A Dynamic Model of Voting*

Arianna Degan
University of Pennsylvania

January 18, 2003

Abstract

We propose and estimate a dynamic model of voting with asymmetric information which incorporates 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, repeated voting, valence.

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

1 Introduction

In representative democracies elected politicians make policy-relevant decisions on behalf of their constituents. The main way citizens participate in the political process and hence may affect policies is through their decisions to go to vote and support a particular candidate. It follows that individual voting behavior may contain information on citizens' political preferences.

Many researchers in political science have focused on the characterization of 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 such as honesty, charisma, integrity, trustworthiness, or leadership). While voters will in general differ with respect to their policy and party preferences, they will typically agree that candidates with relatively higher valence are preferable.[2]

The following observations emerge from data on voting in two consecutive U.S. presidential elections. First, we observe all possible voting profiles: there are individuals who vote for the democratic or republican candidate twice, individuals who vote for the democratic candidate in the first election and for the republican candidate in the second election and vice versa. Second,"voting persistence" and "switching behavior" are both quantitatively significant phenomena: while a large majority of voters, around 80 percent, votes for the same party's candidate in two consecutive elections, the remaining 20 percent votes 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.

In this paper we propose and estimate a dynamic model of voting with asymmetric information which incorporates the three aspects of individual voting behavior mentioned above and allows us to provide an equilibrium interpretation of the empirical evidence. Using individual-level data on voting decisions in two consecutive presidential elections, we identify and estimate (1) the distribution of voters' policy preferences and (2) candidates' valence.

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 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

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

[2] For example, according to Hinich (1982b) valence includes those candidates' characterisitics not related to policies that affect voter evaluations of each candidate, which are beyond that candidates' immediate control. See also, Stokes (1963), Ansolabehere and Snyder (2000). In particular our definition of valence does not include a candidate's ability to implement specific policies, where the opinions of voters may disagree.

is a continuum of voters who care about both the policy position of the winning candidate in each election and his valence. Voters observe candidates' policy positions but not their valence. Voters are heterogeneous with respect to their party identification, information status, policy preferences and demographic characteristics.

We introduce two roles for party identification. First, we allow the distribution of voters' policy preferences to differ by party identification. Second, we assume that party identification has an impact on the access to information. Voters with different party identification may receive different signals on candidates' valence.

The model has a unique equilibrium strategy which induces 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 conditional on individual characteristics. The equilibrium electoral outcome depends on candidates' policy positions and valence, voters' preferences, and the degree of asymmetric information, and reveals information about candidates' valence.

We estimate our structural model using observations on individual 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 preference and provide insights 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 the policy preferences of voters regardless of their party identification (e.g., blacks are more liberal than non-blacks), whereas other characteristics have different effects on the policy preferences of voters depending on their party identification (e.g., while more educated Democrats are relatively more liberal, the more educated among Republicans tend to be relatively more conservative). Our results suggest that it is important to break down the aggregate relationship between individual characteristics 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) had also low valence. This result is perhaps surprising given that Nixon won the 1968 election by a very small margin and then he won again the 1972 election by a margin as large as that of Johnson in 1964 or Roosevelt in 1936. However, the results are consistent with anecdotal accounts of the events surrounding the 1968 and 1972 elections.

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.

Our two main findings can be summarized as follows. First, while in 1968 none of the democratic candidates who participated 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 briefly discuss the relation of our work to the existing literature. First, most of the existing empirical economic 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; 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 which identifies fundamental utility parameters from observed optimal choices and 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. 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" biases (see, e.g., Markus and Converse 1979) and are not interpersonally comparable (see, e.g., Brady 1989), we use this information to externally validate our empirical results instead of using it as an input in the estimation.

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 1984; and Poole 1998).[3] 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 which allow 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.

In section 2 we describe the model. In section 3 we describe the equilibrium and 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 $y_c \in [-1, 1]$, $y_D < y_R$, which corresponds to the traditional liberal-conservative dimension, and valence $x_c \in \{L, H\}$, $L < H$, where L and H denote low and high valence respectively. Both y_c

[3] Sachar (2000) uses data on repeated voting to estimate a model of habit persistence.

and x_c are exogenously given and fixed. In period 2, the incumbent President (that is, the candidate who won 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.[4]

Voters are heterogeneous along three dimensions which we label as *party identification*, *information status* and *policy preferences*. 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 that she considers herself a Democrat; when a voter has a republican party identification, $k_j = r$, it means that she considers herself a Republican; and when a voter has an independent party identification, $k_j = i$, it means that she does not feel attached to any particular party and equivalently she considers herself 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 the 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 she remains uninformed. Information status is fixed during the two periods, that is, if a voter is informed in the first period she will also be informed in the second period. In each period the informed voters receive a signal about candidates' valence. 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, independently on her 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 voters 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.[5] 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

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

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

feel attached to a party are more likely to be informed about that party's candidate.[6] For tractability of the model we assume that party identification is exogenous and for identification purposes we assume that it is fixed.[7] These assumptions can be partially justified by the fact that we restrict our analysis to short-term dynamics.[8]

Each voter j has a preferred policy position, or ideal point, y_j which takes value in the interval $[-1, 1]$. We assume that there is a relationship between voter j's ideal point, her party identification k and her individual characteristics W_j, where W_j belongs to the space of individual characteristics W. In particular, we assume that voters from group k have policy positions distributed according to the function $Y_k(\cdot|W)$ defined in the interval $[-1, 1]$. The assumption that the distribution of voters' ideal points has full support regardless of party identification, is justified by the evidence that voters' self-placement along a liberal-conservative scale is distributed on the full support of such scale irrespective of voters' party identification.[9]

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

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

where y_c^t is the policy position of candidate c at time t, and 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, $s_k^t(.)$, n_k, $Y_k(.)$, $\forall k \in K$).

There is no abstention in the model and voters are assumed to vote sincerely: given their party identification, information status, ideal point, signal, and their beliefs about

[6] This assumption can be justified on the ground 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 which they talk about their party's candidates.

[7] For the purposes of empirical analysis, party identification is measured only once in our dataset. See section 5.1.

[8] 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. Jennings, Stoker, and Bowers (1999) find that party identification and more in general individual political preferences are mainly affected by the political socialization within the family. We are aware of the fact that party identification is subject also to 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 showed 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 voting behavior, candidate evaluations, issues proximity and retrospective performance evaluation".

[9] In the CPS survey respondents are asked to place themselves on a 7-points liberal-conservative scale. We find that all seven points have positive mass irrespective of the party identification group. Similar results are encountered in the summer1969 Gallup opinion polls.

candidates' valence, they vote for the candidate that if elected gives them higher expected utility. As tie breaking rule we assume that when a voter is indifferent between the two candidates she votes for each of them with equal probability.

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 voters observe the identity of the two competing candidates (that is, who is R and who is D) and their policy positions; the informed voters receive a signal about candidates' valence. During period 1 they vote and at the end of the same period they observe the outcome of the election.[10]. They 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.[11] 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 candidates' valence.[12] During period 2 all voters vote and at the end they observe the electoral outcome.

3 Strategies and Equilibrium

Let $\Pi_D^1 \subseteq [0,1]$ be the set of period 1's vote shares for candidate D. A voting strategy for voter j with party identification $k_j = k$ and ideal point y_j is a pair of voting rules (v_{jk}^1, v_{jk}^2) which assign to voter j the candidate to vote for in each period. In particular, the voting rule in period 1, $v_{jk}^1 : S_k \times I -> \{D, R\}$, is defined over the possible signals in that period and on the voter's information status; the voting rule in period 2, $v_{jk}^2 : S_k^2 \times I \times \Pi_D^1 -> \{D, R\}$, is defined over the possible signals in both periods, on the voter's information status and on period 1's vote share.

Let $u_j(x_D^t, x_R^t, y_D^t, y_R^t, y_j) = U_j^R - U_j^D = \lambda(x_R^t - x_D^t) + [(y_j - y_D^t)^2 - (y_j - y_R^t)^2]$ 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 of a voter from group k and information status I about the distribution of candidates' valence.

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

$$v_{jk}^{1*}(s_k^1, I_j) = \begin{cases} R & \text{if } E[u_j(x_D^1, x_R^1, y_D^1, y_R^1, y_j)|s_k^1, I_j, y_D^1, y_R^1, y_j] > 0 \\ D & \text{if } E[u_j(x_D^1, x_R^1, y_D^1, y_R^1, y_j)|s_k^1, I_j, y_D^1, y_R^1, y_j] < 0 \end{cases}$$

and

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

[11] If instead of assuming that voters know the distribution of policy preferences but do not observe the winning candidate's valence we assume that voters 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.

