A Dynamic Model of Voting*

Arianna Degan

University of Pennsylvania

Preliminary version: October, 2002

Abstract

We propose and estimate a dynamic model of voting which incorporates the three main factors affecting voting choices of individual citizens: party identification, policy preferences and candidates' valence. Using individual level data on voting decisions in two consecutive presidential elections we identify and estimate (1) the distribution of voters' policy positions, and (2) candidates' valence. In this respect our analysis differs from existing empirical economic studies of voting which restrict their analysis on single elections and focus on the estimation of policy preferences only.

We estimate the structural model using data on individual voting decisions in the 1968 and 1972 U.S. presidential elections and we use the estimated model to conduct counterfactual experiments to assess the relative importance of candidates' policy positions and valence as well as voters' information on electoral outcomes.

Keywords: party identification, policy preferences, repeated voting, valence.

1   Introduction

In representative democracies elected politicians take policy-relevant decisions on behalf of their constituency. Voters' decisions to support a particular candidate in an election for public office may have important policy consequences. Hence, 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 a relatively higher valence are preferable.

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 vary accross parties.

In this paper we propose and estimate a dynamic model of voting 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. In this respect, our analysis differs from existing empirical economic studies of voting which restrict their analysis on single elections and focus on the estimation of policy preferences only.

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 again for office in the second period and faces a new challenger. There is a continuum of voters who care about both the policy implemented by the winning candidate and his valence. Voters observe candidates' policy positions but not their valence. Voters are heterogeneous with respect to their party identification, information status, policy

---

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

2

preferences and demographic characteristics. We introduce two roles for party identification. First, we assume that there is a relation between voters' party identification and their policy preferences. 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 outcome of an election depends on candidates' policy positions and valence, voters' preferences and the degree of asymmetric information. The model has a unique equilibrium which induces a different voting behavior for each possible combination of candidates' valence in the two elections. The equilibrium implies a probability distribution over voting choices in the two elections conditional on individual characteristics, 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 of 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 the American citizens. For example, 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). Other characteristics however, (e.g., education) have different effects on the policy preferences of voters depending on their party identification. Our results suggest that it is important to break the aggregate relation between individual characteristics and policy preferences by considering the interaction between the effect of individual characteristics, policy preferences and party identification.

Out 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 that he won again the 1972 election by a margin as large as the one of Johnson in 1964 and of Roosevelt in 1936. They are however consistent with anecdotal accounts of the events surrounding the elections we focus on.

In addition to provide 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 and valence as well as voters' information on electoral outcomes.

Before turning our attention to the description of the model, several remarks are in

3

order. First, most of the existing empirical literature on voting bases its estimates of voters' policy preferences on variables containing self-reported information on individual attitudes towards policies and candidates (see, e.g., Cahoon, Hinich and Odershook (1978), Rabinowitz (1978), Poole and Rosenthal (1984), Poole (1998), Enelow and Hinich (1984)). 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 Poole and Rosenthal (1997), Heckman and Snyder(1997), Londregan (1999) and Bailey(2001), to estimate legislators' policy preferences from observed roll-call voting.

Second, while most of the previous empirical analysis of voting focus on single elections (see, e.g., Cahoon,Hinich and Odershook (1978), Rabinowitz (1978), Poole and Rosenthal (1984), Poole (1998), Enelow and Hinich (1984)) we exploit the additional information contained in the sequence of voting choices by the same individuals in two consecutive elections. The variation in the data generated by repeated voting, plus the facts that different individuals face the same candidates in each election and the candidate that wins the first election runs for office also in the second election, allow us to separately identify voters' policy preferences and candidates' valence.

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 and in section 6 we conduct some counterfactual experiments.

2 The Model

There are 2 periods, 1 and 2. In each period there are 2 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 $y_c \in [-1, 1]$,which corresponds to the traditional liberal-conservative dimension,and valence $x_c \in \{L, H\}$, $L < H$, where L and H denote respectively low and high valence. Both y_c 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), reruns for office and faces a new challenger.

There is a continuum of voters with policy preference within the interval $[-1, 1]$. We index each voter by j. Voters observe the candidates' policy positions but they do not observe their valence which is privately known by each candidate. However, voters know the distribution

4

of valence in the population of potential candidates, and we let $q \in (0,1)$ be the probability that a candidate has a high valence.[2]

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 democratic, when a voter has a republican party identification, $k_j = r$, it means that she considers herself republican and when the voter has an independent party identification, $k_j = i$, it means that she does not feel attached to any particular party. 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. The informed voters receive in each period a signal about candidates' valence. 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 be informed also in the second period. Party identification not only affects the probability of becoming informed but also the type of information received. Let s_0^t and s_k^t denote the signal received at time t by an uninformed voter 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 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.[3] For example on his 1981 book Fiorina writes "*All individuals do not receive random samples of political information. One's party identification is no doubt associated with these kind of difference in receipt of information*" (Fiorina 1981,pp81). We model the idea that voters with different party identifications have asymmetric political information by

---

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

[3] Campbell et al.(1960), RePass (1971), Fiorina (1977,1981) , Franklin and Jackon (1983).

assuming that people who feel attached to a party are more likely to be informed about their own party's candidate. We justify this assumption with the fact that typically individuals with a partisan party identification go to their party's conventions, they read partisan newspapers, they are more likely to have friends of the same party with which they talk about their party's candidates. For tractability of the model we assume that party identification is exogenous, while, for identification purposes we assume that it is fixed. These assumptions can be partially justified by the fact that we restrict our analysis to the short-term dynamic.[4]

Party identification is also related to a voter's policy preferences. In particular, within party k, the distribution of ideal points of individuals with characteristics W_j is $Y_k(W_j)$, where W_j belongs to the space of individual characteristics W and Y_k takes value in the interval $[-1, 1]$. We denote the ideal point of individual j, by y_j. The assumption that, for each group k, the distribution of ideal points has full support, is justified by the evidence that voters' self-placement along a liberal-conservative scale is distributed on the full support of such scale irrespectively of voters' party identification.[5]

The utility of the voter depends on the identity of the elected politician. In particular, we assume that a voter's utility for candidate c depends both on the distance between her ideal point and the candidate's policy position and on the candidate's valence,

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

where y_c^t is the policy position of candidate c at time t, and λ is the relative weight that 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 have perfect information on the signal's structure, the distribution of voters across party identification and the group specific distribution of ideal points (that is, they know $n_k, m_k, Y_k(.), \forall k \in K$).

