as a short form for $b(p_{kj}, h)$. We do analogously with the cdf.

[44] In section 5.4 we discuss the robustness of our results to both the assumption on the second parameter of the beta and on the probability of being high valence.

parameter. As we mentioned in section 3, it is impossible to separately identify the weight λ and the perceived maximum possible difference in candidates' valence $(H-L)$, an element independent of the realized valence. A large value of z may be due to voters placing a high weight on valence or to a great potential difference in candidates' valence.[45]

For any two consecutive elections, where in the first period there are two new challengers and in the second period the incumbent runs for reelection, there are potentially eight states to consider. However, as mentioned in section 3, the model puts restrictions on which states are compatible with the observed voting profiles and candidates' configuration. Using the NOMINATE coordinates from table 2, it is easy to verify that the midpoints and gaps in the two periods are such that $m_1 > m_2$ and $g_1 < g_2$. The configurations in 1968 and 1972, together with the fact that we observe all four voting profiles within each group, allow us to exclude states $HHHH$, $HHLH$, $LLHL$, $LLLL$, and $LHLH$ because the derived equilibrium voting profiles are incompatible with the data.[46] We can therefore restrict our attention to three states: $HLLL$, $LHHH$, and $HLHL$. When we estimate the model by conditional maximum likelihood for each of these states, we find that the state corresponding to the highest likelihood is $HLLL$.[47]

[45] We could fix either λ or $(H-L)$ to some arbitrary value, but since these two concepts are related and we are not interested in the estimation of λ and $(H-L)$ per se, we prefer not to take any stand on their value and estimate instead the composite parameter z.

[46] In states $HHHH$ and $HHLH$, Independents and Democrats cannot generate the profile RD because for any value of $z > 0$, the corresponding cut-off points of the informed and uninformed in period 1 are greater than the ones in period 2 ($y_{1i} > y_{2i}$, $y_{1d} > y_{2d}$ and $y_{10} > y_{20}$). In states $LLHL$ and $LLLL$, Democrats cannot generate the profile DR because the conditions on z needed to have $y_{1a} > y_{2a}$ and $y_{10} < y_{20}$ (the only characterization of cut-off points compatible with the profiles of Independents and Republicans) are incompatible. In state $LHLH$ Democrats could generate RD only for values of q that, even if allowed, would impose conditions on z incompatible with the other groups' voting profiles.

[47] Each case requires different restrictions on z in order to generate cut-off points compatible with the four voting profiles. Based on likelihood criteria, given a likelihood at convergence of -885.49 for the state $HLLL$, we could exclude $LHHH$ (with a likelihood at convergence of -896.85) and $HLHL$ (with a likelihood smaller than -1,000). We also compared the goodness-of-fit tests of the model estimated conditional on each of the three states. While when the state is $HLLL$ the estimated model passes the standard Pearson's Chi-Square test on static and dynamic voting profiles both at the aggregate level and by party identification (see section 5.4), it doesn't pass the test on some or

When the state is $HLLL$, in equilibrium the informed voters generate the profiles DD, DR, RR and the uninformed voters the profiles DD, RD, RR. The conditional likelihood function of voting profile V_j of voter j from group k is:

$$L(V_j|\beta_k, \alpha, m_k, z, y, W_j, k, HLLL) =$$
$$\int_{-1}^{1}[m_k I(y_u < y_{2k}) + (1-m_k)I(y_u < y_{10})]^{DDj} \cdot [m_k I(y_{2k} < y_u < y_{1k})]^{DRj} \cdot$$
$$[(1-m_k)I(y_{10} < y_u < y_{20})]^{RDj} \cdot [m_k I(y_u > y_{1k}) + (1-m_k)I(y_u > y_{20})]^{RRj} \cdot b_{kj}(y_u)dy_u$$

where $y_{1k,}$ $y_{2k,}$ y_{10}, y_{20} are the cut-off points corresponding to state $HLLL$ and DD_j, DR_j, RD_j, and RR_j are dummy variables for individual j's voting profile. Using the cdf of the beta distribution $B(\cdot)$, the likelihood can be rewritten as:

$$L(V_j|\beta_k, \alpha, m_k, z, y, W_j, k, HLLL) =$$
$$[m_k B_{kj}(y_{2k}) + (1-m_k)B_{kj}(y_{10})]^{DDj} \cdot [m_k B_{kj}((y_{1k}) - B_{kj}(y_{2k}))]^{DRj} \cdot$$
$$[(1-m_k)(B_{kj}(y_{20}) - B_{kj}(y_{10}))]^{RDj} \cdot [m_k(1 - B_{kj}(y_{1k})) + (1-m_k)(1 - B_{kj}(y_{20}))]^{RRj}$$

The total loglikelihood of j's voting profile unconditional on his party identification can be easily derived as:

$$l(V_j|\beta_d, \beta_r, \beta_i, \alpha, m_d, m_r, m_i, z, y, W_j, HLLL) =$$
$$\sum_j[\ln(L(V_j|\beta_d, \alpha, m_d, z, y, W_j, d, HLLL)) \cdot pidD_j +$$
$$\ln(L(V_j|\beta_r, \alpha, m_r, z, y, W_j, r, HLLL)) \cdot pidR_j +$$
$$\ln(L(V_j|\beta_i, \alpha, m_i, z, y, W_j, i, HLLL)) \cdot pidI_j]$$

where $pidR_j, pidI_j, pidD_j$ are dummies for j's party identification.

5.3 Estimation Results

The estimated values of β_d, β_r, and β_i (table 3) characterize the distribution of voters' policy positions.[48] To interpret the coefficients on demographic characteristics note

all the voting profiles when state $LHHH$ and $HLHL$ are considered respectively.

[48]The covariance matrix of the parameters is calculated as the inverse of the numerical Hessian. For a complete characterization of the estimated distribution of voters' policy positions, we also need the estimate of α and the derived value of h (table 4).

that a bigger value of any coefficient corresponds to a bigger p_{kj} whose effect is to move the mass of the beta distribution to the right. It follows that the higher the coefficient on any individual characteristic the more conservative voters with such a characteristic tend to be.[49]

Table 3. Estimated parameters: policy positions

Variable	Estimate	St.err.	t-stat.
Democrats			
CONST	1.423	0.724	1.966
BLACK	-1.098	0.417	-2.634
EDUH	0.020	0.095	0.212
EDUC	-0.428	0.195	-2.199
SOUTH	0.448	0.155	2.898
FEMALE	-0.022	0.081	-0.275
MINCOME	0.223	0.112	1.987
AGE	0.005	0.033	0.151
Republicans			
CONST	2.542	0.436	5.833
BLACK	-0.701	0.339	-2.067
EDUH	-0.238	0.123	-1.938
EDUC	-0.006	0.095	-0.066
SOUTH	0.002	0.066	0.029
FEMALE	-0.153	0.100	-1.531
MINCOME	0.096	0.095	0.972
AGE	0.061	0.034	1.792
Independents			
CONST	2.338	0.480	4.866
BLACK	-1.201	0.446	-2.695
EDUH	0.085	0.119	0.715
EDUC	-0.334	0.139	-2.396
SOUTH	0.281	0.161	1.743
FEMALE	-0.236	0.107	-2.199
MINCOME	-0.024	0.091	-0.265
AGE	0.002	0.029	0.066

The results on policy preferences are the following. There are some characteristics that affect voters' policy preferences similarly regardless of their party identification and other characteristics whose effects differ with party identification. BLACK has

[49]This is true when we make the comparison for fixed h.

a notable (negative) effect on all groups. SOUTH has a very strong (positive) effect on Democrats and an almost significant effect on Independents, while it is has no effect on Republicans. EDUC has a significant (negative) effect on Democrats and Independents, while EDUH has a significant (negative) effect on Republicans. This means that among Democrats and Independents the most educated voters are more liberal than their less educated counterparts, while among Republicans the opposite is true, that is, the least educated are more liberal. FEMALE is (negatively) significant only among Independents and slightly significant among Republicans. AGE does not help to explain policy preferences. MINCOME has a relatively significant effect only on Democrats.