[12] The signal is reduntant for the informed voters of the same party as the incumbent.

$$v_{jk}^{2*}(s_k^1, s_k^2, I_j, \pi_D^1) = \begin{cases} R & \text{if } E[u_j(x_D^2, x_R^2, y_D^2, y_R^2, y_j)|s_k^1, s_k^2, I_j, \pi_D^1, y_D^2, y_R^2, y_j] > 0 \\ D & \text{if } E[u_j(x_D^2, x_R^2, y_D^2, y_R^2, y_j)|s_k^1, s_k^2, I_j, \pi_D^1, y_D^2, y_R^2, y_j, \pi_D^1] < 0 \end{cases}$$

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. The two elections are linked by the effect of the aggregate outcomes on beliefs. However, sincere voting and the fact that the incumbent's type and policy position as well as voter's information status are constant over time imply that period 1's voting behavior is independent of period 2's. Notice that even though the equilibrium strategy is unique, the individual's actual voting choice depends on the realized information status and on which of the finite and discrete combination 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 $x = \{x_D^1, x_R^1, x_D^2, x_R^2\}$ as its generic element. The result of proposition 1 together with the assumption that the utility is quadratic with respect to the distance 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 is a cut-off strategy. For any voter* j*, party identification* k*, ideal point* y_j*, signal* s_k^t*, information status* I_j *and aggregate vote share* π_D^{t-1}*, there exists a cut-off point* $\overline{y}(s_k^t, I_j, \pi_D^{t-1})$ *such that: if* $y_j < \overline{y}(s_k^t, I_j, \pi_D^{t-1})$ *it is optimal to vote* D*, if* $y_j > \overline{y}(s_k^t, I_j, \pi_D^{t-1})$ *it is optimal to vote* R*, if* $y_j = \overline{y}(s_k^t, I_j, \pi_D^{t-1})$ *the voter is indifferent and will vote* $D(R)$ *with probability* $\frac{1}{2}$.
(ii) Each cut-off point is of the form: $\overline{y}(s_k^t, I_j, \pi_D^{t-1}) = m_t - \frac{\lambda\Delta_j^t}{g_t}$*,where*
$\Delta_j^t = E[(x_R^t - x_D^t)|s_k^t, I_j, \pi_D^{t-1}]$ *is the expected difference in valence between candidate* R *and* D *perceived by individual* j *at time* t;
$m_t = \frac{y_R^t + y_D^t}{2}$ *is the "midpoint" of the candidates' policy positions ; and*
$g_t = 2(y_R^t - y_D^t)$ *is the "gap" between the two candidates' policy positions.*

Proof: Voter j's optimal strategy is to vote $R(D)$ if $Eu_j > 0 (< 0)$. Voter j will vote $R(D)$ 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 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. Proposition 2 tells us that if a voter's ideal point in any period is to the left of her cut-off point, she will vote for D, and if it is to the right of her cut-off point, she will vote for R. Therefore, the locations of a voter's cut-off points in the two periods and their relationship with respect to the voter's ideal point determine the voter's dynamic voting choice. Let $V = \{RR, RD, DR, DD\}$ be the set of dynamic optimal voting profiles, where RR denotes the profile of a voter who voted for the republican candidate in two consecutive elections, RD denotes the profile of a voter who voted for the republican candidate in the first period

and for the democratic candidate in the second period and so on. 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. Take and informed voter j with party identification k and ideal point y_j. When $y_{2k} < y_j < y_{1k}$ her voting profile will be DR, when $y_{1k} < y_j < y_{2k}$ her voting profile will be RD. Analogously, the uninformed voter will generate the profile DR when $y_{20} < y_j < y_{10}$ and RD when $y_{10} < y_j < y_{20}$.

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, on her information and on the candidates' actual valence through the realized signals and the information contained in period $1'$s electoral outcome. In each period t, any cut-off point can be expressed just 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. 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 the possible states which are compatible with the observed voting patterns. If, for example, 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 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, RR. 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 the incompatibility would arise from the behavior of the independents. Specifically, the cut-off points of an informed Independent would be $y_{1i} = m_1$ and $y_{2k} = m_2$, while the cut-off points of an uninformed Independent would be $y_{1o} = m_1$ and $y_{2o} = m_2 - \frac{z}{2g_2}$. Since $y_{1i} > y_{2i}$ and $y_{10} > y_{20}$, Independents in equilibrium cannot generate RD. In general in our model, to generate both switching patterns (DR and RD) within each group, we need the cut-off points of informed and uninformed voters to be such that they can generate different switching behaviors (either $y_{2k} < y_{1k}$ and $y_{20} > y_{10}$ or $y_{2k} > y_{1k}$ and $y_{20} < y_{10}$, $\forall k$). We define a configuration to be the pair of the candidates' policy positions in the two elections and the identity of the incumbent. The details about which states are compatible with different configurations and observing all four voting profiles are contained in Appendix B.
Proposition 2 also allows us to derive the following result about electoral outcomes.

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

Proof: See Appendix A.

The intuition is that for any possible state, the equilibrium share is uniquely determined by the fraction of voters with different party identifications, their probability of being informed

and the distribution of policy positions.[13] 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 what is the vote share corresponding to any particular candidates' valence in period 1. Such share will be different for any valence pair so that by observing π_D^1 voters know x_D^1 and x_R^1.

Corollary: *In equilibrium, π_D^1 completely reveals the incumbent's valence.*

The result 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 of the corollary, 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 outcome.

4 Discussion of Modeling Choices

Before turning to the empirical analysis it is important to stress that there are two main key features in the data. The most important feature is that we observe all four voting profiles (DD, DR, RD, RR) both at aggregate level and within each group k. The second feature of the data is that the four voting profiles are heterogeneous across groups, that is the frequency of each profile changes with party identification. Our model must be able to generate 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, voters' heterogeneity is only due 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 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).[14] 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 generate DD) or DR and RR (when the informed generate RR). To see why, without loss of generality consider the problem faced by a voter from group d when D is the incumbent in period 2. The voter, in each period $t \in \{1, 2\}$, by voting R receives an expected utility of $qH + (1-q)L$ regardless of her information status.[15]

[13] There is a particular combination of parameters in which uninformed voters cannot distinguish between the state in which in period 1 both candidates have a low valence and the one in which both candidates have a high valence. See Appendix A for the details.

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

[15] The only information a voter has about the opposite party's candidate is the prior distribution of types,

In the first period, an uninformed voter will randomize between the two candidates with probability $(\frac{1}{2}, \frac{1}{2})$ because she receives an expected utility of $qH+(1-q)L$ when she votes for either candidate. An informed voter, on the other hand, will vote for D ($H > qH+(1-q)L$) when D has high valence (when she receives the signal $s_d^1 = \{H, 0\}$) and she will vote for R ($L < qH+(1-q)L$) when D has low valence (when she receives the signal $s_d^1 = \{L, 0\}$). In the second period both the informed and uninformed voter know the incumbent's valence. By voting D, a Democrat can receive an expected utility of L when the incumbent has a low valence and H when the incumbent has a high valence. Solving for the optimal voting choice in all the possible states, it turns out that, when the democratic candidate's valence is H the informed can only generate the profile DD and the uninformed the profiles DD and RD. Analogously, when the democratic candidate's valence is L, the informed can only generate the profile RR and the uninformed the profiles DR and RR.

When voters do not care about valence ($\lambda = 0$) we go back to the standard one-dimensional spatial model of voting. 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 $-(y_j - y_D^t)^2 < -(y_j - y_R^t)^2$ she votes for R; she 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 median policy position.[16] 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 then, since period 2's cut-off point is smaller (greater) than period 1's, the model will only generate DD, RR, DR (DD, RR, RD).

A peculiar 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 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 accross 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.[17]

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

[17] In the context our model we could explain the difference in voting patterns across groups only if voters were sorting among party identifications on the basis of their individual characteristics, and if these had opposite 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 enough to justify differences in policy positions across party identifications.

5 Empirical Analysis

5.1 Data

We focus our attention on the 1968 and 1972 U.S. Presidential elections.[18] 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 unsuccessful presidential candidate in 1960, is the candidate for the republican party. In 1968 Nixon wins the election with 43.42% of the popular vote, to Humphrey's 42.72%.[19] 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, with 60.69% of the popular vote, and then resigns in 1974 after Watergate.

We use two sources of data. The first is the 1972 Center for Political Studies (later National Election Studies) data. The second is the Poole and Rosenthal's DW-NOMINATE data on legislators' coordinates.

The 1972 CPS dataset is particularly appropriate for the estimation of our model for different reasons. It is an individual-level dataset which contains observations on voting choices in 1968 and 1972.[20] 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 also contains data on individuals' socio-demographic characteristics and party identification.

We use variables for voters' party identification, their voting choices, their individual characteristics, and candidates' positions.[21] 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.[22] We use a dummy variable in each election, R_{68} and R_{72}, which

[18] There are two main reasons why we concentrate on two periods. First, we have data on individuals' repeated voting, of a reasonable sample size, only for 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.

[19] In 1968, George Wallace, a third candidate, receives the remaining 13.53% of the popular vote. We focus only on the two major parties' candidates.

[20] 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 missreporting in our context, we looked at the 1972-76 panel data and compared the reported vote after the 1972 elections and the recall question on 1972's vote asked before 1976 election. In our selected sample only 7% of respondent missreported their vote. Himmelweit, Biberman, and Stockdale (1978) analyze the vote bias in recalls.