---

[4]Party identification is supposed to capture a long term attachment to a party. 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 (expecially in the years covered by our analysis). Besides, 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".

[5]In the CPS survey respondents are asked to place themself on a 7-points liberal-conservative scale. We find that all the seventh points have positive mass irrespectively of the party identification group. Similar results are encountered in the Gallup's opinion polls.

There is no abstention in the model and voters are assumed to vote sincerely: given their beliefs about candidates' valence, their party identification, their information status, their ideal point and their signal, they vote for the candidate that if elected gives them the higher expected utility.[6] 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 is known before the beginning of period 1. At the beginning of period 1 voters observe the identity of the two competing candidates and their policy positions; some voters become informed and receive a signal about the candidates' valence. During period 1 they vote and at the end of the same period they observe the outcome of the election (the voting share and the identity of the winning candidate).[7] At the beginning of period 2 voters observe the identity of the new challenger and his policy position. They do not directly observe the incumbent's valence but they update their beliefs using the information contained in period 1' s voting share and signal.[8] The informed receive another signal on candidates' valence.[9] During period 2 all voters vote and at the end they observe the electoral outcome.

3   Strategies and Equilibrium

Let S_k be the signal space for group k, and $\Pi_D^1 \subseteq [0,1]$ be the set of period 1's voting 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 $(v_j^1, v_j^2)_k$ which assign to voter j the candidate to vote for in each period. In particular, the voting rule in period 1, $v_j^1 : S_k \times I -> \{D, R\}$, is defined over the possible signals in that period and on the voter's information status; the voting rule in period 2, $v_j^2 : S_k \times I \times \Pi_D^1 -> \{D, R\}$, is defined over the possible signals in both periods, on the voter's information status and on period 1's voting 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.

---

[6]The literature is divided with respect to the assumption of sincere versus strategic voting. For example the assumption of strategic interaction among voters requires a high degree of voters' sophistication and coordination. Besides, sincere voting allow us to avoid to deal with the continuum of equilibria that strategic voting would otherwise generate in a context of a continuum of voters.

[7]The voting share is a sufficient statistic for the electoral outcome.

[8]If, instead of assuming that voters know the distribution of policy preferences but do not observe the winning candidate valence, we assume that voters can observe the winning candidate's valence through his behavior while in office, the equilibrium voting behavior for each state of the world wouldn't be affected.

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

Proposition1: *The unique equilibrium strategy profile with sincere voting* $\{(v_j^1, v_j^2)_k\}_{k=\{d,r,i\}}^*$ *is characterized as follows. For any voter* $j \in [-1, 1]$ *with party identification* $k \in K$ *and with policy position* $y_j \in [-1, 1]$:

$$v_j^1(s_k^1, y_j, I_j)^* = \begin{cases} R & \text{if } Eu_j(x_D^1, x_R^1, y_D^1, y_R^1, y_j | s_k^1, I_j, y_D^1, y_R^1, y_j) > 0 \\ D & \text{if } Eu_j(x_D^1, x_R^1, y_D^1, y_R^1, y_j | s_k^1, I_j, y_D^1, y_R^1, y_j) < 0 \end{cases}$$

and

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

Were the expectation is taken with respect to the beliefs $\{P_{k,I}\}_{(k=\{d,r,i\}, I=\{1,0\})}$ *on the distribution of candidates' types 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 together with 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 although the equilibrium strategy is unique, the individual's actual voting choice depends on the realized information status and on the state of the world, that is 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, and $x = \{x_D^1, x_R^1, x_D^2, x_R^2\}$ 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_j, *ideal point* y_j, *signal* s_k^t, *information status* I_j *and aggregate voting share* π_D^t, *there exists a cut-off point* $\overline{y}(k_j, s_k^t, I_j, \pi_D^t)$ *such that:* *if* $y_j < \overline{y}(k_j, s_k^t, I_j, \pi_D^t)$ *it is optimal to vote* D, *if* $y_j > \overline{y}(k_j, s_k^t, I_j, \pi_D^t)$ *it is optimal to vote* R, *if* $y_j = \overline{y}(k_j, s_k^t, I_j, \pi_D^t)$ *the voter is indifferent and will vote* $D(R)$ *with probability* $\frac{1}{2}$.
(ii) Each cut-off point is of the form: $\overline{y}(k_j, s_k^t, I_j, \pi_D^t) = m_t + \frac{\lambda E_j^t}{g_t}$, *where*
$E_j^t = E(H - L | k_j, s_k^t, I_j, \pi_D^t)$ *is individual* j*'s expected difference in abilities at time* t *(between candidate* R *and* D*);*
$m_t = \frac{y_R^t + y_D^t}{2}$ *is the "midpoint" of the candidates' policy positions ;*
$g_t = 2(y_R^t - y_D^t)$ *is the "gap" between the two candidates' policy positions.*

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

Proposition 2 tells us what are the elements that affect the equilibrium voting strategy, therefore, it can be used to analyze the possible voting profiles that can emerge in equilibrium. A first thing to notice is that only the elements which enter in the expression for the cut-off points are relevant to the equilibrium strategy. In each period t, any cut-off point can be expressed just as a combination of the midpoint between the two candidates' position, m_t, their gap, g_t, and the weighted expected difference in valence λE_j^t. For any information status and signal it is easy to show that λE_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. Voter $j's$ cut-off points will depend on candidates' policy positions, on her information and on candidates' actual valence through the realized signals and the information contained in period $1's$ electoral outcome. Proposition 2 says that, at each period, if a voter's ideal point 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. The locations of a voter's cut-off points in the two periods and their relatioship 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 voting profile, 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_{to} be the cut-off point at time t respectively for an informed voter from group k and for an uninformed voter. 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 $y_{2o} < y_j < y_{1o}$ and RD when $y_{1o} < y_j < y_{2o}$.