Interestingly, despite the high correlation between income level and education, income level has a separate and opposite effect from education. While Democrats with a high education are more liberal, those with the highest income level are more conservative. Note also that none of the individual characteristics are very significant among Republicans, probably due to the small number of observations with any profile other than *RR*.

A complementary way to analyze policy preferences is to look at the plots, both at the aggregate level and by party identification, of the estimated marginal distributions of voters' ideal points by characteristics (figures 1.1-1.26 in Appendix C). All the marginal distributions are in line with the above results. Blacks are more liberal than non-blacks both in the aggregate (figure 1.1) and within each party identification (figures 1.7, 1.13, and 1.19). Southerners are more conservative than non-southerners (figure 1.3). However, this effect is very significant for Democrats, less so for Independents, and not significant at all for Republicans (figures 1.9, 1.21, and 1.15). This result is consistent with the division that was occurring during those years between southern and northern Democrats and with the known fact that southerners were generally more conservative than northerners. Although at the aggregate level highly educated voters are more conservative than their less educated counterparts (figure

1.2), education has different effects across party identifications (figures 1.8, 1.14, and 1.20). Even though women are more liberal than men among both Republicans and Independents (figures 1.16 and 1.22), the effect of gender disappears in the aggregate (figure 1.4). As expected from the point estimates, Democrats with an income level lower than the median are more liberal than those with higher income levels (figure 1.11). Age doesn't have a separate effect on policy preferences. However, Independents older than 62 appear to be more conservative than young Independents.

The above results on the estimated policy preferences lead us to the following three observations. First, considering the relatively homogeneous demographic composition of different groups, our results indicate that to understand the relationship between demographics and policy preferences it is important to consider their interaction with party identification. Second, overall, Democrats' policy preferences are more heterogeneous than Independents', which in turn are more heterogeneous than Republicans' (figure 1.25). Third, the aggregate distribution of ideal points is relatively conservative (figure 1.26). Both last results are in accordance with the self-reported liberal-conservative view of the population.[50]

It is important to point out that in our estimation we do not use any a priori information on individuals' political preferences (such as self-reported preferences towards candidates, policies, or parties). Rather, in order to estimate voters' preferences we apply a revealed-preference approach which relies solely on observed individual voting choices, and we use individual self-reported preferences only to externally validate our results. Yet, maybe surprisingly, most of the above results are in line with what other political-sociological studies have said about political preferences of American voters.[51] In addition, our results allow us to disentangle the effect that each charac-

[50] According to the self-reported positions on a 7-point liberal-conservative categorization from the 1972 CPS data, 24% of respondents report to be liberal, 41% conservative, and the remaining 35% report to be middle-of-the-road; the corresponding proportions among Republicans are 9%, 60%, and 31%; among Democrats 33%, 26%, and 41%; and among Independents 29%, 37%, and 34%. In the summer 1969 Gallup opinion polls, 23% of respondents report to be conservative, 28% moderately conservative, 18% moderately liberal, 15% liberal, and 15% don't know.

[51] See, e.g., Miller and Shanks (1996) and Scammon and Wattenmberg (1971).

teristic has on voters' policy preferences and get further insight to the relationship between demographics and political views of the American electorate.

The estimated voters' probabilities of being informed (m_d, m_r, and m_i in table 4) indicate that Democrats and Republicans were more likely to be informed than Independents (68%, 29%, and 18% probability respectively) and that Democrats were more likely to be informed than both Republicans and Independents.[52] As we would expect if we assume that there is some cost of gathering information, we estimate that even if Independents have richer information (they receive signals on both candidates), they have a smaller probability of receiving such information compared to the other groups.

Table 4. Estimated parameters (continued)

Variable	Estimate	st. err.
m_d	0.6799	0.069
m_r	0.2870	0.0897
m_i	0.1869	0.0295
z	0.8098	0.2274
h	4.6183	3.0256
LogLikelihood		-885.49

Our estimated state of the world is $HLLL$, meaning that we estimate Humphrey to have high valence and McGovern and Nixon to have low valence. Most experts would agree $HLLL$ is an accurate reflection of the actual state. Humphrey did not win the 1968 presidential election, but he was highly respected and very experienced having served in the Senate almost continuously from 1949 until he died in 1978. Scammon and Wattenberg (1971, p.172) seem to be in perfect agreement with our findings stating that Humphrey ". . . was perceived, finally, as hard-hitting, intelligent, and forceful, as well as a nice guy."

Nixon's success was due mainly to the coupling of his relatively moderate economic policy with his ability to deal with foreign affairs; however, as Watergate revealed, he

[52]The parameters m_d, m_r, m_i, z, h are transformations of directly estimated parameters. Their standard errors are calculated using the delta method.

was not a trustworthy politician. Even before Watergate the wide use of impoundment of funds, veto power, and administrative discretion, as well as the secret bombing of Cambodia in 1969 may point to Nixon being a low-valence politician.[53] More objective evidence of Nixon's low valence include his "dirty-tricks team," an old institution in American politics that was taken to new heights in 1972; and the "Plumbers," an apparatus originally organized to plug leaks that eventually expanded its activities to include a variety of secret and illegal operations.[54]

George McGovern, a very liberal candidate not supported by mainstream Democrats, easily won the democratic nomination basically because he was the only candidate left after Nixon's negative campaign forced Muskie then Jackson and finally Humphrey out of the race. The Eagleton case and the O'Brien case are only a few of the many examples of McGovern's inability to make clear decisions and to keep promises.[55]

Our findings on candidates' valence imply that Nixon was elected President in 1968 even though he was a low-valence politician running against a high-valence challenger (Humphrey). Even more interestingly is that the 1972 election was a landslide for Nixon even though all voters knew of his low valence. Our estimated model provides an equilibrium interpretation of such outcomes as well as of the observed group-specific voting patterns.

According to our estimated model, the main factors allowing Nixon to win both elections were a conservative constituency together with voters' lack of information in 1968 about his low valence, and having an extremely liberal and low-valence demo-

[53]The secret bombing of Cambodia doesn't neccessarily indicate Nixon's low valence, but it is one example of Nixon's tendency to take action in secrecy. Also, the frequent exercise of veto power and impoundements of funds may indicate Nixon's low valence to the extent to which they demonstrate his inability to constructively negotiate with the Congress, a skill required of a good leader.

[54]Among the most famous episodes related to the "dirty-tricks team" and the "Plumbers"are: the secret efforts to discredit and defame Daniel Ellsberg (who published a top secret study on the origin and conduct of the war in Vietnam in the New York Times); the fake cable created in an attempt to link President J. Kennedy to the assassination of south Vietnam's President; the illegal money-collection during the 1972 campaign; and the predatory strategy of Nixon's 1972 committee for reelection, which attempted to sabotage the campaigns of the top democratic front-runners in the hopes of forcing them out of the race and leaving a weak democratic opponent for the general election.

[55]See, White (1973).

cratic opponent in 1972.

Among Democrats, who we found to be relatively conservative, the informed who voted for Humphrey in 1968 and switched to Nixon in 1972 (20%) were voters with a moderate policy view who knew Humphrey had high valence and McGovern had both low valence and an extremely liberal policy position. When faced with candidates with equal valence (Nixon and McGovern), they preferred to vote for the more conservative candidate (Nixon). The vote of informed Republicans and Independents with voting profile DR, who both knew Nixon had low valence, was driven mainly by policy concerns in both periods. The uninformed voters who switched their vote from R to D were slightly conservative voters that given their symmetric prior on valence, voted for Nixon in 1968 driven by policy concerns.[56] However, Nixon's revealed low valence turned their votes toward McGovern in 1972 despite his liberal policy position.