[21] 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 makes the problem of ex-post rationalization less serious.

[22] There is also a party identification variable which uses a 7-points categorization and takes into consid-

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-periods voting profile. 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.[23] We use two dummies for education level: EduH for education levels strictly lower than high school degree, and EDUC for education levels equal to or greater than college degree.[24] We use AGE as a continuous scaled variable for the respondent's age and a dummy variable, MINCOME, for income levels greater than the median.

Table 1 Descriptive statistics

Variables	**All**	**r**	**d**	**i**
R68	655	344	111	200
R72	716	339	177	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[a]	240	104	92	44
MINCOME	615	222	214	179
TOT	1083	360	428	295

[a] Number of respondents with age strictly greater than 61.

The original data set consists of 2705 observations. We select the respondents who voted either for a republican or for a democratic candidate in both 1968 and 1972 and for which we have data on individual characteristics and party identification. The resulting sample contains 1083 observations.[25] Table 1 contains descriptive statistics of the selected sample

eration the strenght of party identification. We choose the 3-point categorization first of all because our model doesn't incorporate the strenght of party affiliation, and second because our measure is more stable over time.

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

[24] The excluded category includes college degree and some college.

[25] Of the 2705 initial observations, 2285 have both a pre and post-election interview. Of these, the respondents who voted in both 1968 and 1972, were 1312. Only 1246 answered to who they voted for. Of these, 1113 voted either Republicans or Democrats in both elections. For 30 of them we have missing data on individual characterisitics.

from the first source of data.

As a measure of candidates' positions on a liberal-conservative scale we use the first dimension of legislators' coordinates estimated with the DW-NOMINATE "constant model" by Poole and Rosenthal. The DW-NOMINATE is a dynamic model that estimates, separately for the House and Senate, legislators' coordinates on a two-dimensional policy space using data on roll call voting. Similarly, the coordinates of Presidents are estimated using their support roll calls.[26] The "constant model" is a version of the DW-NOMINATE model in which candidates' coordinates are constrained to remain constant over the whole candidate's career.

Table 2.Presidential candidate's coordinates[27]

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

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 legislators that served in different Congresses. In particular, we can compare the coordinates of the elected Presidents with the one of their challengers who typically served in Congress in different years.[28] We report the coordinates on a liberal-conservative scale for the presidential candidates considered in our study in table 2.[29]

5.2 Estimation procedure

In the model voters know their own ideal points but the econometrician doesn't. Since we are interested in the link between individual characteristics and policy preferences for each group of voters, we assumed that voter j's ideal point is a party-specific, non-deterministic function of her characteristics W_j and party identification k_j. Here we fix 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_{jk}, r_{jk}).[30] We parametrize the first parameter of such distribution to j's characteristics W_j (BLACKS, EDUH, EDUC, SOUTH, FEMALE, AGE, MINCOME) and we restrict the coefficients r_{jk} to be the same both within and across groups. Since $p_{jk} > 0, r_{jk} > 0$ our parametrization becomes

[26] 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).

[27] The DW-NOMINATE "costant model" coordinates were available on-line from *htpp://voteview.uh.edu.*

[28] A third reason is that the two dimensions are estimated one at a time so that we can just take the first dimension. Poole and Rosenthal (1997, 1999).

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

[30] We choose the beta distribution because it is the more flexible distribution and it is defined on a finite support.

$$p_{jk} = \exp(\beta_k W_j)$$
$$r_k = \exp(r)$$

where β_k and r are preference parameters to be estimated.[31]

Candidates' positions are exogenous parameters which are available in the data. For any candidates' positions and state of the world, the unique equilibrium with sincere voting induces a different voting behavior. We cannot estimate valence directly because different candidates' valences lead to different equilibrium voting behavior and consequently to different conditions on the parameters that we want to estimate. However, since we observe an ex-post voting behavior and for any state there is a unique equilibrium voting behavior, we can estimate the parameters of the model by maximum likelihood conditional on each state. We then pick as estimate of valence the state whose equilibrium maximum likelihood is the highest, which indicates 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 $B_{jk}(\cdot)$ we can write the likelihood of the observed voting profile V_j of individual j from group k conditional on the state $x = (x_D^1, x_R^1, x_D^2, x_R^2)$ as:[32]

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

We fix q to 0.5, which is equivalent to an uninformative prior, and we estimate β_k, r, m_d, m_r, m_i and z conditional on the state x.[33] The parameter z is a composite parameter. As we mentioned is section 3, it is impossible to separately identify the weight λ and the perceived maximum possible difference in candidates' valence $(H-L)$, which is independent on the realized valence. A big value of z may be due to the fact that voters give a high weight to valence or that the potential difference in candidates' valence is big. We could fix either λ or $(H - L)$ to be some arbitrary value but since these two concepts are somehow 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 to estimate instead the composite parameter z.

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, the model puts restrictions on which states are compatible with the observed voting profiles and candidates' configuration. Using the NOMINATE coordinates from table 2, the midpoints and gaps in the two periods are such that $m_1 > m_2$ and $g_1 < g_2$.

[31] In general it is very hard to separately identify the first and second parameter of the beta distribution. Some sort of restrictions are neccessary. We could have opted for different parametrizations but ours is probably the more flexible since we want to allow both the mean and the variance to differ across party identifications.

[32] We use b_{jk} 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 $b(p_k(W_j), r_k)$. The analogous for the cdf.

[33] It is not necessary to fix q. However, when we fix q, we improve the precision of the estimates for the other parameters.

The configuration in 1968 and 1972 together with the fact that in the data 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.[34] We can therefore restrict our attention to three states: $HLLL$, $LHHH$, and $HLHL$.[35] We estimate the model by maximum likelihood separately conditional on each of these states and we obtain that the state corresponding to the highest likelihood is $HLLL$.[36] We report the equilibrium cut-off points corresponding to this state in table 3.

Table 3 HLLL: cut-off points

group	**cut-off point 1**	**cut-off point2**
d	$0.0555 + z/3.164$	-0.008
r	$0.0555 + z/3.164$	$-0.008 + z/3.672$
i	0.0555	-0.008
$I = 0$	0.0555	$0.008 + z/3.672$

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

$$L(V_j|\beta_k, r, m_k, z, y, W_j, k, HLLL) =$$
$$\int_{-1}^{1}[\Pr(DD|y_u)^{DDj} \cdot \Pr(DR|y_u)^{DRj} \cdot \Pr(RD|y_u)^{RDj} \cdot \Pr(RR|y_u)^{RRj}] \cdot b_{jk}(y_u)du$$

where the probabilities of each voting profiles are calculated using the cut-off points corresponding to the state $HLLL$ and DD_j, DR_j, RD_j, and RR_j denote the dummy variables for individual j's voting profile. Using a different notation we can rewrite the above likelihood as:

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

[34] In states $HHHH$ and $HHLH$ Independents cannot generate the profile RD. For any value of $z > 0$, both $y_{1i} > y_{2i}$ and $y_{10} > y_{20}$. In states $LLHL$ and $LLLL$ Democrats cannot generate the profile DR because the conditions that we would need to impose on z to have $y_{1a} > y_{2a}$ and $y_{10} < y_{20}$ are incompatible. In state $LHLH$ Democrats could generate RD only for $q > 0.86$. Even if we were allowing such value of q, the conditions on z imposed by the profiles of Democrats and Republicans would be incompatible with the ones of Independents. We do not have to worry about the particular case in which the aggregate share corresponding o the period 1's state HH is equal to the aggregate share corresponding to period 1's state LL (see Appendix A), because the above inconsistencies remain.

[35] See Appendix B.

[36] Each case requires different restrictions on z in order to generate cut-off points compatible with the four voting profiles. In the first case $z \in (0.23; 3.701)$, in the second case $z \in (1.45; 3.64)$, in the third case $z \in (0.23, 0.27)$. 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 of lees than -1000).

which using the cdf of the beta distribution can be rewritten as :

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

The likelihood of individual $j's$ voting profile unconditional on her party identification and the total loglikelihood can be written respectively as:

$L(V^j|\beta_d, \beta_r, \beta_i, r, m_d, m_r, m_i, z, y, W_j, HLLL) =$
$[L(V_j|\beta_d, r, m_d, z, y, W_j, HLLL)]^{pidRj}.$
$[L(V_j|\beta_r, r, m_r, z, y, W_j, HLLL)]^{pidDj}.$
$[L(V_j|\beta_i, r, m_{i,}, z, y, W_j, HLLL)]^{pidIj}$

$l(V_j|\beta_d, \beta_r, \beta_i, r, m_d, m_r, m_i, z, y, W_j, HLLL) = \sum_j [$
$\ln(L(V_j|\beta_d, r, m_d, z, y, W_j, d, HLLL)) \cdot pidD_j +$
$\ln(L(V_j|\beta_r, r, m_r, z, y, W_j, r, HLLL)) \cdot pidR_j +$
$\ln(L(V_j|\beta_i, r, m_i, z, y, W_j, i, HLLL)) \cdot pidI_j]$
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 4) characterize the distribution of voters' policy positions.[37] To interpret the coefficients on individual characteristics note that a bigger value of any coefficient corresponds to a bigger p_{jk} 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 are voters with such characteristic.[38]

[37] For a complete characterization of the estimated distribution of voters' policy positions we need also the estimate for r (table 5).