One very important 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' position 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

9

a high valence (state $HHHH$) and the incompatibily 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 $y_{10} > y_{20}$, the independents in equilibrium cannot generate RD. The details about which states are compatible with the candidates' configuration (that is their policy positions and the identity of the incumbent) and observing all four voting profiles are contained in appendix B.

Proposition 2 also allows us to derive the following result about the electoral outcome.

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.[10] Voters know n_d, n_r, n_i, m_d, m_r, m_i and Y_d, Y_r, Y_i. 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 voting 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, use in equilibrium 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 the four voting profiles (DD, DR, RD, RR) both at aggregate level and within each group k. The second

---

[10]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.

feature of the data is that the four voting profiles are heterogeneous across groups, that is the frequency of each profile changes by party identifications. Our model must be able to generate equilibrium predictions coherent with such features. In particular, we can show that, within our framework, each element of the model (that is the policy dimension, the valence dimension, asymmetric information and party identification) is necessary for such a purpose, that is, whenever we drop any one of these element 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, the one of voters belonging to the party whose candidate will become incumbent in period 2, that can generate at most two voting profiles (DD and RD or DR and RR).[11] The informed voters of such 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). Let us explain why this is the case. Without loss of generality, consider the problem faced by a voter from group d when D is the incumbent in period 2. The voter, at each period $t \in \{1,2\}$, by voting R receives an expected utility of $qH + (1-q)L$ (remember that the only information about the opposite party's candidate is just the prior distribution of types, independently on the information status,). In the first period, the 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. The informed, on the other hand, when D has a high valence (that is when she receives the signal $s_d^1 = \{H, 0\}$) will vote for D ($H > qH + (1-q)L$) and when D has a low valence (that is when she receives the signal $s_d^1 = \{L, 0\}$) she will vote for R ($H > qH + (1-q)L$). In the second period both the informed and the uninformed voter, knows the incumbent's valence. By voting D, a democratic voter 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.

---

[11]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 belonging to d vote R and the informed voters belonging to r vote D.

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 median voter, where the median voter is the one with median policy position (the median is taken with respect to the aggregate distribution of policy positions). With no utility for valence, party identification looses the role of intermediator of information 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 group. This implies that any kind of switching behavior is unidirectional. If candidate D is the incumbent and candidate R in period 2 has a policy position to the left (right) of candidate R's in the period 2 then, period 2's cut-off point will be smaller (greater) than period 1's and 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. When we eliminate asymmetric information within group, that is 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 accross groups unless people perfectly sort among party identifications on the base of their individuals' characteristics.[12] Party identification, however, is not strictly necessary. It is necessary only within the strict context of our model where the probability of receiving a signal is not correlated with the policy positions but it would not be necessary if we allow otherwise.

---

[12] Although there is some sorting among party identifications based on individual characteristics, such as for example race, sorting on individual characteristics is not extreme enough to justify difference in policy positions.

5   Empirical Analysis

5.1   Data

We focus our attention on the 1968 and 1972 US Presidential elections[13]. In 1968 there are two new candidates running for office: Hubert Humphrey and Richard Nixon. Humphrey, a Senator from Minnesota since 1949 and Vice President during the 1964's term, is the running candidate for the democratic party. Nixon, a Senator form California from 1951 to 1953, Vice-President from 1953 to 1960 and unsuccessful presidential candidate in 1960, is the running candidate for the republican party. In 1968 Nixon wins the elections, he reveive the 43.42% and Humphrey the 42.72% of the popular vote.[14] In 1972 Nixon runs as incumbent candidate 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 with a great margin, 60.69% of popular vote, and then resigns in 1974 after the discovery of the events involved in Watergate. Consistently with our model, the 1968-1972 episode is such that in the first election there are two new contenders.

We use two sources of data. The first is the Center of Political Studies (later National Election Studies) data on the 1972 Presidential Elections. The second is the Poole-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 contextual questions on 1972's voting choice and retrospective questioning on the 1968's.[15] It has two different half samples, each representative of the cross-section in 1972, which can be used to test the performance of the model out of sample. It contains data on individual's socio-demographic characteristics and party identification. We use data on party identification, voting choice, candidates' position, individual characteristics.[16] In all CPS/NES studies respondents are

---

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

[14]In 1968 there is actually a third candidate, George Wallace who receives the remaining 13.53% of the popular vote. However, we focus only on the two major parties' candidates.

[15]Himmelweit, Biberman and Stockdale (1978) analyze the vote bisa related to recall questions; Wright (1993) analyzes the measuring error in vote choice in NES.

[16]Each survey contains a pre-election and a post-election wave. Question on party identification and retrospective voting are asked in the pre-election wave to make the problem of ex-post rationalization less serious.

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.[17] We use data on who the respondents voted for in 1968 and 1972. In particular, we use a dummy variable in each election, R_{68} and R_{72}, which takes the value one if the respondent voted for the republican candidate and zero if the respondent voted for the democratic's.[18] Analogously we use dummy variables, RR, RD, DR, DD, for the respondent two-periods voting profile. We use a dummy variable, BLACK, which takes the value one if the respondent is black and zero otherwise 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. We chose two different indicators of education level: one for education levels strictly lower than high school degree, EDUH, and the other for education levels greater than college degree, EDUC. We use AGE as a continuous scaled variable for the respondent's age and a dummy variable, MINCOME, for income level greater than the median.

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.[19] Table 1 describes the data that we use from the first source.