We can also explain why, although Nixon won in 1972 with a much greater margin than in 1968, the proportion of Republicans who voted for Nixon in 1972 is smaller than in 1968. In the first election most Republicans were unaware of Nixon's low valence, and they voted for him based on policy considerations. In 1972 some of these voters, after learning that Nixon had low valence, preferred to vote for McGovern despite his extremely liberal policy position.

5.4 Goodness of Fit

To assess whether our model can reproduce the quantitative features of the data, we need to compare the predicted voting profiles to the observed ones. We report the actual and fitted voting profiles at the aggregate level and by party identification in tables 5, 6, and 7. The "actual" column reports the frequency in the data (overall or by party identification) of each voting profile. The "predicted" column reports the estimated probability of each voting profile, calculated by integrating the individual probability of such a profile over voters (overall or within a particular group).

[56] According to the candidates' coordinates from table 2, middle-of-the road voters are closer to Humprhey than to Nixon.

Table 5. Aggregate dynamic voting profiles

Profiles	Actual	Fitted
DD	27.33	27.46
DR	12.19	12.10
RD	6.56	6.46
RR	53.92	53.98
$\mathcal{X}^2_{(3)}$		.0303

Table 6. Aggregate static voting profiles

1968	Actual	Predicted	1972	Actual	Predicted
D	39.52	39.56	D	33.89	33.93
R	60.48	60.44	R	66.11	66.07
$\mathcal{X}^2_{(1)}$		.0007	$\mathcal{X}^2_{(1)}$		.0007

Table 7. Dynamic voting profiles by party identification

Profiles	Actual	Predicted	Actual	Predicted	Actual	Predicted
	Party D	Party D	Party R	Party R	Party I	Party I
DD	53.97	53.73	1.94	2.84	19.66	19.41
DR	20.09	20.43	2.50	1.78	12.54	12.59
RD	4.67	4.88	3.89	3.05	12.54	12.92
RR	21.26	20.97	91.67	92.32	55.25	55.08
$X^2_{(3)}$		.0828		2.900		.044

We perform standard goodness-of-fit tests on both dynamic and static voting profiles and report the relative Pearson's Chi-Square test at the end of each table.[57] Table 5 shows the results for the four dynamic voting profiles at the aggregate level. The value of the test indicates that our model cannot be rejected by the data at conventional significance levels.

Table 6 shows the aggregate static voting patterns in 1968 and 1972. The model predicts perfectly both electoral outcomes. In neither year can we reject our model. Analogous results hold for the dynamic voting patterns within each party identification (table 7). The model captures both party loyalty and switching behavior, and their differences across parties.

[57]The test on dynamic voting profiles is a $\mathcal{X}^2$ with three degress of freedom, while the test on the static voting profiles is a $\mathcal{X}^2$ with one degree of freedom. The critical values at 5% are 7.81 and 3.84 respectively.

An additional and complementary way to assess the ability of a model to generate predictions quantitatively coherent with empirical evidence is to verify how it performs out of sample. We make two different types of out-of-sample predictions.

First, we estimate the model on either one of the half samples of the original data set and make out-of-sample predictions on the other half. Second, using the estimated parameters from the original sample, we perform out-of-sample predictions on voting behavior of voters in the 1968 and 1972 elections using data from a different source, the General Social Survey (GSS).

The reader can find the results of these out-of-sample predictions in Appendix D. The model performs relatively well on the half samples (tables D.1-D.6), predicting the aggregate outcomes of the elections in 1968 and 1972 and the dynamic voting patterns of Democrats and Republicans on both half samples. The model doesn't pass the goodness-of-fit test on the dynamic voting profiles of Independents only by a tiny margin. The imperfect prediction of the behavior of Independents prevents the model from passing the test on the aggregate dynamic voting profiles.[58]

Our model performs very well on the GSS data (tables D.7-D.9). The estimated model perfectly predicts the electoral outcomes in both 1968 and 1972, and it cannot be rejected on either the aggregate or the group-specific dynamic voting profiles .[59]

5.5 Robustness

While in section 4 we discussed the elements of the model necessary to explain the qualitative features of the data, here we discuss their importance to our empirical analysis and the robustness of the results to some of the assumptions made in the estimation. Some of these assumptions are necessary for identification purposes, others allows us to obtain a better fit of the data or to improve the precision of the estimates.

[58]The imperfect prediction is probably affected by the small sample size of the two half samples (577 and 506 observations respectively).

[59]It fails to pass the test for Independents by just a few decimal points.

In our model, information status (that is, whether a voter receives a signal) is assigned at the beginning of the first period and remains constant thereafter. This is a critical assumption allowing us to identify the probability of being informed. In fact, we can identify m_d, m_r, and m_i because by holding the information status of each voter fixed, informed and uninformed voters generate opposite switching patterns.

We assumed that each policy position with positive mass has a corresponding fraction of voters m_k that becomes informed.[60] This assumption on information status is sufficient to guarantee that the aggregate vote share perfectly reveals information on candidates' valence. Moreover, because of this assumption, the aggregate voting patterns put restrictions directly on the probability of being informed.

None of the assumptions on party identification are necessary for identification, rather, they help to better explain the features of the data.[61] It would be difficult to explain differences in dynamic voting patterns across parties without these assumptions. We estimated a model in which we completely eliminate the role of parties. The model in which voters' policy positions are drawn from a common distribution ($\beta_k = \beta, \forall k$) and each voter has the same probability of becoming informed ($m_k = m$, $\forall k$) and the same signal ($s_k^t = \{x_D^t, x_R^t\}$, $\forall k$) is rejected by both the likelihood-ratio test and the goodness-of-fit test. The data reject a model in which party identification is not taken into account. We reach similar results even if we shut down each of the first two elements of party identification (different β_k and m_k) one at a time.[62] On the basis of the goodness-of-fit test on dynamic voting profiles and on the likelihood-ratio

[60] The probability of being informed may be otherwise related to policy preferences or to observable characteristics or may be just a random effect.

[61] The assumptions we refer to are: the relationship between policy positions and party identification, the party specific probability of becoming informed, and the asymmetry in signals accross parties.

[62] The assumption that individuals with a partisan party identification receive only the signal about their own party's candidate implements the idea that citizens with a partisan party identification are more likely to receive information about their own party's candidate. We cannot reject our model specification when we compare it with the model where all informed voters receive both signals regardless of their party identification. If we fix z, we can also estimate a model where informed partisans can receive a signal either on their own candidate or on both. We cannot reject the hypothesis that partisans receive only one signal.

test, we reject both the model in which preferences are constrained to be the same across parties and the one in which there is a common probability of being informed.[63]

When we estimated the model with the probability that a candidate has high valence (q) free, we obtained a point estimate of .53 (very close to our uninformative prior). However, although the qualitative results did not change, the estimates of all the other coefficients became much less precise.[64]

In the specification of the distribution of voters' ideal points, we have parametrized the first parameter of the beta while we constrained the second parameter to a constant common to all party identification groups. The estimate of h is rather imprecise. Although we restricted the second coefficient of the beta not to be party specific, the identification of h separately from the constant terms in p_{kj} is tenuous. In our context, such identification is complicated by the fact that we also estimate the parameter z, which enters in the expression for the cut-off points. Some sort of restrictions are necessary. We could have opted for different parametrizations, but ours is probably the most flexible, since we want to allow both the mean and the variance to differ across party identifications. In addition, the restriction on h make the coefficients on individual characteristics for different party identification groups more comparable.

By abstracting from abstention, what we estimate is the distribution of policy positions of citizens who go to vote. In fact, we have restricted the analysis to citizens who voted in both 1968 and 1972 and who voted either for the republican or the democratic candidate. However, we have conducted out-of-sample predictions on voting choices of citizens that we observe voting only in 1968 or in 1972 (tables D.10 and D.11 in Appendix D) and the results suggest that our estimates are robust to the exclusion from the sample of people who voted in only one election.[65]

[63]The loglikelihood of the model in which there is a unique distribution of policy preferences is -1100.95, while in the model in which there is a common probability of becoming informed it is -911.10.