[38] This is true when we make the comparison for a fixed r.

Table 4 Estimated policy position parameters

VARIABLE	ESTIMATE	ST-DEV	t-STAT
Democrats			
_CONST	1.423	0.812	1.752
BLACK	-1.098	0.445	-2.463
EDUH	0.020	0.094	0.214
EDUC	-0.428	0.206	-2.080
SOUTH	0.448	0.166	2.694
FEMALE	-0.022	0.079	-0.279
MINCOME	0.223	0.116	1.916
AGE	0.005	0.032	0.151
Republicans			
_CONST	2.542	0.478	5.315
BLACK	-0.701	0.393	-1.782
EDUH	-0.238	0.145	-1.633
EDUC	-0.006	0.126	-0.049
SOUTH	0.002	0.141	0.013
FEMALE	-0.153	0.120	-1.274
MINCOME	0.095	0.117	0.816
AGE	0.061	0.041	1.500
Independents			
_CONST	2.338	0.521	4.486
BLACK	-1.201	0.482	-2.490
EDUH	0.085	0.128	0.666
EDUC	-0.334	0.148	-2.262
SOUTH	0.281	0.182	1.539
FEMALE	-0.236	0.111	-2.122
MINCOME	-.024	0.087	-.273
AGE	0.002	0.028	0.0611

The results on policy preferences are the following. There are some characteristics that have a similar effect on the policy preferences of voters regardless of their party identification and other characteristics that have different effects on the policy preferences of voters depending on their party identification. RACE has a big (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 counterpart, 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 doesn't

help to explain policy preferences. MINCOME has a relatively significant effect only on Democrats. It is interesting to notice that despite the high correlation between income level and education, income level has a separate and opposite effect from education. While higher education makes Democrats more liberal, having a high level of income makes them more conservative. Note also that none of the individual characteristics are very significant among Republicans, probably due to the small number of observations with profile different from *RR*.

A complementary way to analyze policy preferences is to look at the plots of the estimated marginal distributions of voters' ideal points by characteristics both in the aggregate and by party identification (figures 1-26). All the marginal distributions are in line with the above results. Blacks are more liberal than non-blacks both in the aggregate (figure 1) and within each party identification (figures 7-9). Southerners are more conservative than non-southerners (figure 3), however the effect is very significant for Democrats, less so for Independents, and it is not significant at all for Republicans (figures 13-15). 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 in general more conservative than northerners. Although at the aggregate level highly educated voters are more conservative than their less educated counterparts (figure 2), education has a different effect across party identification (figures 10-12). Even though women are more liberal than men among both Independents and Republicans (figures 16-18), the effect of gender disappears in the aggregate (figure 4). As expected from the point estimates, Democrats with an income level lower than the median are more liberal than those with the highest income level (figure 19). Age doesn't have a separate effect on policy preferences; however, Independents more than 62 years old appear to be more conservative than young Independents. Such an effect is probably due to the correlation between age and education.

Considering the relatively homogeneous demographic composition of different groups, our results indicate that to understand the relationship between individual characteristics and policy preferences it is important to consider their interaction with party identification.[39] Overall, Democrats' policy preferences are more heterogeneous than Independents', which in turn are more heterogeneous than Republicans' (figure 25); also the aggregate distribution of ideal points is relatively conservative. Both results are is in accordance with the self-reported liberal-conservative view of the population.[40]

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,

[39] Race is the only individual characterisitic whose distribution among party identifications is very asymmetric.

[40] According to the self-reported positions on a 1-7 points liberal-conservative categorization from the 1972 CPS data, 24% of respondents reports to be liberal, 41% conservative and the remaining 35% report to be middle of the road; the corresponding proportion among republicans are 9%, 60% and 31%; among Democrats are 33%, 26% and 41%; and among Independents are 29%, 37% and 34%. In the summer 1969 Gallup opinion polls, 23% of respondents reports to be conservative, 28% moderately conservative, 18% moderately liberal, 15% liberal and 15% don't know.

policies or parties). Rather, in order to estimate voters' preferences we apply a revealed-preference approach which relies on observed individual voting choices only, 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 sociological studies have said about political preferences of American voters.[41] In addition, our results allow us to disentangle the effect that each characteristic has on voters' policy preferences and get further insight on the relation 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 5), indicate that Democrats and Republicans are more likely to be informed than Independents (68%, 29%, and 18% probability respectively) and that Democrats are more informed than both Republicans and Independents. As we would expect if we think 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. The estimate of r is rather imprecise. Although we restricted r_k not to be party specific, the identification of r separately from the constant terms in p_{jk} is tenuous. It is also hard to separate the effect of the mean of the distribution of policy positions from z, which enters in the expression of the cut-off points. We are mainly interested in the qualitative effects of individual characteristics on policy preferences and on their relative effect among party identifications. Such results as well as the results on the probability of being informed are not sensitive to values of z within its confidence interval. Different values of z only cause the distribution of policy positions to either shrink or spread over the support leaving the main results unchanged. Analogous results are true when we estimate the model for different values of r.

Table 5 Estimated Parameters (Continue)

VARIABLE	ESTIMATE	ST-DEV
m_d	0.6799	0.0724
m_r	0.2870	0.1019
m_i	0.1869	0.0297
z	0.8098	0.2545
$r_d = r_r = r_i$	4.6183	3.368
LogLikelihood		-885.49

Our estimated state of the world is $HLLL$, which means that we estimate Humphrey to have high valence, 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 elections but he was very experienced and respected having served in the Senate almost without interruption from 1949 until he died in 1978. Scammon

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

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 and his extreme ability to deal with foreign affairs; however, as Watergate revealed, he was not a trustworthy politician. Even before Watergate, the secret bombing of Cambodia in 1969, the wide use of impoundment of funds, of veto power, of administrative discretion, may point to Nixon being a low-valence politician.[42] More objective evidence of Nixon's low valence are his "dirty-tricks team", an old institution in American politics that was taken to new heights in 1972; and the "Plumbers", an apparatus organized to plug leaks but that expanded its activities to include a variety of other secret and illegal operations. Famous are the secret effort to discredit and defame Daniel Ellsberg (who published on the New York Times a top secret study on the origin and conduct of the war in Vietnam), the fake cable created in the attempt to link President J. Kennedy to the assassination of south Vietnam's President, and the illegal money-collection during the 1972 campaign. The predatory strategy for Nixon's reelection in 1972 included many attempts to sabotage the campaign of the top democratic front runners in the hope of forcing them out of race so as to face a weak democratic opponent in the general election. Muskie dropped out of race after few primaries followed by Jackson and Humphrey.

George McGovern, a very liberal candidate without the support of the mainstream democrats easily won the democratic nomination basically because he was the only candidate left. The O'Brien case, the Eagelton case and the Salinger affair are only few of the many examples of McGovern inability to make clear decisions and to maintain promises.[43]

Our findings on candidates' valence imply that Nixon was elected President in 1968 even though he was a low-valence politician and he was running against a high valence challenger (Humphrey). What is more interesting is that the 1972 election was for Nixon a landslide

[42] The secret bombing of Cambodia doesn't neccessarily indicate Nixon's low valence but it is one example that shows Nixon's tendency to take actions in big secrecy. Also, the frequent use of impundements of fund and exercise of veto power may indicate Nixon's low valence to the extent in which they are due to his inability to constructively negotiate with the Congress, as a good leader is supposed to do.

[43] After the convention closes, the new candidate assembles the National Committee of his party, appoints a new chairman and vice-chairman. During the convention McGovern asked Lawrence O'Brien twice to be the new chairman. O'Brien was ready to accept upon the closing of the convention when McGovern, after talking to his people, retreated the offer. McGovern chose as his running mate Senator Thomas Eagelton. Only after the decision was maken, Eagelton was found to have had episodes of serious mental illness, been hospitalized three times, and received electrical shock therapy twice. After these facts became known, McGovern mantained his support for Eagelton until he was forced by the pressure of the press, his party and public opinion to withdraw him. McGovern incoherence and untrustworthiness appear evident in his way of facing the problem: one day he called Eagelton saying that although he had 30 editorials against him he was 1000 % with him; that night he informed the press in South Dakota that he was reconsidering the nomination. Mcgovern asked Pierre Salinger, whom he had previously repudiated as vice-chariman of the Democratic National Commettee, to go to Paris to meet with the North Vietnamese to negotiate for peace and the release of American prisoners. The trip didn't lead to any positive results and when it was discovered by the press, McGovern publicly denied having anything to do with it.

even if 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 that allowed Nixon to win both elections were: the conservative constituency, voters asymmetric information, in particular voters' lack of information on Nixon's low valence in 1968, and having an extremely liberal and low-valence democratic candidate in 1972.