Table 1 Descriptive statistics

---

[17]There is also a party identification variable which uses a 7-points categorization and takes into consideration 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, second, because our measure is more stable over time and less responsive to short-term shocks.

[18]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.

[19]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.

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*	240	104	92	44
MINCOME	615	222	214	179
TOT	1083	360	428	295

* We reported the number of respondents with age greater or equal than 62.

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 votes to roll calls. 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.

Such measures of candidates' coordinates are particularly appealing for three reason. 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 serve in Congress in different years. Third, the two dimensions are estimated once at a time and the particular structure makes them enter separately in the likelihood so that we can just take the first dimension.[20] Table 2 reports the coordinates on a liberal-conservative scale in the interval $[-1, 1]$ for the

---

[20] Poole K and Rosenthal H, 1997, 1999.

presidential candidates considered in our study.[21]

<div align="center">

Table 2.Presidential candidate's coordinates[22]

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

</div>

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's 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 draw from a beta distribution with support $[-1, 1]$ and parameters (p_{jk}, r_{jk}).[23] We parametrize the first parameter of such distribution to j's characteristics W_j (BLACKS, EDUH, EDUC, SOUTH, FEMALE, AGE, MINCOME) and we further restrict the coefficients r_{jk} to be the same within and across groups. Since $p_{jk} > 0, r_{jk} > 0$ our parametrization becomes

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

where β_k and r are preference parameters to be estimated.[24]

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' valence lead to different equilibrium voting behavior and consequently to different conditions on the parameters that we want to estimate. However, since for any state there is a unique equilibrium voting behavior, we can estimate the parameters of the model by

---

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

[22]Such coordinates were downloadable from K. Poole's home page. I thank K Poole for his collaboration in obtaining such data and for his useful suggestions.

[23]We choose the Beta distribution which is the most flexible distribution and it is defined on a finite support.

[24]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.

maximum likelihood conditional on each state. We then pick as estimate of valence the state whose equilibrium maximum likelihood is the highest. 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:[25]

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

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 candidate's valence x.[26] 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 potentially candidates' valence can be very different. We could fix either λ or $(H - L)$ to 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 the composite parameter z.[27]

For any two consecutive elections where in the first period there are two new challengers and in the second period the incumbent rerun for office there are potentially 8 states to consider, however, the model puts restrictions on which states are compatible with the observed voting profiles and the observed candidates' policy positions. Using the NOMINATE candidates' coordinates for the presidential candidates in 1968 and 1972, the midpoints and gaps in the two periods are such that $m_1 > m_2$ and $g_1 < g_2$. Such configuration of candidates' positions together with the fact that in the data we observe all four voting profiles within each group allow us to exclude the states $HHHH$, $HHLH$, $LLHL$, $LLLL$, $LHLH$ because the derived equilibrium voting profiles are incompatible with the data.[28] We can

---

[25] We use b_{jk} to indicate the density function of 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.

[26] It is not necessary to fix q. However, when we fix it, we improve the precision of the estimates for the other parameters. We fix q to 0.5 both because it corresponds to an uninformative prior (hence it is a neutral choice if we do not have any apriori information) and because when we estimate q we cannot reject the hypotesis that it is equal to 0.5.

[27] If we fix λ and we estimate $(H-L)$, the estimate of $(H-L)$ is very sensitive to the particular parametrization of λ. The same is true the other way around.

[28] In the states $HHHH$ and $HHLH$ and $LHLH$, the independents cannot generate the profile RD. For any value of $z > 0$, both $y_{1i} > y_{2i}$ and $y_{10} > y_{20}$. In the states $LLHL$ and $LLLL$ the democrats cannot

then restrict to three states: $HLLL$, $LHHH$ and $HLHL$.[29] 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 $HLLL$.[30] Table 3 shows the equilibrium cut-off points corresponding to this state.

Table 3 HLLL: cut-off points

group	cut point 1	cut 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:

$$L(V_j|\beta_k, m_k, z, y, W_j, k, HLLL) =$$
$$\int_{-1}^{1}[\Pr(DD|y_u)^{ddj} \cdot \Pr(DR|y_u)^{drj} \cdot \Pr(RD|y_u)^{rdj} \cdot \Pr(RR|y_u)^{rrj}] \cdot b_{jk}(y_u)du$$

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, rr_j, are dummies for individual j's voting profile. Using a different notation we can rewrite the above likelihood as:

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

which using the Beta cdf can be rewritten as :

---

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 . We do not have to worry about the particular case in which the aggregate share when the two candidates in the first period are both a high valence is equal to the aggregate share when the two candidates in the first period have both a low valence, appendix A, because the above inconsistencies remain.

[29] See Appendix B.

[30] 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).

$L(V_j|\beta_k, m_k, z, y, W_j, k, HLLL) =$

$[m_k B_{jk}(y_{2k}) + (1 - m_k)B_{jk}(y_{10})]^{ddj} \cdot [m_k B_{jk}((y_{1k}) - B_{jk}(y_{2k}))]^{drj} \cdot$

$[(1 - m_k)(B_{jk}(y_{20}) - B_{jk}(y_{10}))]^{rdj} \cdot [m_k(1 - B_{jk}(y_{1k})) + (1 - m_k)(1 - B_{jk}(y_{20}))]^{rrj}$