[64]On the basis of likelihood-ratio test we couldn't reject our restriction that q equals 0.5.

[65]Notice that the sample of 1968-only voters includes observations for which we have missing values in 1972. The same is true for the sample of 1972-only voters, where the missing values in 1968 also include those that voted for Wallace.

Similarly, to assess the robustness of our estimates to the abstraction from Wallace's candidacy in 1968 (in 1972 there was no relevant third party candidate), we use our estimated model to predict how the voters who chose Wallace in 1968 voted in 1972 (table D.12). It seems that our estimated model is able to predict how these individuals voted in 1972.[66] We are aware that this result is not sufficient to claim that abstracting from a third party's candidate does not bias our results. However, we find the results of this out-of-sample prediction (table D.12) comforting because they indicate that the non-inclusion of Wallace does not affect the ability of the estimated model to predict individual voting behavior.[67]

6 Counterfactual Experiments

As pointed out in section 5.3, the estimated model tells us that Nixon wins the election in 1968 even though he has low valence and faces a high-valence opponent. It also tells us that Nixon is reelected in 1972 despite the fact that voters know he is a low-valence politician. These outcomes arise from the combination of a conservative constituency, the particular degree of asymmetric information about candidates' valence, and the trade-off between policy and valence in voters' utility function. Since voters have preferences over both candidates' policy positions and their valence, based on these findings alone we cannot draw any immediate conclusion on the relative desirability of alternative electoral outcomes.

In this section we use our estimated model to conduct counterfactual experiments allowing us to shed some light on the efficacy of the electoral process in selecting the "best" candidate and on the so called incumbency advantage.

We determine the most liberal policy position that would have allowed a high and a low-valence democratic candidate to defeat Nixon in 1968 and in 1972. Then we

[66] We slightly underpredict the proportion of Democrats that voted for the republican candidate.

[67] It is not clear how we should modify the information structure to include a third candidate in our model. Should Independents receive information only on the third party or on all candidates? In addition, from a pratical point of view we cannot include Wallace in our empirical analysis because we do not have a measure of his policy position.

match this with the policy positions of the candidates participating in the democratic primaries. We find that in 1968 none of the existing candidates held such a policy position.[68] We obtain a different result in 1972. While McGovern would have lost the election even if he had a high valence, Humphrey and Muskie could both have defeated Nixon provided that they had high valence.[69] Additionally, there were two other candidates, Jackson and Lindsay, that according to our model could have defeated Nixon independently of their valence.[70]

This result is in line with the argument that after 1968 the efficacy of the primary elections system in selecting the "best" candidate declined dramatically. The McGovern-Fraser Commission introduced a reform to the democratic presidential nominating process making the system more democratic since 1972. The commission "recommended" adopting proportional representation in the allocation of delegates for the various contenders based on the proportion of the popular vote they received in the primary; it required, in convention and caucus states, that 75% of delegations be chosen at district level; and it fixed quotas to guarantee the representation of minority groups at the national convention.[71] From 1968 to 1972 the percentage of delegates nominated through primaries went from 40% to more than 60%. From a theoretical point of view, it is well known that plurality rule in an environment with more than two candidates does not necessarily lead to the selection of the "best"

[68]Only a high-valence democratic candidate with a policy position greater than -.1349 could have defeated Nixon in 1968. The main democratic candidates in 1968 were Johnson (who resigned), Robert Kennedy (who was assassinated during the primaries), and McCarthy. Their policy positions are respectively -.235, -.468, and -.369.

[69]In 1972 a high-valence candidate would have won the election had he had a policy position greater than -.416. A low-valence candidate would have needed a policy position greater than -.2385. We know from our estimates that Humphrey had high valence.

[70]The main democratic candidates in the 1972 primary elections were Humphrey, Muskie, Jackson, and Lindsay, candidates whose policy positions (apart from Lindsay's) are respectively -.34, –.328, and -.205. We do not have the policy position of Lindsay on a comparable scale since he served the House of Representatives and not the Senate. However, looking at his coordinate on the common-space model (Poole, 1998), where the coordinates of the members of the House of Representatives and of the Senate are scaled on a common-space, we can claim that Lindsay had a policy position conservative enough to allow him to defeat Nixon in 1972 (he switched party affiliation from Republican to Democrat in 1972).

[71]See, e.g., White (1972) and Davis (1997).

candidate.[72] Also, simulation studies show that the process of selecting presidential nominees in the U.S. is highly unpredictable.[73] The new system favored the nomination of extremely liberal candidates who didn't necessarily have a high valence.[74] Our result is also in line with the fact that Nixon's campaign for reelection in 1972 was effective in inducing his strongest democratic opponents to drop out of the race.

In the second counterfactual experiment, we compare the realized electoral outcomes in 1968 and 1972 with those corresponding to a scenario where all voters have perfect information on candidates' valence. The outcome of this experiment allows us to assess the role played by asymmetric information in the electoral process. We find that while in 1972 McGovern would have still lost the election, in 1968 Humphrey would have won.

Similarly, within the asymmetric information environment, we can assess the effect different probabilities of being informed have on the electoral outcome. This is relevant since parties can affect such probabilities during a campaign. We do not need to find stories about Nixon and other Republicans affecting the information during the campaign.[75] It is well known that since the beginning of his political career, Nixon adopted the tactic of "discredit your opponent."[76] Two examples are the series of "dirty tricks" played on Ed Muskie, the democratic front-runner in 1972 and Nixon's denigrating campaigning against McGovern in 1972.[77]

While we cannot find any combination of probabilities that would have helped

[72] See, e.g., Mueller (1989).

[73] See, e.g., Merril (1988) and Cooper and Munger (2000).

[74] See, e.g., Davis (1997).

[75] See White (1969, 1973) and Genovese (1990).

[76] New York Times, April 24, 1994.

[77] "In February of 1972, voters in New Hampshire, site of the first primary, received late night phone calls from people claiming to represent the 'Harlem for Muskie Committee' promoting the candidacy of Muskie. . . .Shortly after the Florida primary, letters were mailed to Democrats on stationary stolen from Muskies' headquarters, with 'vote for Muskie' message and containing vicious lies about Muskies' Democratic opponents. . . . Perhaps the most damaging trick on Muskie took place just prior to the New Hampshire primary. The conservative newspaper *The Manchester Union Leader* published a letter signed by a Paul Morrison accusing Muskie of insulting Canadian-Americans, calling them 'Canucks', and accusing Muskie's wife of being an alchoolic who would walk up and down the aisles of planes drunk, encouraging people to tell dirty jokes" Genovese (1990, p.183).

McGovern win the election in 1972, we do find different information structures that would have made Humphrey win the election in 1968. For example, even if not fully informed, had Republicans and Independents been more informed, Humphrey would have been elected President in 1968. According to our estimates of the probabilities of being informed, Republicans and Independents were more susceptible to the lack of information. They voted for Nixon both because of his policy position and because they did not think he was a low-valence politician.

Finally, we use our estimated model to make an observation regarding the so called incumbency advantage. A large literature on incumbency advantage has been developed based on the empirical observation that incumbent politicians have a higher probability of being elected than new challengers. One possible explanation for such an advantage is that, in the presence of moral hazard and adverse selection, an efficient electoral process guarantees that good-quality (high-valence) politicians are elected with a higher probability than bad quality (low-valence) ones.[78] Another possible explanation is that risk averse voters prefer to elect a known incumbent rather than a new, and hence "more risky," challenger.[79] Our estimated model indicates that although low-quality incumbents can be reelected, and can even win with a wider margin than in the election in which they were first appointed, it does not necessarily follow that the incumbent has an advantage. In our model had Nixon not been incumbent in 1972 but a new challenger with unknown valence, he would have won the presidential election with a wider margin than he actually did.[80] In fact in our model a low-valence incumbent has a disadvantage, but despite this (as in 1972) he can be reelected by a large margin. A high-valence incumbent has an advantage, but likewise he may be defeated.[81] Our finding that incumbents can have an advantage as well as a

[78]See Banks and Sundaram (1998). Ferejhon (1986) has a model with only moral hazard where politicians who exert higher effort have a higher probability of being reelected.