Among Democrats, that we found to be relatively conservative, the informed who voted Humphrey in 1968 and switched to Nixon in 1972 (20%) were voters with a moderate policy view who knew that Humphrey had high valence and that McGovern had low valence as well as an extremely liberal policy position. When faced with candidates with equal valence they preferred to vote for the more conservative candidate (Nixon). The vote of informed Republicans and Independents, who both knew that Nixon had low valence, was driven mainly by policy concerns in both periods. The uninformed voters who switched their vote in the opposite direction (RD) were slightly conservative voters who in the first period voted for Nixon driven by policy concerns but switched to McGovern in the second period, despite his liberal policy position, due to the revealed low valence of Nixon. We can also explain why, although Nixon won in 1972 with a much greater margin than in 1968, the proportion of Republicans who voted Nixon in 1972 is smaller than in 1968. The reason is that most Republicans in 1968 were unaware of Nixon's low valence. They voted for Nixon in 1968 on the basis of policy considerations but some of them, after learning that he has low valence, preferred to vote for McGovern despite his extremely liberal policy position.

5.4 Robustness

While in section 4 we discussed the elements of the model which are necessary to explain the qualitative features of the data, here we discuss the ingredients that are important from an empirical point of view. They can be grouped in two categories: information status and party identification. Some are strictly required for identification purposes and others are required to obtain a better fit of the data.

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 which allows 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.

Another assumption related to the information status is that corresponding to each policy position with positive mass there is a fraction m_k that becomes informed.[44] This assumption is sufficient to guarantee that the aggregate vote share perfectly reveals information on candidates' valence. Moreover, because of this assumption, the aggregate voting patterns directly put restrictions on the probability of being informed.

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

None of the assumptions on party identification are necessary for identification; rather, they help to explain better the features of the data.[45] 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 each voter has the same probability of becoming informed ($m_k = m$, $\forall k$), the same signal ($s_k^t = \{x_D^t, x_R^t\}$, $\forall k$) and in which voters policy positions are drawn from a common distribution ($\beta_k = \beta$ $\forall k$), is rejected by both likelihood test and 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 element of party identification at a time. On the basis of goodness of fit test on dynamic voting profiles and on likelihood test, we reject both the model in which preferences are constrained to be the same accross parties and the one in which there is a common probability of being informed.[46] The assumption that individuals with a partisan party identification receive only the signal about their own party's candidate is not an ad hoc nor an identifying assumption. We estimated a model where all the informed voters, independently on their party identification receive signals about both candidates. Such a model leads to a smaller likelihood, therefore, the data seem to support our assumption on the party-specific signal structure, that is the idea that party identification works like an information selection device which restricts the kind of information that one is exposed to or willing to absorb.[47]

5.5 Goodness of Fit

To assess whether our model can reproduce the quantitative features of the data we want to measure how close the predicted voting profiles are to the observed ones. We report the actual and fitted voting profiles on aggregate and by party identification in tables 6.1, 6.2 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. This is calculated by integrating over voters (overall or within a particular group) the individual probability of such profile.

We perform standard goodness of fit tests on both dynamic and static voting profiles and report the relative $\mathcal{X}^2$ test at the end of each table. Table 6.1 shows the results for the four dynamic voting profiles at aggregate level. The value of the test indicates that our model cannot be rejected by the data. Table 6.2 shows the aggregate static voting patterns in 1968 and 1972. The model predicts perfectly both electoral outcomes. In neither year we can reject our model. Analogous results hold for the dynamic voting patterns within each party identification (table 7). The model captures both the effect of party loyalty and switching behavior, and their differences accross parties.

[45] 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 the signals accross parties.

[46] The likelihood of the model in which there is a unique distribution of policy preferences has a loglikelihood of -1100.95 while the one in which there is a common probability of becoming informed is -911.10.

[47] The likelihood of the model with 2 signals and our original model are -887.63 and -885.49 respectively.

Table 6.1 Aggregate Actual and Predicted
Dynamic Voting Profiles 1968-72

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

*The critical value at 5% for a $X^2$with three degrees of freedom is 7.81

Table 6.2 Aggregate Actual and Predicted 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

**The critical value at 5% for a $X^2$with one degrees of freedom is 3.84

Table 7 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	2.59	1.78	12.54	12.59
RD	4.67	4.88	3.90	3.05	12.54	12.92
DR	21.26	20.97	91.67	92.32	55.25	55.08
$\mathcal{X}^{2*}_{(3)}$		.0828		2.900		.044

An additional and complementary way to assess the ability of a model to generate predictions quantitatively coherent with the 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 dataset and make out of sample prediction on the other half.[48] Second, we use the estimated parameters from the original sample to perform out of sample predictions on the voting behavior of individuals in 1968 and 1972 elections using data from a different source, the General Social Survey (GSS). We report the results of such exercise on the two half samples in tables 8.1,

[48]In 1972 CPS survey respondents are given two types of forms. One part of respondents answers to Form I and the other to From II. There are two types of half samples. The first type of half sample has some respondents receiving Form I and others Form II in the pre-election interview. The second type of half sample has the first half receiving Form I and the second half receiving Form II in the pre-election inteview. We use the second type.

8.2, 9, 10.1, 10.2 and 11; and the results of the same exercise on GSS data in tables 12.1, 12.2 and 13.[49]

Table 8.1 Out of sample prediction on the second half sample:
Aggregate Profiles: Dynamic Profiles 1968-72

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
$\mathcal{X}^{2*}_{(3)}$		8.132

Table 8.2 Out of sample prediction on the second half sample:
Aggregate Profiles: Static Profiles

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
$\mathcal{X}^{2\ **}_{(1)}$		3.064	$\mathcal{X}^{2\ **}_{(1)}$		1.450

Table 9 Out of sample prediction on the second half sample:
Profiles by Party Identification

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

[49] Observations on voting decisions in 1968 and 1972 elections are available from respondents interviewed in 1973. The final sample consists of 639 observations.

Table 10.1 Out of sample prediction on the first half sample:
Aggregate Profiles: dynamic Profiles 1968-72

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 10.2 Out of sample prediction on the first half sample:
Aggregate Profiles: Static Profiles

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 11 Out of sample prediction on the first half sample:
Profiles by Party Identification:

Profiles	Actual	Predicted	Actual	Predicted	Actual	Predicted
	Party D	Party D	Party R	Party R	Party I	Party I
DD	53.48	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

The model performs relatively well on the half samples. It predicts the aggregate outcome of the election in 1968 and 1972 (tables 8.2 and 10.2), and the dynamic voting patterns of Democrats and Republicans (table 9 and 11) on both half samples. The model doesn't seems to perfectly predict the behavior of Independents in the half samples. Indeed, by a small margin, the model doesn't pass the goodness of fit test on the dynamic voting profiles of Independents in the two half samples. The imperfect prediction of the behavior of Independents prevents the model to pass the test on the aggregate dynamic voting profiles (tables 8.1 and 10.1).

Table 12.1 Out of sample prediction on GSS data:
Aggregate profiles: Dynamic Profiles 1968-72

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 12.2 Out of sample prediction on GSS data:
Aggregate Profiles: Static 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 13 Out of sample prediction on GSS data
Profiles by Party Identification

Profiles	Actual	Predicted	Actual	Predicted	Actual %	Predicted
	Party D	Party D	Party R	Party R	Party I	Party I
DD	66.78	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

Our model performs very well on the GSS data. The estimated model perfectly predicts the electoral outcomes in both 1968 and 1972 (table 12.2), and it cannot be rejected on either the aggregate or the group-specific dynamic voting profiles (table 12.1 and 13).[50]

6 Counterfactual Experiments

As pointed out in section 5.3, the 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 that he is a low-valence politician. These outcomes arise from the combination of a conservative constituency, the particular degree of

[50] It doesn't pass the test for Independents just for few decimal points.

asymmetric information about candidates' valence and the trade-off between policies, and valence in voter's 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 which allow us to shed some light on the efficacy of the electoral process to select the "best" candidate and on the so called incumbency advantage.

We determine what is the most liberal policy position that would have allowed a high and low-valence democratic candidate to defeat Nixon in 1968 and in 1972. Then, we 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.[51] 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.[52] Additionally, there were other two candidates, Jackson and Lindsay, that according to our model could have defeated Nixon independently of their valence.[53]

This result is in line with the argument that after 1968 the efficacy of the primary elections system to select the "best" candidate declined dramatically. The McGovern-Fraser Commission introduced a reform of the democratic presidential nominating process in 1972 which made the system more democratic. The commission "recommended" to adopt proportional representation in the allocation of delegates for the various contenders based on their proportion of the popular vote in the primary; in convention and caucus states it required that 75% of delegations be chosen at district level; and it fixed quotas to guarantee the representation of minority groups at the national convention.[54] 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" candidate.[55] Also, simulation studies show that the process of selecting presidential nominees in the U.S. is highly unpredictable.[56] The new system favored the nomination of

[51] Only a high valence democratic candidate could have defeated Nixon in 1968 and he should have had a policy position greater than -.1349. 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.