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

$L(V^j|\beta_d, \beta_r, \beta_i, m_d, m_r, m_i, z, y, W_j, HLLL) =$

$\quad [L(V_j|\beta_d, m_d, z, y, W_j, HLLL)]^{pidRj} \cdot$

$\quad [L(V_j|\beta_r, m_r, z, y, W_j, HLLL)]^{pidDj} \cdot$

$\quad [L(V_j|\beta_i, m_i,, z, y, W_j, HLLL)]^{pidIj}$

$l(V_j|\beta_d, \beta_r, \beta_i, m_d, m_r, m_i, z, y, W_j, HLLL) = \sum_j [$

$\quad \ln(L(V_j|\beta_d, m_d, z, y, W_j, d, HLLL)) \cdot pidD_j +$

$\quad \ln(L(V_j|\beta_r, m_r, z, y, W_j, r, HLLL)) \cdot pidR_j +$

$\quad \ln(L(V_j|\beta_i, m_i, z, y, W_j, i, HLLL)) \cdot pidI_j]$

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

5.3   Estimation Results

Table 4 shows the estimates for β_d, β_r, β_i. To interpret the coefficients on individual's characteristics note that a bigger value 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 is the coefficient on any individual characteristic the more conservative are voters with such characterisitic[31]. 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 voters with a higher level of education are more liberal than those with a mid or lower level of education, while, among Republicans the opposite is true, that is, it is the least educated that are most 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 besides the high correlation between income level and education, income level have a separate and opposite effect from education. While higher education

---

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

makes Democrats more liberal, high level of income makes them more conservative. Note that none of the individual characteristics are very significant among Republicans probably due to the small number of observations with profile different from RR.

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

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 by party identification and on aggregate, figure 1-26. All the marginal distributions are on line with the above results. Blacks are more liberal than non-blacks both on aggregate (figure1) and within each party identification (figure 7-9). Southerns are more conservative than non-southerns (figure 3), however the effect is very significant for Democrats followed by Independents while it is not significant at all for Republicans (figure 13-15). This result is consistent with the division that was occurring during those years between southern and non-southen Democrats and with the known fact that southerns were in general more conservative than northerns. Although at aggregate level higher educated voters are more conservative than mid-educated and lower educated voters respectively (figure 2), education has a different effect across party identification (figure 10-12). Even if women are more liberal than men both among Independents and Republicans (figure 16-18), the effect of gender disappears on aggregate (figure 4). As expected from the point estimates, Democrats with an income level lower than the median are more liberal than those with a highest level(figure 19). Age doesn't have a separate effect on policy preferences, however, Independents voters more than 62 years old appear more conservative than the youngest ones. Such effect is probably due to the correlation between age and education. Considering the relative homogeneous demographic composition of different groups, these results seem to imply that there is a strong sorting of individuals among party identifications on the base of unobservables characteristics.[32] If we look at just the aggregate relationships between policy positions and individual characteristics we can arrive to misleading conclusions. Political candidates who choose their platform in order to target particular groups should take such differences into consideration. Overall, Democrats have more heterogeneous preferences on the policy spectrum than Independents, than Republicans (figure 25) whose distribution of ideal points is relatively concentrated to the right. In coherence with the findings of other studies and with the self-reported liberal-conservative view of the population, our estimated agregate distribution of ideal points is relatively conservative.[33]

Most of the above results are on line with what other sociological studies have said about political preferences of American voters.[34] In addition, our results allow to disentangle the effect that each characteristics has on voters' policy preferences and get further insight on

---

[32]Except with respect to race, whose distribution among parties is clearly asymmetric.

[33]Rabinowitz 1978, finds a conservative distribution of policy preferences in 1968. CPS self-reported position on a 7 point liberal-conservative categorization in 1972. Gallup's polls in 1968.

[34]For a discussion, among others, see Scammon and Wattenmberg (1971), Miller and Shanks (1996).

the relation between demographics and political views of the American citizens. It is also very important to stress that we do not use any apriori information on individuals' political preferences to get such results, we rather apply a revealed preference approach in which we use only data on how individuals vote in two consecutive elections and some of their characteristics.[35]

Table 5 shows the estimates for the remaining parameters. The estimated probabilities of receiving information on candidates' valence (m_d, m_r, m_i) indicate that Democrats and Republicans are more likely to be informed than Independents (68% and 29% probability respectively against 18%) 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 a 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 really imprecise. Although we restricted r_k not to be party specific we cannot unambiguously separately identify the constant term in p_{jk} and r. 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. When we jointly estimate p_{jk}, r and z, the estimate of z is not too precise while when we fix r to some value not only the estimate of z becomes very precise but we also eliminate all the correlation between p_{jk} and z. 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 causes 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 (Cntd)

---

[35] In particular we didn't use any of the variables available in the CPS survey which are related to the proximity of individuals' to candidates and parties: feeling termomenters, positions on different issues etc. Such self-reported measures of proximity to candidates are known to suffer from the "projection" and "persuasion" hypothesis and are likely to be non interpersonally comparable.

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 $HLLL$. This means that we estimate Humphrey to have a high valence, McGovern and Nixon to have a low valence. Let us stress that the result on Nixon's valence implies that Nixon has been elected as President in 1968 even if he was a low valence politician, but more interestingly he has been reelected in 1972 even after the voters learned about his 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. As Scammon et al. write in their book, "...the Humphrey campaign finally began 'putting it together, that the Vietnam issue was muted and that the candidate was perceived, finally, as hard-hitting, intelligent, and forceful, as well as a nice guy".[36]

Nixon's success was due mainly to the coupling of his relatively moderate economic policy and his extreme ability to deal with foreign affairs, but, as Watergate revealed, he was not a trustworthy politician. The secret bombing of Cambodia in 1969 is an example of his effort to overexpand the President's war power; impoundment of funds was a way to take the power of the purse from Congress; he made a wide use of veto power, of administrative discretion, of executive orders and secrecy. After the leakage of information in 1971 about the secret bombing of Cambodia, Nixon and Kissinger started a series of illegal wiretapping and they organized a special apparatus "the Plumbers" designated to plug possible leaks but "this clandestine group began to do more than originally charged".[37] His "attack group" lead by Colson was specialized in the collection of information on democratic rivals to be used later to discredit them. His "dirty-tricks team", an old institution in American politics that reached high level in 1964, was taken to new heights in 1972. 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) and the fake cable created in the attempt to

---

[36]Scammon and Wattenberg 1971.