[79]See Bernhardt and Ingberman (1985).

[80]In our sample the voting share for Nixon in 1972 had he been a new challenger would have been 71.61% instead of the 66.11% that he received (in our sample) as an incumbent.

[81]When we apply our model to the 1976 and 1980 elections, we estimate Ford to have low valence and both Carter and Reagan to have high valence. This is an example in which a high-valence

disadvantage is in line with the results of the model proposed by Fiorina (1981) where voters care about both the expected future policy of an elected politician and his past record. Similar to our model where the incumbent has a disadvantage (advantage) when he has low (high) valence, in his model the incumbent has a disadvantage (advantage) when he has a bad (good) policy record.

7 Conclusions

In this paper we propose and estimate a dynamic model of voting with asymmetric information incorporating the three main factors affecting individual voting behavior: policy preferences, candidates' valence, and party identification. The estimation is based on a revealed preference approach. We use the structure of the model plus individual-level data on voting choices in two consecutive elections to uncover fundamental utility parameters and other unobservable elements. In particular, we estimate the distribution of voters' policy positions, candidates' valence, and the probability of voters of being informed.

The estimated distribution of voters' policy positions allows us to quantify the effect of demographic characteristics on voters' policy preferences conditional on their party identification. Our results indicate the importance of party identification to the relationship between demographics and policy preferences. We find that overall Democrats' policy preferences are more heterogeneous than Independents', which in turn are more heterogeneous than Republicans', and that the aggregate distribution of ideal points is relatively conservative.

Our estimates of valence indicate that Humphrey (the democratic candidate in 1968) had high valence, McGovern (the democratic candidate in 1972) had low valence, and Nixon (the republican candidate in 1968 and again, as incumbent President, in 1972) also had low valence. This result is perhaps surprising given that Nixon won the 1968 election by a very small margin but went on to win the 1972 election by a

incumbent, Carter, is defeated.

margin as large as that gained by Johnson in 1964 or Roosevelt in 1936. The results are, however, consistent with anecdotal accounts of the events surrounding the 1968 and 1972 elections.

We estimate that Democrats have a higher probability of being informed than Republicans who, in turn, have a higher probability of being informed than Independents (68%, 29%, and 18% probability respectively).

We use the estimated candidates' valence, voters' probability of being informed, and distribution of voters' policy positions to provide an equilibrium interpretation of the observed voting patterns and electoral outcomes. In addition, we use the estimated model to conduct counterfactual experiments allowing us to assess the relative importance of candidates' policy positions, candidates' valence, and voters' information on electoral outcomes as well as to evaluate the performance of the electoral process.

First, we calculate the most liberal policy position that would have allowed a high and a low-valence democratic candidate to win the elections in 1968 and in 1972 respectively. Then, we verify whether any of the democratic candidates at the primary elections had the required policy position. We find that while in 1968 none of the democratic candidates participating in the democratic presidential primaries could have defeated Nixon, virtually all of the participants in the 1972 democratic primaries (other than McGovern) could have won the presidency. Second, we analyze the effect of asymmetric information on the outcome of the two elections and find that had all voters been aware of Nixon's low valence, Humphrey would have been elected President in 1968. Third, we provide an example where, although the incumbent President is reelected, the incumbent has a disadvantage.

In this paper we have investigated individual voting behavior in consecutive elections for the same public office. The methodology we propose is quite general and can be extended to address a variety of issues related to voting in multiple elections.

References

[1] Ansolabehere, Stephen, and James M. Jr Snyder. 2000. "Valence Politics and Equilibrium in Spatial Election Models." *Public Choice,* 103: 327-336.

[2] Bailey, Michael. 2001. "Ideal Point Estimation with A Small Number of Votes: A Random-Effects Approach." *Political Analysis,* 9: 192-210.

[3] Bernhardt, Daniel M., and Daniel E. Ingberman. 1985. "Candidate Reputation and the Incumbent Effect." *Journal of Public Economics,* 27: 47-67.

[4] Black, D. 1958. "Theory of Committees and Elections." *Cambridge University Press.*

[5] Brady, Henry E. 1989. "Factor and Ideal Point Analysis for Interpersonal Incomparable Data." *Psychometrika,* 54: 181-202.

[6] Brady, Henry E. 1990. "Traits Versus Issues: Factor Versus Ideal-Point Analysis of Candidate Thermometer Ratings." *Political Analysis,* 2: 97-129.

[7] Cahoon, Lawrence, Melvin J. Hinich, and Peter C. Ordeshook. 1978. "A Statistical Multidimensional Scaling Method Based on the Spatial Theory of Voting." In *Graphical Representation of Multivariate Data,* ed. P.C. Wang. *New York: Academic Press.*

[8] Campbell, Angus, Phillips E. Converse, Warren E. Miller, and Donald E. Stokes. 1960. "The American Voter." *New York: Wiley.*

[9] Coombs, Clyde. 1964. "A Theory of Data." *New York: Wiley.*

[10] Cooper, Alexandra, and Michael C. Munger. 2000. "The (un)predictability of Primaries With Many Candidates: Simulation Evidence." *Public Choice,* 103: 377-355.

[11] Cuckierman, Alex. 1991. "Asymmetric Information and the Electoral Momentum of Public Opinion Polls." *Public Choice,* 70: 181-213.

[12] Davis, James W. 1997. "U.S. Presidential Primaries and the Caucus-Convention System." *Greenwood Press.*

[13] Degan, Arianna. 2003. "Essays on Voting." *Ph.D. Dissertation. University of Pennsylvania.*

[14] Downs, A. 1957. "An Economic Theory of Democracy." *New York: Harper and Row.*

[15] Enelow, James M., and Melvin J. Hinich. 1982a. "Ideology, Issues and The Spatial Theory of Elections." *American Political Science Review,* 76: 493-501.

[16] Enelow, James M., and Melvin J. Hinich. 1982b. "Non-Spatial Candidate Characteristics and Electoral Competition." *Journal of Politics,* 44: 115-130.

[17] Enelow, James M., and Melvin J. Hinich. 1984. "A Spatial Theory of Voting: An Introduction." *New York: Cambridge University Press.*

[18] Ferejohn, John. 1986. "Incumbent Advantage and Electoral Control." *Public Choice,* 50: 5-25.

[19] Fiorina, Morris P. 1977. "An Outline for a Model of Party Choice." *American Journal of Political Science,* 21: 601-626.

[20] Fiorina, Morris P. 1981. "Retrospective Voting in American National Elections." *New Haven University Press.*

[21] Green, Donald P., and Bradley Parmquist. 1990. "On Artifacts of Partisan Instability." *American Journal of Political Science,* 34: 872-902.

[22] Groseclose, Tim. 1999. "Character, Charisma, and Candidate Location: Downsian Models When One Candidate Has a Valence Advantage." *Unpublished Manuscript, Stanford University.*

[23] Heckman, James J., and James M. Jr. Snyder. 1997. "Linear Probability Models of the Demand for Attributes with an Empirical Application to Estimating the Preferences of Legislators." *Rand Journal of Economics,* 28: 142-189.

[24] Himmelweit, Hilde T., Marianne Jaeger Biberian, and Janet Stockdale. 1978. "Memory for Past Vote: Implications of a Study of Bias in Recall." *British Journal of Political Science*, 8: 365-384.

[25] Jackson, John E. 1975. "Issues, Party Choices, and Presidential Votes." *American Journal of Political Science,* 19: 161-185.

[26] Jackson, John E., and Charles H. Franklin. 1982. "The Dynamics of Party Identification." *American Political Science Review,* 77: 957-973.