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

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

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

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

[56] See Merril (1988), and Cooper and Munger (2000).

extremely liberal candidates who didn't necessarily have a remarkable valence.[57] Our result is also in line with the fact that most of the 1972 Nixon's campaign was directed in forcing the strongest democratic candidates out of race in the primary elections.

In the second counterfactual experiment, we compare the realized electoral outcomes in 1968 and 1972 with the ones 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 elections, in 1968 Humphrey would have won.

Similarly, within the asymmetric information environment, we can assess the effect the probability of being informed has on the electoral outcome. This is relevant since parties can affect such probabilities during a campaign. We do not need to find stories about Republicans and Nixon affecting the information during the campaign[58]. It is known that since Nixon appeared on the political scene, he adopted the tactic of "discredit your opponent."[59] Some examples are the series of "dirty tricks" played on Ed Muskie, the democratic front runner in 1972 and his denigrating campaigning against McGovern in 1972.[60] While we cannot find any combination of probabilities that would have helped McGovern win the elections in 1972, we do find different information structures that would have made Humphrey win the election in 1968. In particular, 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 offer some considerations on the so called incumbency advantage. A large literature on incumbency advantage has been developed based on the empirical observation that most of the times incumbent politicians have a higher probability of being elected than new challengers. One possible explanation of 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) politicians.[61] Another explanation has to do with risk averse

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

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

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

[60] "In February of 1972, voters in New Hampshire, site of the first primary, received late night phone calls from people claiming to represent '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....Pheraphs 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).

[61] 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.

voters who prefer to elect a known incumbent than a new, hence "more risky", challenger.[62] Our estimated model indicates that although low quality incumbents can be reelected and they can be reelected with a wider margin than in the election in which they were first appointed, it does not necessarily follow that there is an incumbency 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.[63] In our model a low-valence incumbent has a disadvantage, but (as in 1972) he can be reelected by a large margin. A high valence incumbent has an advantage but he may be defeated.[64] Our finding that incumbents can have an advantage as well as a disadvantage is in line with the results of the model proposed by Fiorina (1981) where voters care about both the expected future policy of elected politicians and their past record. Similarly to our model where the incumbent has a disadvantage (advantage) when he has low (high) valence, in his model the incumbent has a disadvantage 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 which incorporates 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 individual characteristics on voters' policy preferences conditional on their party identification. Our results indicate that to understand the relation between policy preferences and individual characteristics it is important to consider their interaction with party identification.

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

[62] Bernhardt and Ingberman (1985).

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

[64] When we apply our model to the 1976 and 1980 elections we estimated Ford to be low valence and both Carter and Reagan to be high valence. This is an example in which a high valence incumbent, Carter, is defeated.

We estimate that Democrats have a higher probability of being informed than Republicans, which 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 preferences, to provide an equilibrium interpretation of the observed voting patterns and electoral outcomes. In addition, we use the estimated model to conduct counterfactual experiments which allow us to assess the relative importance of candidates' policy positions and valence as well as voters' information on electoral outcomes and to evaluate the performance of the electoral process.

First, we calculate what should have been 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 who participated 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 we find that had all voters been aware that Nixon was a low-valence candidate, Humphrey would have been elected president in 1968. Third, we provide an example in which, although the incumbent President is reelected in office even if he has low valence, 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 for different public offices. In future work we plan to estimate a model of voting in presidential and congressional U.S. elections and address the issue of split-ticket voting (i.e., voting for different parties in multiple simultaneous 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, M. Daniel, and Daniel E. Ingberman. 1985. "Candidate Reputation and the Incumbent Effect." *Journal of Public Economics 27: 47-67.*

[4] Brady Henry E. 1989. "Factor and Ideal point analysis for interpersonal incumparable data". *Psychometrika 54: 181-202.*

[5] 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.*

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

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

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

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

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

[11] Enelow, James M., and Melvin J. Hinch. 1982b. "Nonspatial Candidate Characterisitics and Electoral Competition." *Journal of Politics* 44: 115-130.

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

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

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

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

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

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

[18] 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.*

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

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

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

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

[23] 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.*

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

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

[26] Merril, Samuel III. 1988. "Making Multicandidate Elections More Democratics." *Princeton Univesrity Press.*

[27] Miller, W and J.M Shanks. 1996. "The New American Voter." *Harvard University Press.*

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

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

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

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

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

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

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

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

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

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

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

[39] Persson, Torsten, and Guido Tabellin. 2000. "Political Economics: Explaining Economic Policy." MIT Press.

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

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

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

Appendix A

We let, as we did throughout the whole paper, $z = \lambda(H - L)$. 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 expected vote share conditional on each state, in large samples, is the same as the actual vote share conditional on the same state. Let $C_1 = [n_d(1 - m_d)Y_d(m_1) + n_r(1 - m_r)Y_r(m_1) + n_i(1 - m_i)Y_i(m_1)]$

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 important the incumbent's type) is perfectly revealed by the vote share.

Appendix B

We can classify all possible configurations in four cases:
(1) $m_1 > m_2$ and $g_1 < g_2$, R is the incumbent;
(2) $m_1 < m_2$ and $g_1 > g_2$, R is the incumbent;
(3) $m_1 > m_2$ and $g_1 > g_2$, D is the incumbent;
(4) $m_1 < m_2$ and $g_1 < g_2$, D is the incumbent.[65]

For each of these cases, we analyze what are the states that "potentially" can generate all four voting profiles.[66] We report in tables B.1.-B.4. the cut-off points corresponding to such states for each of the cases listed above. States $HLHL$ and $HLLL$ in case (1), $HLHL$ and $HLHH$ in case (3), and states $HHHH$ and $HLHH$ in case (4) can generate all four profiles when the informed voters generate the profile DR and the uninformed voters generate the profile RD, that is, when $y_{1k} > y_{2k}$ and $y_{10} < y_{20}$. Analogously, states $LHHH$ and $LHLH$ in case (1), $HHHH$ and $LHHH$ in case (2), and states $LHLL$ and $LHLH$

[65] In the first case the democratic candidate in the second period has a policy position more liberal than the democratic candidate in the first period; in the second case the democratic candidate in the second period has a policy position more conservative than the democratic candidate in the first period; in the third case the republican candidate in the second period has a policy position more liberal than the republican candidate in the first period; and in the fourth case the republican candidate in the second period has a policy position more conservative than the republican candidate in the first period.

[66] Remember that the incumbent's type and position are fixed so that for each case there are eight possible states to analyze.

in case (4) can generate all four profiles when the informed voters generate the profile RD and the uninformed voters generate the profile DR, that is, when $y_{1k} < y_{2k}$ and $y_{10} > y_{20}$.

Table B.1. Cut-off points: case (1)

$HLHL$	t=1	t=2	$HLLL$	t=1	t=2
d	$m_1 + \frac{z(1-q)}{g_1}$	$m_2 + \frac{z}{g_2}$	d	$m_1 + \frac{z(1-q)}{g_1}$	m_2
r	$m_1 + \frac{zq}{g_1}$	$m_2 + \frac{zq}{g_2}$	r	$m_1 + \frac{zq}{g_1}$	$m_2 + \frac{zq}{g_2}$
i	$m_1 + \frac{z}{g_1}$	$m_2 + \frac{z}{g_2}$	i	$m_1 + \frac{z}{g_1}$	m_2
I=0	m_1	$m_2 + \frac{zq}{g_2}$	I=0	m_1	$m_2 + \frac{zq}{g_2}$
$LHHH$	t=1	t=2	$LHLH$	t=1	t=2
d	$m_1 - \frac{zq}{g_1}$	m_2	d	$m_1 - \frac{zq}{g_1}$	$m_2 - \frac{z}{g_2}$
r	$m_1 - \frac{z(1-q)}{g_1}$	$m_2 - \frac{z(1-q)}{g_2}$	r	$m_1 - \frac{z(1-q)}{g_1}$	$m_2 - \frac{z(1-q)}{g_2}$
i	$m_1 - \frac{z}{g_1}$	m_2	i	$m_1 - \frac{z}{g_1}$	$m_2 - \frac{z}{g_2}$
I=0	m_1	$m_2 - \frac{z(1-q)}{g_2}$	I=0	m_1	$m_2 - \frac{z(1-q)}{g_2}$

Table B.2. Cut-off points: case (2)

$HHHH$	t=1	t=2	$LHHH$	t=1	t=2
d	$m_1 + \frac{z(1-q)}{g_1}$	m_2	d	$m_1 - \frac{z}{g_1}$	m_2
r	$m_1 - \frac{z(1-q)}{g_1}$	$m_2 - \frac{z(1-q)}{g_2}$	r	$m_1 - \frac{z(1-q)}{g_1}$	$m_2 - \frac{z(1-q)}{g_2}$
i	m_1	m_2	i	$m_1 - \frac{z}{g_1}$	m_2
I=0	m_1	$m_2 - \frac{z(1-q)}{g_2}$	I=0	m_1	$m_2 + \frac{z(1-q)}{g_2}$