[37]Genovese 1990.

link President J.Kennedy to the assassination of south Vietnam's president. "In addition to the normal and expected work of the President's reelection effort, a subterranean operation of illegal money-collecting and dirty tricks was employed. In an effort to accumulate enough money to run the campaign, the president's money collectors went beyond the bounds of pressuring potential donors to extorting funds".[38] The strategy for Nixon's reelection in 1972 starts with a campaign in the democratic primaries against the strongest democratic candidates in the hope of forcing them out of race so as to face a weak democratic opponent in the general elections. Muskie dropped out of race after the New Hampshire primary, followed by Jackson and Humphrey. The only candidate left was George McGovern, a very liberal candidate without the support of the mainstream Democrats. McGovern's weakness started to appear at the time of the California's primaries where, during a public debate with Humphrey, he showed a lack of a clear and organized political program. "Starting right after his election the public started to question the largest thrust of his campaign: credibility and competence".[39] The O'Brien's case, the Eagelton's case and the Salinger's affair are only few of the many examples of his inability to take clear decisions and to maintain promises.[40]

We can now interpret the observed voting patterns in the light of the results of our model. Taking into consideration the fact that we found Democrats to be relatively conservative, we can say that the informed Democrats who vote Humphrey in 1968 and switch to Nixon in 1972 (20.7%) are voters with a moderate policy view who know that Humphrey has

---

[38]Genovese 1990.

[39]White T (1973).

[40]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'Brein 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 retreived the offer.

Mcgovern chose as his running mate Senator Thomas Eagelton. Only after the decision was taken, his team made investigations on the record of Eagelton and found that he had episodes of serious menthal illness, he had been hospitalized three times and he had received electrical shock twice. After these facts became known, McGovern mantained his support for Eagelton as long as he was forced by the pressure of the press, his party and the public opinion to withdraw him. His 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, the same day at night he informed the press in South Dakota of his intention of reconsidering the nomination.

Pierre Salinger, previously repudiated by McGovern as vice-chariman of the Democratic National Commettee was asked by the same McGovern 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 it was discovered by the press. McGovern publicly denied to have anything to do with it.

a high valence and that McGovern has a low valence. When faced with candidates with equal valence they prefer to vote for the more conservative candidate (Nixon). The vote of informed Republicans and Independents, who both know abouth Nixon's low valence, is driven mainly by policy concerns in both periods. Those who switch vote in the opposite direction (RD) are slightly conservative voters who, given a symmetric prior on candidates' valence, vote Nixon in the first period, driven by policy concerns, and switch to McGovern in the second period, after having observed Nixon's low valence, besides his liberal policy position driven by valence concerns. We can also explain why although Nixon wins in 1972 with a much greater margin than in 1968, the proportion of Republicans who votes Nixon in 1972 is smaller than in 1968. The reason is that most of Republicans, whose distributions of ideal points is relatively conservative, in 1968 are unaware of Nixon's low valence. They vote for him in 1968 on the basis of policy considerations but some of them in 1972, after learning about Nixon's valence, prefer to vote for McGovern besides his extremely liberal policy position.

5.4   Robustness

While in section 3 we discuss 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 the information status, whether a voter receive 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 , m_i because by holding the information status of each voter fixed, informed and uninformed generate opposite switching patterns.

Another assumption related to the information status is that at each point with positive mass there is a fraction m_k that becomes informed.[41] This assumption is sufficient to guarantee that the aggregate voting 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.

None of the assumptions on party identification are necessary for identification. They

---

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

help to explain better the features of the data.[42] 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 a draw from the 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.[43] 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, on the basis of likelihood criteria, our assumption on the party specific signal structure is not rejected by the data.[44] The data support our idea that party identification works like an information selection device which restricts the kind of information that one is willing to absorb.

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. Table 6.1, 6.2 and 7 show the actual and fitted voting profiles on aggregate and by party identification. 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's probability of such profile. We perform standard goodness of fit test 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 4 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

---

[42]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 signal accross parties.

[43]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.

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

aggregate static voting patterns in 1968 and 1972 separately. The model predicts perfectly the two 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 the differences accross parties.

An additional and complementary way to assess the goodness of fit of a model 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.[45] 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 the General Social Survey (GSS).

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

---

[45] In NES studies respondents are usually given two types of forms. In any interview 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. We couldn't estimate the model on the half sample of the first type becase after using our sample selection criteria we were left with an insufficient sample. In the first half all the blacks Independents vote DD. In the second half all southern Republicans vote RR.

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

Tables 8.1, 8.2, 9, 10.1, 10.2 and 11 show the results of such exercise on the two half samples. Tables 12.1, 12.2 and 13 show the results of the same exercise on GSS data.[46] The model performs relatively well on the half samples. It predicts the aggregate outcome of the election in 1968 and 1972 (tables 8.2, 8.4 and tables 10.2, 10.4) and the dynamic voting patterns of Democrats and Republicans (table 9 and 11) on both half samples. However, the model doesn't capture perfectly the behavior of the Independents, consequently, on the half samples, it doesn't pass the goodness of fit test by a very small margin on either the Independents' or the aggregate profiles.

Our model performs very well on the GSS data. Table 12.1 and 13 show that we cannot reject the model on either the aggregate or the group specific dynamic voting profiles.[47] Table 12.2 shows that the model perfectly predicts the electoral outcome in both 1968 and 1972.