[27] Jacobson, Gary. 1990. "The Electoral Origin of Divided Government: Competition in the U.S. House of Representatives 1990." *Boulder, CO: Westview.*

[28] Jennings, M. Kent, Laura Stoker, and Jake Bowers. 1999. "Politics Across Generations." *Paper presented at the annual meeting of the American Political Science Association in Atlanta.*

[29] Jones, Calvin C., and Benjamin I. Page. 1975. "Reciprocal Effects of Policy Preferences, Party Loyalties and the Vote." *American Political Science Review,* 73: 1071-1089.

[30] Lewis, Jeffrey B. 2001. "Estimating Voter Preference Distributions from Individual-level Data." *Political Analysis,* 9: 275-297.

[31] Londregan, John. 2000. "Estimating Legislator's Preferred Points." *Political Analysis,* 8: 35-56.

[32] Markus, Gregory B., and Phillips E. Converse. 1979. "A Dynamic Simultaneous Equation Model of Electoral Choice." *American Political Science Review,* 73: 1055-1070.

[33] Merril, Samuel III. 1988. "Making Multicandidate Elections More Democratic." *Princeton University Press.*

[34] McCarty Nolan M., Keith T. Poole, and Howard Rosenthal. 1997. "Income Redistribution and the Realignment of American Politics". *American Enterprise Institute for Public Policy Research.*

[35] Miller, Warren E. and J. Merril Shanks. 1996. "The New American Voter." *Harvard University Press.*

[36] Mueller, Dennis C. 1989. "Public Choice II." *New York: Cambridge University Press.*

[37] Page, Benjamin, and Calvin C. Jones. 1979. "Reciprocal Effects of Policy Preferences, Party Loyalties and the Vote." *American Political Science Review,* 74: 1071-1089.

[38] Poole, Keith T. 1998. "Recovering a Basic Space from a Set of Issues Scales." *American Journal of Political Science,* 43: 954-993.

[39] Poole, Keith T., and Howard Rosenthal. 1984. "US Presidential Elections 1968-80: A Spatial Analysis." *American Journal of Political Science,* 28: 282-312.

[40] Poole, Keith T., and Howard Rosenthal. 1997. "Congress: A Political Economic History of Roll Call Voting." *New York: Oxford University Press.*

[41] Poole, Keith T., and Howard Rosenthal. 1999. "D-Nominate after 10-years: A Comparative Update to Congress: A Political Economic History of Roll Call Voting." Available at *http://voteview.uh.edu/prapsd99.pdf.*

[42] Rabinowitz, George. 1978. "On the Nature of Political Issues: Insights from a Spatial Analysis." *American Journal of Political Science,* 22: 793-817.

[43] RePass, David E. 1971. "Issue Salience and Party Choice." *American Political Science Review,* 65: 389-400.

[44] Ricker, William H. 1986. "The Art of Political Manipulation." *New Haven: Yale University Press.*

[45] Scammon, Richard.M., and Ben J. Wattenberg. 1971. "The Real Majority." *New York: Coward-McCann.*

[46] Shachar, Roni. 2000. "Party Loyalty as Habit Formation." Forthcoming *Journal of Applied Econometrics.*

[47] Stokes, Donald E. 1963. "Spatial Models of Party Competition." *American Political Science Review,* 57: 368-377.

[48] Stokes, Donald E. 1992. "Valence Politics." In "Electoral Politics," ed. Dennis Kavanagh. *Oxford: Clarendon Press; New York: Oxford University Press.*

[49] White, Theodore. 1969. "The Making of the President, 1968." *New York Atheneum Publishers.*

[50] White, Theodore. 1973. "The Making of the President, 1972." *New York Atheneum Publishers.*

[51] Wright, Gerald C. 1992. "Errors in Measuring Vote Choice in the National Election Studies 1952-88." *American Journal of Political Science,* 37: 291-316.

8 Appendix A

As we did throughout the whole paper, we let $z = \lambda(H - L)$. For any given group-specific probability of being informed, distribution of ideal points, candidates' positions in period 1, and proportion of voters in each group, it is possible to calculate the expected vote shares in the first period conditional on candidates' valence. Because within each party identification we have a continuum of voters, and at each policy point with positive mass there is a constant fraction of voters that becomes informed, the actual vote share corresponding to the realization of any given state, in large samples, is the same as the expected vote share conditional on the same state. Let $C_1 = [n_d(1 - m_d)Y_d(m_1) + n_r(1 - m_r)Y_r(m_1) + n_i(1 - m_i)Y_i(m_1)]$

Table A.1. Vote shares

state at t=1	share π_D^1
HH	$C_1 + [n_d m_d Y_d(m_1 + \frac{z}{2g_1}) + n_r m_r Y_r(m_1 - \frac{z}{2g_1}) + n_i m_i Y_i(m_1)]$
HL	$C_1 + [n_d m_d Y_d(m_1 + \frac{z}{2g_1}) + n_r m_r Y_r(m_1 + \frac{z}{2g_1}) + n_i m_i Y_i(m_1 + \frac{z}{g_1})]$
LH	$C_1 + [n_d Y_d(m_1 - \frac{z}{2g_1}) + n_r m_r Y_r(m_1 - \frac{z}{2g_1}) + n_i m_i Y_i(m_1 - \frac{z}{g_1})]$
LL	$C_1 + [n_d m_d Y_d(m_1 - \frac{z}{2g_1}) + n_r m_r Y_r(m_1 + \frac{z}{2g_1}) + n_i m_i Y_i(m_1)]$

It can be shown that while $\pi_D^1(HL) > \pi_D^1(HH) > \pi_D^1(LH)$ and $\pi_D^1(HL) > \pi_D^1(LL) > \pi_D^1(LH)$, we cannot say anything a priori about $\pi_D^1(HH)$ and $\pi_D^1(LL)$. However, these two shares will be different except for the particular case in which $[n_d m_d Y_d(m_1 + \frac{z}{2g_1}) + n_r m_r Y_r(m_1 - \frac{z}{2g_1})] = [n_d m_d Y_d(m_1 - \frac{z}{2g_1}) + n_r m_r Y_r(m_1 + \frac{z}{2g_1})]$. It follows that the state (and, more importantly, the incumbent's type) is perfectly revealed by the vote share with the only exception given by parameters satisfying the above condition.

9 Appendix B

The 1972 CPS survey was conducted by the Center for Political Studies of the Institute for Social Research at the University of Michigan as part of a series of studies on national elections produced by the political behavior program of the Survey Research Center.

The survey was administered to 2705 respondents and contains 1070 variables. The primary objective of this study was the analysis of the current attitudes and voting patterns of a cross-section of American citizens. The respondents were sent

questionnaires in two waves: the first before the election (between September 1 and November 6) and a second after the election (between November 7 and February 13).[82] The number of respondents that have both a pre-election and a post-election interview was 2285.

Of central interest for our analysis are the variables on voting choices in the 1968 and 1972 presidential elections. In the 1972 pre-election interview, respondents were asked whether they voted in the 1968 national election and who they voted for in the presidential election; analogous questions are asked in the post-election wave for the 1972 election.

In our analysis we concentrate on respondents who participated in both the 1968 and 1972 elections, who voted either for the republican or the democratic candidate, and for whom we have information about their race, gender, age, region, education, income, and party identification.

Of the 2285 observations, the respondents who voted in both 1968 and 1972 numbered 1312. Table B.1 shows the extent of abstention in the initial sample. Only 1246 disclosed who they voted for. Of these, 1113 voted either for a republican or democratic candidate in both elections. For 30 of them we have missing data on individual characteristics. It follows that our final sample has 1083 observations.