Table B.3. Cut-off points: case (3)

$HLHL$	t=1	t=2	$HLHH$	t=1	t=2
d	$m_1 + \frac{z(1-q)}{g_1}$	$m_2 + \frac{z(1-q)}{g_2}$	d	$m_1 + \frac{z(1-q)}{g_1}$	$m_2 + \frac{z(1-q)}{g_2}$
r	$m_1 + \frac{zq}{g_1}$	$m_2 + \frac{z}{g_2}$	r	$m_1 + \frac{zq}{g_1}$	m_2
i	$m_1 + \frac{z}{g_1}$	$m_2 + \frac{z}{g_2}$	i	$m_1 + \frac{z}{g_1}$	m_2
I=0	m_1	$m_2 + \frac{z(1-q)}{g_2}$	I=0	m_1	$m_2 + \frac{z(1-q)}{g_2}$
$LHLL$	t=1	t=2	$LHLH$	t=1	t=2
d	$m_1 - \frac{zq}{g_1}$	$m_2 - \frac{zq}{g_2}$	d	$m_1 - \frac{zq}{g_1}$	$m_2 - \frac{zq}{g_2}$
r	$m_1 - \frac{z(1-q)}{g_1}$	m_2	r	$m_1 - \frac{z(1-q)}{g_1}$	$m_2 - \frac{z}{g_2}$
i	$m_1 - \frac{z}{g_1}$	m_2	i	$m_1 - \frac{z}{g_1}$	$m_2 - \frac{z}{g_2}$
I=0	m_1	$m_2 - \frac{zq}{g_2}$	I=0	m_1	$m_2 - \frac{zq}{g_2}$

Table B.4. Cut-off points: case (4)

$HHHH$	t=1	t=2	$HLHH$	t=1	t=2
d	$m_1 + \frac{z(1-q)}{g_1}$	$m_2 + \frac{z(1-q)}{g_2}$	d	$m_1 + \frac{z(1-q)}{g_1}$	$m_2 + \frac{z(1-q)}{g_2}$
r	$m_1 - \frac{z(1-q)}{g_1}$	m_2	r	$m_1 + \frac{zq}{g_1}$	m_2
i	m_1	m_2	i	$m_1 + \frac{z}{g_1}$	m_2
I=0	m_1	$m_2 + \frac{z(1-q)}{g_2}$	I=0	m_1	$m_2 + \frac{z(1-q)}{g_2}$

We report in table B.5. the probability of each voting profile when the uninformed voters generate the profile RD and the informed voters generate the profile DR . We report in table B.6. the analogous probabilities when the uninformed voters generate the profile RD an the informed voters generate the profile DR.

In order to generate all four voting profiles the cut-off points must satisfy some conditions that guarantee that informed and uninformed voters generate opposite switching behavior and that the probabilities in table B.5 and B.5 are well defined.

It can be shown that for any generic configuration, the states not included in the corresponding table (among tables B.1.- B.4.) are incompatible with observing all four profiles because the conditions that guarantee switching profiles in both directions within each party identification cannot be satisfied by the candidates' configuration or are not compatible with each other.

Table B.5.

$\Pr(DR\|\text{d}) = m_d[Y_d(y_{1d}) - Y_d(y_{2d})]$
$\Pr(RD\|\text{d}) = (1-m_d)[Y_d(y_{20}) - Y_d(y_{10})]$
$\Pr(DD\|\text{d}) = m_d Y_d(y_{2d}) + (1-m_d)Y_d(y_{10})$
$\Pr(RR\|\text{d}) = m_d(1-Y_d(y_{1d})) + (1-m_d)(1-Y_d(y_{20}))$
$\Pr(DR\|\text{r}) = m_r[Y_r(y_{1r}) - Y_r(y_{2r})]$
$\Pr(RD\|\text{r}) = (1-m_r)[Y_r(y_{20}) - Y_r(y_{10})]$
$\Pr(DD\|\text{r}) = m_r Y_r(y_{2r}) + (1-m_r)Y_r(y_{10})$
$\Pr(RR\|\text{r}) = m_r(1-Y_r(y_{1r})) + (1-m_r)(1-Y_r(y_{20}))$
$\Pr(DR\|\text{i}) = m_i[Y_i(y_{1i}) - Y_i(y_{2i})]$
$\Pr(RD\|\text{i}) = (1-m_i)[Y_i(y_{20}) - Y_i(y_{10})]$
$\Pr(DD\|\text{i}) = m_i Y_i(y_{2i}) + (1-m_i)Y_i(y_{10})$
$\Pr(RR\|\text{i}) = m_i(1-Y_i(y_{1i})) + (1-m_i)(1-Y_i(y_{20}))$

Table B.6.

$\Pr(RD\|\text{d}) = m_d[Y_d(y_{2d}) - Y_d(y_{1d})]$
$\Pr(DR\|\text{d}) = (1-m_d)[Y_d(y_{10}) - Y_d(y_{20})]$
$\Pr(DD\|\text{d}) = m_d Y_d(y_{1d}) + (1-m_d)Y_d(y_{20})$
$\Pr(RR\|\text{d}) = m_d(1-Y_d(y_{2d})) + (1-m_d)(1-Y_d(y_{10}))$
$\Pr(RD\|\text{r}) = m_r[Y_r(y_{2r}) - Y_r(y_{1r})]$
$\Pr(DR\|\text{r}) = (1-m_r)[Y_r(y_{10}) - Y_r(y_{20})]$
$\Pr(DD\|\text{r}) = m_r Y_r(y_{1r}) + (1-m_r)Y_r(y_{20})$
$\Pr(RR\|\text{r}) = m_r(1-Y_r(y_{2r})) + (1-m_r)(1-Y_r(y_{10}))$
$\Pr(RD\|\text{i}) = m_i[Y_i(y_{2i}) - Y_i(y_{1i})]$
$\Pr(DR\|\text{i}) = (1-m_i)[Y_i(y_{10}) - Y_i(y_{20})]$
$\Pr(DD\|\text{i}) = m_i Y_i(y_{1d}) + (1-m_i)Y_i(y_{20})$
$\Pr(RR\|\text{i}) = m_i(1-Y_i(y_{2d})) + (1-m_i)(1-Y_i(y_{10}))$

Also, some of the "potential" states imply restrictions on the parameters that may not

be satisfied by some particular configuration. For example in state $LHLH$ of case (1) Democrats cannot generate both vote switching when $qg_2 > g_1$ (as it is in the episodes we consider).