---

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

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

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

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

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 Party D	Predicted Party D	Actual Party R	Predicted Party R	Actual % Party I	Predicted 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

6 Counterfactual Experiments

Using our model we can conduct interesting counterfactual experiments. The model tells us that Nixon wins the election in 1968 even though he has low valence and faces a high valence his 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 asymmetric information on the candidates' valence and the trade-off between policies and valence in voter's utility function. Using

our model and the estimated parameters we can determine what is the most liberal policy position that would have allowed a democratic candidate to defeat Nixon. We find that a high valence democratic candidate would have been able to defeat Nixon had he had a policy position greater than -.1349. When we match this with the policy positions of the Democratic candidates in the 1968 Democratic primaries we find that none of the existing candidates held such a policy position.[48] We can conduct a similar experiment in 1972. This time a high valence candidate (as was Humphrey) 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. This means that had McGovern had a high valence he would have lost the election anyway. When we match these thresholds with the policy positions of the democratic candidates in the 1972 presidential primaries we find that Humphrey, Muskie and McCarthy could all have defeated Nixon provided that they had a high valence. Additionally there were other two candidates like Jackson and Lindsay that according to our model could have defeated Nixon independently on their valence.[49] This result is in line with the argument that the system of primary elections is inefficient. It is well known that majority rule in a context with many candidates does not necessarily leads to efficient outcomes.[50] The process of selecting presidential nominees in the US is highly unpredictable. The Reform of the Democratic primaries in 1972 made the system more democratic and extended the representation at the national convention to minority groups. The new system favors the nomination of extremely liberal candidates not necessarily with a remarkable valence. "If more than two parties or candidates are expected, then the vote-maximizing position is not close to your opponents, but well away from them", Tullok 1967.

An additional experiment we perform is to 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. While we cannot find any combination of probabilities that would have helped McGovern win the election in 1972, we do find different information structures that would have made Humphrey win the election in 1968. In particular, had Republicans and

---

[48]The main Democratic candidates in 1968 were Johnson (who resigned), B. Kennedy (who was assassinated during the primaries), McCarthy. Their policy positions are respectively -.235, -.476 and -.369.

[49]The main Democratic candidates in the 1972 primary elections were Humphrey, Muskie, McCharty, Jackson, Sanford, Chisholm and Linsday. The policy positions of the first five are respectively -.34, −.328, .369, -.205, -.255. We do not have the policy position of Chisholm and Linsday on a comparable scale since they served the House of Representative and not the Senate. However looking at their coordinate on the common-space model (Poole 1998) we can claim that Linsday had a very conservative position (he switched party affiliation from Republican) while Chisholm had a policy position even more liberal than McGovern's.

[50]Mueller (1989), Merril Samuel (1988).

Independents been more informed, Humphrey would have been elected as President. The estimated probabilities of being informed imply that Republicans and Independents were more susceptible to this 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. We do not need to find stories about Republicans and Nixon affecting the information during the campaign[51]. It is know that since Nixon appeared on the political scene, he adopted the tactic of "discredit your opponent"[52]. Some examples are the series of "dirty tricks" to Ed Muskie, his front runner in 1972 and his denigrating campaigning against G. McGovern in 1972.[53]

Last, we can use our model to make an assessment about the incumbency advantage. Starting from the empirical observation that most of the time incumbent politicians have a higher probability of being elected than new challengers, a large literature on the incumbency advantage has been developed. One possible explanation of such advantage is that in presence of moral hazard and averse selection an efficient electoral process guarantees that good politicians are elected with a higher probability.[54] Another explanation has to do with risk averse voters who prefer to elect a known incumbent than a new, hence "more risky", challenger.[55] Here we want to offer an opposite point of view: although low quality incumbents can be reelected and they can be reelected with a greater margin than in the election in which they were first appointed, we cannot say 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 bigger margin than what he ac-

---

[51]Genovese M. (1990), White T. (1969,1973).

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

[53]"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". Micheal Genovese (1990).

Famous are also the early morning meeting of the "attack group" with the intent to prepare Nixon's campaign speeches: "statements, attacks and rispostes designed to make McGovern look like a fool". Micheal Genovese (1990).

[54]Banks and Sundaram (1998). Farejhon (1986) has a model with only moral hazard where politicians that exert higher effort have a higher probability of being reelected.

[55]Bernhardt and Ingberman (1985).

tually did.[56] In our model a low valence incumbent has a disadvantage, but as it happened in 1972 he can be reelected with a big margin. A high valence incumbent has an advantage but he may be defeated.[57]

References

[1] Ansolabehere Stephen and Snyder James M. Jr (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, May, 192:-210.*

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

[4] Jones Calvin C and Page Benjamin I.(1975), "Reciprocal effects of Polocy Preferences, Party Loyalties and the Vote", *American Political Science Review, 73,1071-1089.*

[5] Cahoon Lawrence,.Hinich Melvin J. and Ordeshook Peter C. (1978), "A Statistical Multidimensional Scaling Method Based on the Spatial Theory of Voting", *Graphical representation of multivariate data.*

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

[7] Cuckierman Alex (1991), "Asymmetric information and the electoral momentum of public opinion pools", *PublicChoice, 70, 181-213.*

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

[9] Enelow James M, Hinch Melvin (1982), "Ideology, Issues and The Spatial Theory of Elections", *APSR* 1982, vol 76

---

[56] In our sample, the voting share for Nixon in 1972 had he been a new challenger would have been 71.608%, instead of the 66.11% that he got by being an incumbent.

[57] When we apply our model to the 1976 and 1980 elections we find that Ford has a low valence and both Carter and Reagan have a high valence. this is an example in which a high valence incumbent, Carter, is defeated.

[10] Enelow James M, Hinch Melvin (1982), "Nonspatial Candidate Characterisitics and Electoral Competition", *Journal of Politics*, vol. 44, 115-130.

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

[12] Fiorina Morris P. (1977), "An Outline for a Model of Party Choice", *AJPS, 21, 601-626.*

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

[14] Green Donald. Philip and Parmquist Bradley.(1990), "Of Artifacts and Partisan Instability", *AJPS , 34 872-902.*

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

[16] Heckman James J. and Snyder James M. Jr. (1997), "Linear Probability Models of the Demand for Attributes with an Empirical Application to Estimating the Preferences of Legislators", *Rand Journal of Economics.*

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

[18] Jackosn John E. (1975), "Issues, Party Choices, and Presidential Votes*", Americam Journal of Political Science, 161-185.*

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

[20] Lewis Jeffrey B (2001), "Estimating Voters' Preference Distribution from Individual-level Data", *Unpublicshed Manuscript, Princeton University, Department of Politics*

[21] Londregan John and Romer Thomas, "Polarizazion, Incumbency and the Personal Vote", in *Political Economy: Institutions, competition, and Representation,.* Ed by W A Barnett,M hinich and N Schofield. *New york: Cambridge University Press, 1993.*

[22] Londregan John (1999), "Estimating Legislator's Preferred Points", *Political Anyalisis* 8:1, 35-56.