Table B.1. Abstention

1968\1972	vote 1972	abstain 1972	missing	Total
vote 1968	1312	174	1	1487
abstain 1968	77	164	0	241
missing	273	283	1	557
Total	1662	621	2	2285

[82] In 1972 CPS survey respondents are given two types of forms. One part of respondents answer to Form I and the other to Form II. There are two types of half samples. The first type has some respondents receiving Form I and others Form II in the pre-election interview. The second type has the first half receiving Form I and the second half receiving Form II in the pre-election inteview. In the half-sample out-of-sample predictions (section 5.4) we use the second type.

10 Appendix C: Figures

Figure 1.1. Ideal Points by Race: Aggregate

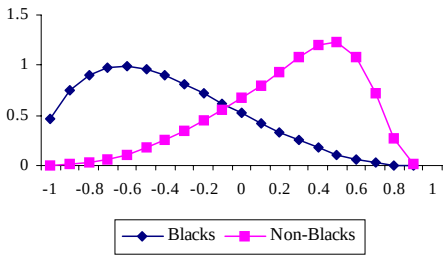


Figure 1.2 . Ideal Points by Education: Aggregate

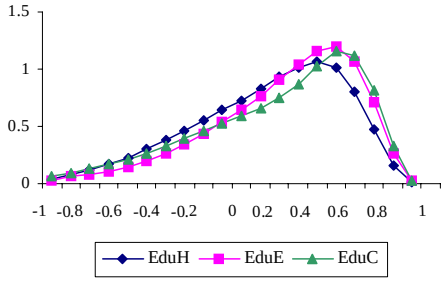


Figure 1.3. Ideal Points by Region: Aggregate

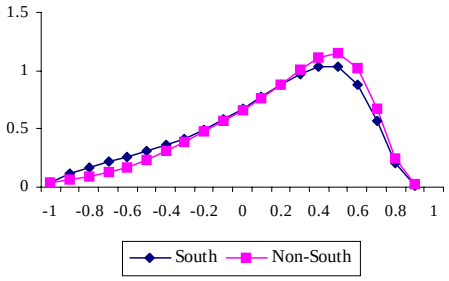


Figure 1.4. Ideal Points by Gender: Aggregate

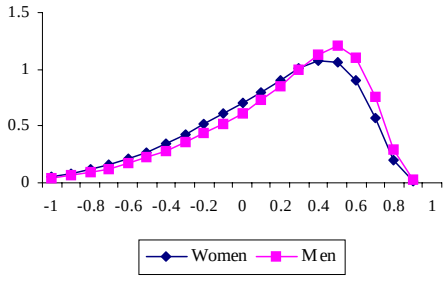


Figure 1.5. Ideal Points by Income: Aggregate

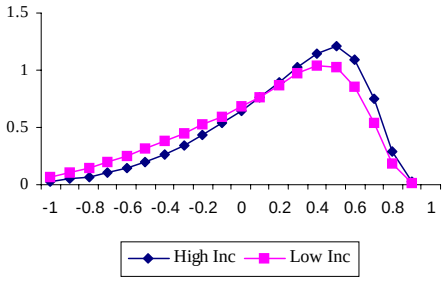


Figure 1.6. Ideal Points by Age: Aggregate

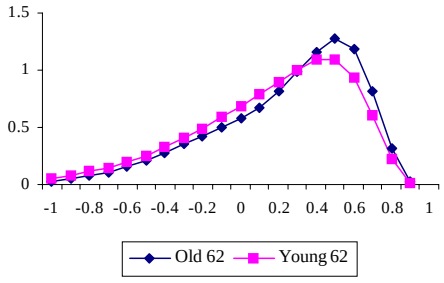


Figure 1.7. Ideal Points by Race: Democrats

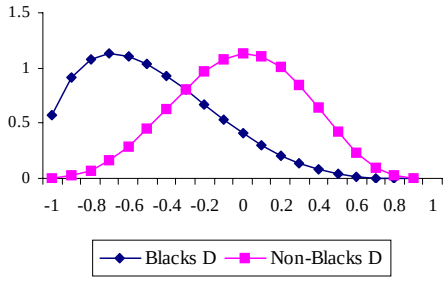


Figure 1.8. Ideal Points by Education: Democrats

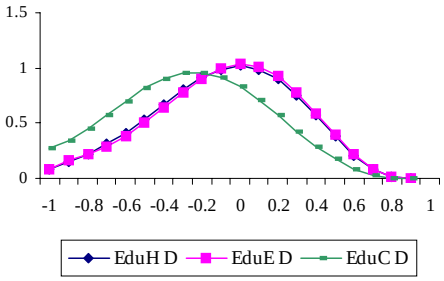


Figure 1.9. Ideal Points by Region: Democrats

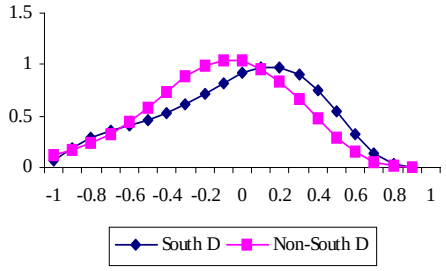


Figure 1.10. Ideal Points by Gender: Democrats

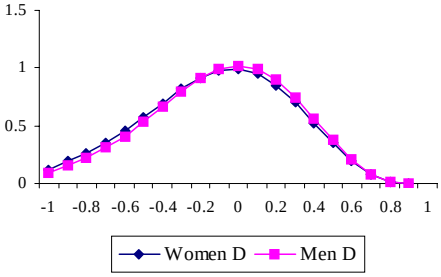


Figure 1.11. Ideal Points by Income: Democrats

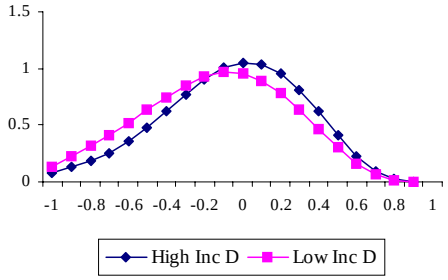


Figure 1.12. Ideal Points by Age: Democrats

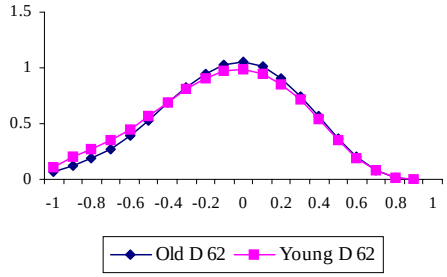


Figure 1.19. Ideal Points by Race: Independents

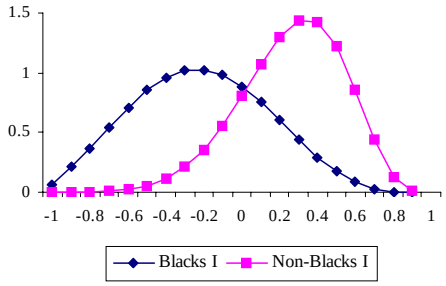


Figure 1.20. Ideal Points by Education: Independents

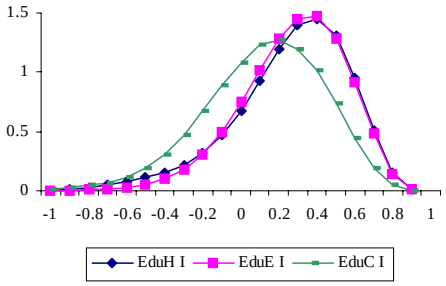


Figure 1.21. Ideal Points by Region: Independents

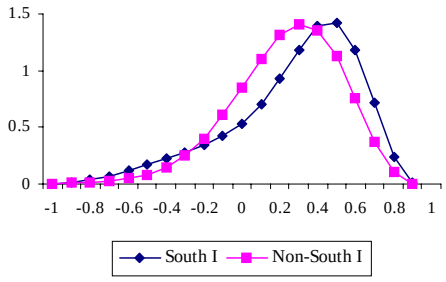


Figure 1.22. Ideaal Points by Gender: Independents

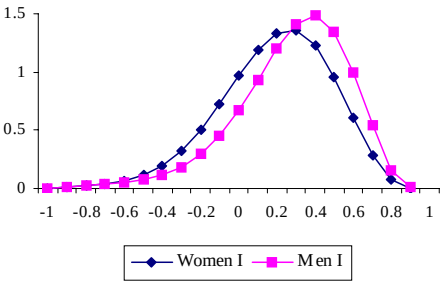


Figure 1.23. Ideal Points by Income: Independents

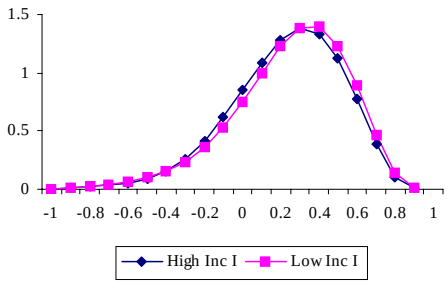


Figure 1.24. Ideal Points by Age: Independents

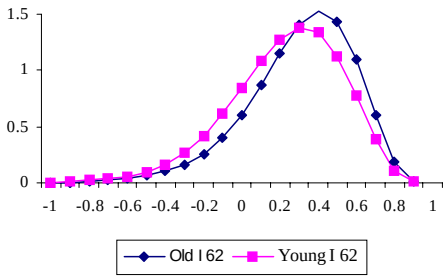


Figure 1.25. Distribution of Ideal Points by Pid

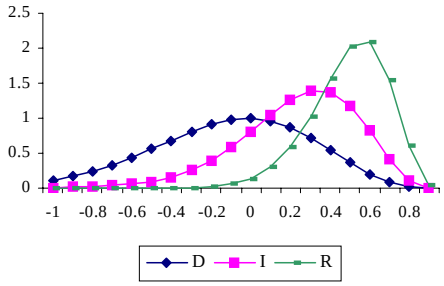
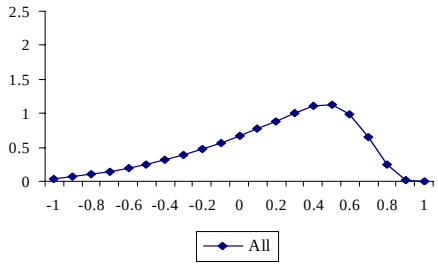


Figure 1.26. Aggregate Distribution of Ideal Points



11 Appendix D: Tables

Table D.1. Prediction of aggregate dynamic voting profiles on the second half sample

Profiles	Actual	Predicted
DD	28.66	24.03
DR	12.06	12.92
RD	5.34	7.47
RR	53.95	55.57
nobs	506	506
$X^2_{(3)}$		8.132

Table D.2. Prediction of aggregate static voting profiles on the second half sample

1968	Actual	Predicted	1972	Actual	Predicted
D	40.71	36.96	D	33.99	31.51
R	59.29	63.04	R	66.01	68.49
$X^2_{(1)}$		3.064	$X^2_{(1)}$		1.450

Table D.3. Prediction of dynamic voting profiles by pid on the second half sample

Profiles	Actual	Predicted	Actual	Predicted	Actual	Predicted
	Party D	Party D	Party R	Party R	Party I	Party I
DD	54.55	48.26	2.33	2.35	24.26	16.18
DR	20.71	22.26	2.33	2.00	11.76	13.15