Figure 1. Aggregate distribution of ideal points by race

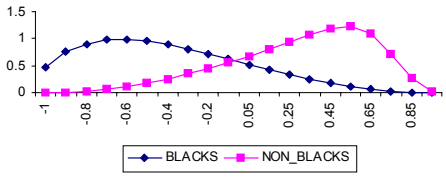

Figure 2. Aggregate distribution of ideal points by education

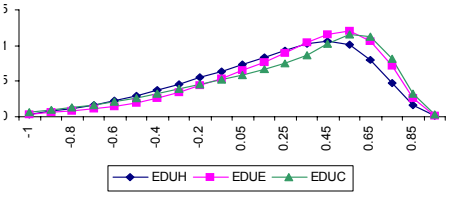

Figure 3. Aggregate distribution of ideal points by region

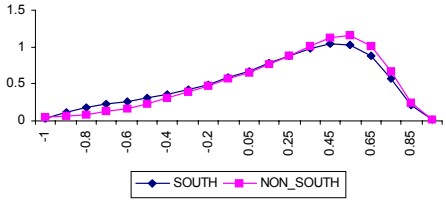

Figure 4. Aggregate distribution of ideal points by gender

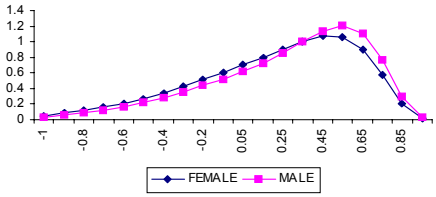

Figure 5. Aggregate distribution of ideal points by income level

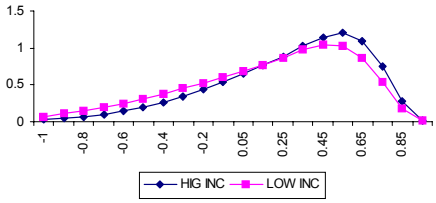

Figure 6. Aggregate distribution of ideal points by age

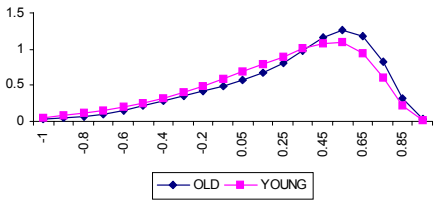

Figure 7. Democrats: distribution of ideal points by race

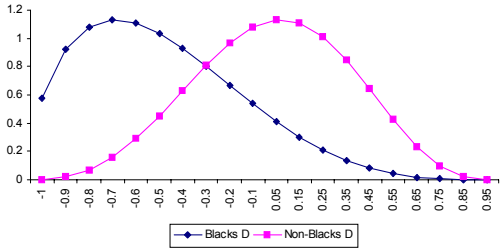

Figure 8. Republicans: distribution of ideal points by race

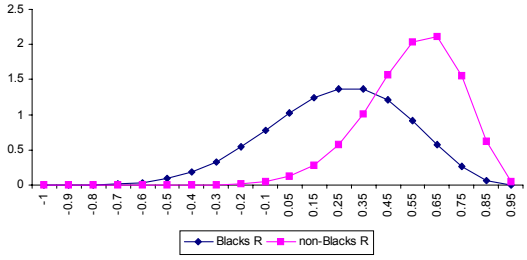

Figure 9. Independents: distribution of ideal points by race

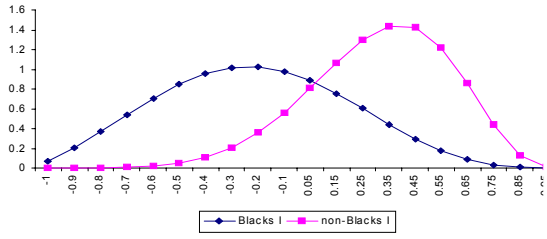

Figure 10. Democrats: distribution of ideal points by education

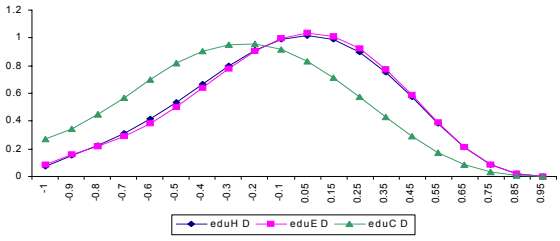

Figure 11. Republicans: distribution of ideal points by education

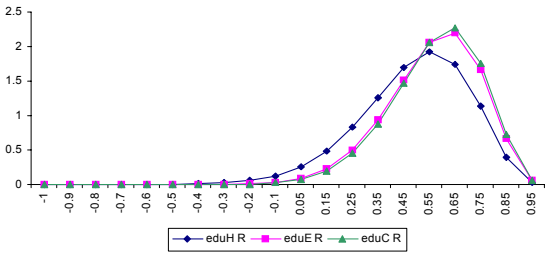

Figure 12. Independents: distribution of ideal points by education

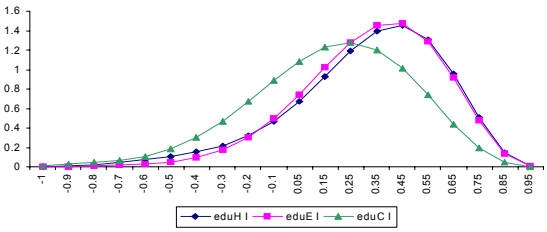

Figure 13. Democrats: distribution of ideal points by region

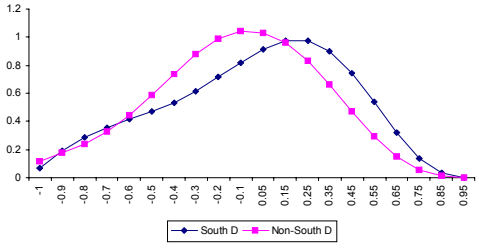

Figure 14. Republicans: distribution of ideal points by region

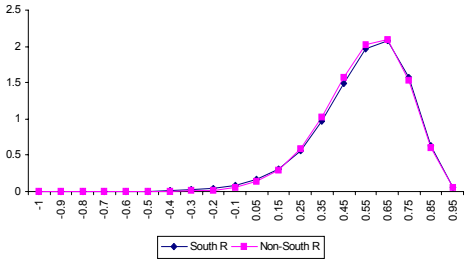

Figure 15. Independents: distribution of ideal points by region

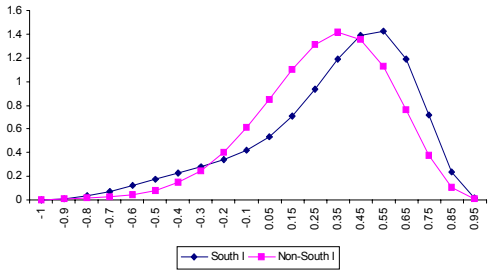

Figure 16. Democrats: distribution of ideal points by gender

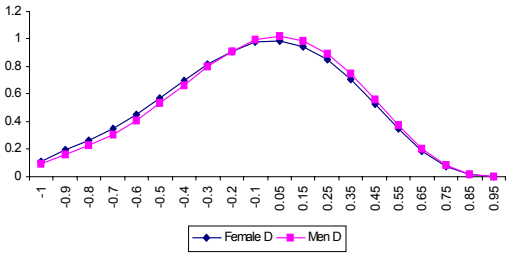

Figure 17. Republicans: distribution of ideal points by gender

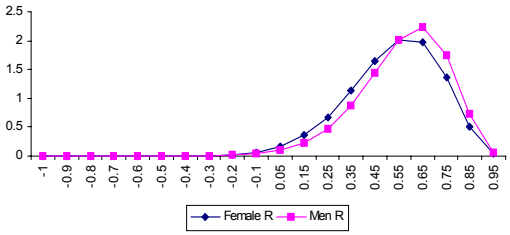

Figure 18. Independents: distribution of ideal points by gender

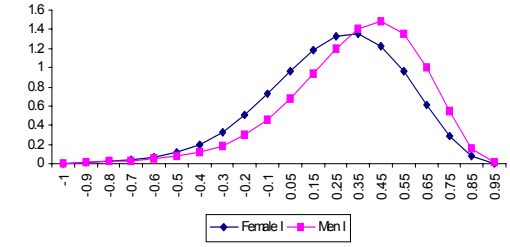

Figure 19. Democrats: distribution of idal points by income level

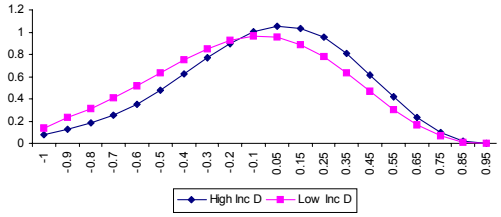

Figure 20. Republicans: distribution of ideal points by income level

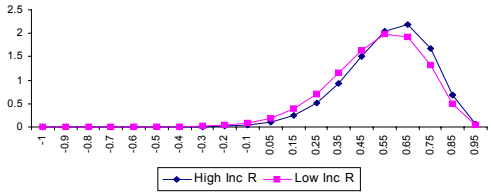

Figure 21. Independents: distribution of ideal points by income level

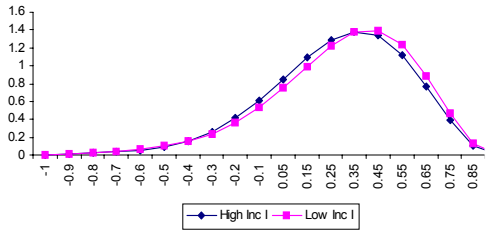

Figure 22. Democrats: distribution of ideal points by age

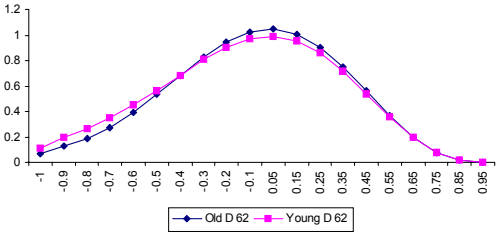

Figure 23. Republicans: distribution of ideal points by age

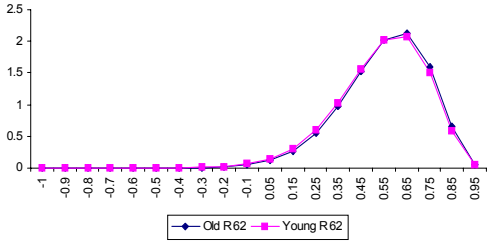

Figure 24. Independents: distribution of ideal points by age

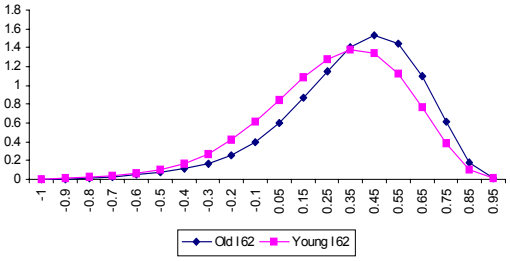

Figure 25: Distribution of ideal points by party identification

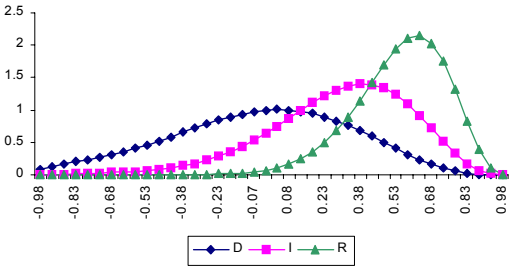

Figure 26: Agregate distribution of ideal points