RD	4.54	5.64	4.07	3.21	8.09	15.54
RR	20.20	23.85	91.28	92.44	55.88	55.13
nobs	198	198	172	172	136	136
$\mathcal{X}^2_{(3)}$		3.362		0.5158		10.559

Table D.4. Prediction of aggregate dynamic voting profiles on the first half sample

1968-72	Actual	Predicted
DD	26.17	30.47
DR	12.31	11.36
RD	7.63	5.31
RR	53.90	52.86
nobs	577	577
$\mathcal{X}^2_{(3)}$		9.92

Table D.5. Prediction of aggregate static voting profiles on the first half sample

1968	Actual	Predicted	1972	Actual	Predicted
D	38.47	41.83	D	33.80	35.78
R	61.53	58.17	R	66.20	64.22
$\mathcal{X}^2_{(1)}$		2.674	$\mathcal{X}^2_{(1)}$		.987

Table D.6. Prediction of dynamic voting profiles by pid on the first half sample

Profiles	Actual	Predicted	Actual	Predicted	Actual	Predicted
	Party D	Party D	Party R	Party R	Party I	Party I
DD	53.48	57.32	1.60	2.89	15.72	24.24
DR	19.57	18.96	2.66	1.50	13.21	12.03
RD	4.78	4.11	3.72	2.34	16.35	10.54
RR	22.17	19.61	92.02	93.26	54.72	53.18
nobs	230	230	188	188	159	159
$\mathcal{X}^2_{(3)}$		1.6568		4.334		10.105

Table D.7. Prediction on GSS data: aggregate dynamic voting profiles

1968-72	Actual	Predicted
DD	33.49	31.84
DR	11.79	11.56
RD	4.40	6.34
RR	50.31	50.26
nobs	639	639
$\mathcal{X}^2_{(3)}$		4.3648

Table D.8. Prediction on GSS data: aggregate static voting profiles

1968	Actual	Predicted	1972	Actual	Predicted
D	45.28	43.40	D	37.89	38.18
R	54.72	56.60	R	66.11	61.82
$\mathcal{X}^2_{(1)}$		0.925	$\mathcal{X}^2_{(1)}$		0.0216

Table D.9. Prediction on GSS data: dynamic voting profiles by pid

Profiles	Actual	Predicted	Actual	Predicted	Actual	Predicted
	Party D	Party D	Party R	Party R	Party I	Party I
DD	65.78	60.63	0.51	2.98	22.03	21.00
DR	15.59	18.29	1.53	1.77	17.51	12.40
RD	4.18	4.38	2.04	3.10	7.34	12.83
RR	14.45	16.69	95.92	92.15	53.11	53.77
nobs	263	263	196	196	177	177
$\mathcal{X}^2_{(3)}$		3.016		5.090		7.992

Table D.10. Prediction of voting profiles of 1968-only voters

Profiles	Data	Model	Data	Model	Data	Model	Data	Model
	All	All	PartyD	PartyD	PartyR	PartyR	PartyI	PartyI
D	47.51	45.87	78.57	75.80	11.90	7.26	27.27	29.65
R	52.49	54.13	21.43	24.20	88.10	92.74	72.73	70.35
nobs	181	181	84	84	42	42	55	55
$\mathcal{X}^2_{(1)}$		0.197		0.3527		1.348		0.1485

Table D.11. Prediction of voting profiles of 1972-only voters*

Profiles	Data	Model	Data	Model	Data	Model	Data	Model
	All	All	Party$D$	Party$D$	Party$R$	Party$R$	Party$I$	Party$I$
D	40.71	36.91	58.79	54.96	9.30	10.69	35.33	28.90
R	59.29	63.09	41.21	45.04	90.70	89.31	64.67	71.10
nobs	452	452	199	199	86	86	167	167
$\mathcal{X}^2_{(1)}$		2.801		1.180		0.173		3.36

*Includes respondents that voted for Wallace in 1968

Table D.12. Prediction of voting profiles of voters who voted for Wallace in 1968[83]

Profiles	Data	Model	Data	Model	Data	Model	Data	Model
	All	All	PartyD	PartyD	PartyR	PartyR	PartyI	PartyI
D	21.19	26.91	25.45	39.30	7.20	5.00	23.26	20.23
R	78.81	73.09	74.55	60.71	92.80	95.00	76.74	79.77
nobs	118	118	55	55	20	20	43	43
$\mathcal{X}^2_{(1)}$		1.967		4.42		0.145		0.243

[83]Among the respondents for whom we know all individual characteristics, there are 118 who voted for Wallace in 1968 and in 1972 voted for one of the two major parties' candidates (only 3 voted for the independent candidate in 1972).