[23] Markus Gregory B and Converse Phillips.E. (1979), "A Dynamic Simultaneous Equation Model of Electoral Choice", *APSR vol .73.*

[24] Merril S.III "Making Multicandidate Elections More Democratics", *Princeton University Press 1988.*

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

[26] Mondak Jeffery.J (1995(., "Competence, Integrity, and the Electoral Success of Congressional Incumbents", *The Journal of Politics, Vol.57, Nov.,1043-69.*

[27] Mueller Dennis C. "Puclic Choice II", *New York, Cambridge University Press1989.*

[28] Page Benjamin and Jones Calvin C (1979), "Reciprocal Effects of Policy Preferences, Party Loyalties and the Vote", *APSR vol. 74, 1071-1089.*

[29] Poole Keith T. (1998), "Recovering a Basic Space from a Set of Issues Scales", *AJPS, vol 43, 954-993.*

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

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

[32] Poole Keith T and Rosenthal Howard (1984), "US Presidential Elections 1968-80: A Spatial Analysis.", *AJPS, 282-312.*

[33] Rabinowitz George (1978)., "On the Nature of Political Issues: Insights from a Spatial Analysis", *AJPS, 22, 793-817.*

[34] RePass David E (1971), "Issue Salience and Party Choice", *APSR, 389-400*

[35] Scammon Rischard.M. and Wattenberg Ben J., "The Real Majority", 1971.

[36] Shachar Ronny, "Party Loyalty as Habit Formation", forthcoming *Journal of Econometrics.*

[37] Stokes Donald (1963) "Spatial Models of Party Competition", *APSR*, 57, 368-377.

[38] Persson Torssenn and Tabellini Guido, "Political Economics: Explaining Economic Policy", 2000, MIT Press.

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

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

[41] Wright Gerald C. (1992),"Errors in Meausring Vote Choice in the National Election Studies 1952-88", *AJPS, vol.37, 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' position in period 1 and proportion of voters in each group, it is possible to calculate the expected voting shares in the first period conditional on the candidates' valence. Because within each party identification we have a continuum of voters and at each point with positive mass a constant fraction of voters becomes informed, the expected voting share conditional on each state, in large samples, is the same as the actual voting 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 apriori between $\pi_D^1(HH)$ 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 that the state (and more important the incumbent's type) is perfectly revealed by the voting share.

Appendix B

We can classify all the possible combinations of candidates' policy positions and identity of the incumbent in four main categories: (1) $m_1 > m_2$ and $g_1 < g_2$, R is the incumbent and the

democratic candidate in the second period has a position more liberal than the democratic candidate's in the first period, $y_D^2 < y_D^1$; (2) $m_1 < m_2$ and $g_1 > g_2$, R is the incumbent and the democratic candidate in the second period has a position more conservative than the democratic candidate's in the first period, $y_D^2 > y_D^1$; (3) $m_1 > m_2$ and $g_1 > g_2$, D is the incumbent and the republican candidate in the second period has a position more liberal than the republican candidate's in the first period, $y_R^2 < y_R^1$; (4) $m_1 < m_2$ and $g_1 < g_2$, D is the incumbent and the republican candidate in the second period has a position more conservative than the republican candidate's in the first period, $y_R^2 > y_R^1$. We analyze for each of this categories what are the states that can generate the 4 voting profiles.[58]

B.1. R is the incumbent: $m_1 > m_2$ and $g_1 < g_2$

The only states that potentially can generate the 4 voting profiles are: $HLHL$, $HLLL$, $LHHH$, $LHLH$.

In the first case, $HLHL$, the cut-off points of informed and uninformed are the following: $y_{1d} = m_1 + \frac{z(1-q)}{g_1}$; $y_{1r} = m_1 + \frac{zq}{g_1}$; $y_{1i} = m_1 + \frac{z}{g_1}$; $y_{10} = m_1$; $y_{2d} = y_{2i} = m_2 + \frac{z}{g_2}$; $y_{20} = y_{2r} = m_2 + \frac{zq}{g_2}$. In the second case , $HLLL$, the cut-off points of informed and uninformed are the following: $y_{1d} = m_1 + \frac{z(1-q)}{g_1}$; $y_{1r} = m_1 + \frac{zq}{b_1}$; $y_{1i} = m_1 + \frac{z}{g_1}$; $y_{10} = m_1$; $y_{2d} = y_{2i} = m_2$; $y_{20} = y_{2r} = m_2 + \frac{zq}{g_2}$. In these first two cases, the informed generate the profile DR and the uninformed the profile RD, that is $y_{1k} > y_{2k}$ and $y_{10} < y_{20}$.

In the third case , $LHHH$, the cut-off points of informed and uninformed are the following: $y_{1d} = m_1 - \frac{zq}{g_1}$; $y_{1r} = m_1 - \frac{z(1-q)}{g_1}$; $y_{1i} = m_1 - \frac{z}{g_1}$; $y_{10} = m_1$; $y_{2d} = y_{2i} = m_2$; $y_{20} = y_{2r} = m_2 - \frac{z(1-q)}{g_2}$. In the fourth case , $LHLH$, the cut-off points of informed and uninformed are the following: $y_{1d} = m_1 - \frac{zq}{g_1}$; $= y_{1r} = m_1 - \frac{z(1-q)}{g_1}$; $y_{1i} = m_1 - \frac{z}{g_1}$; $y_{10} = m_1$; $y_{2d} = y_{2i} = m_2 - \frac{z}{g_2}$; $y_{20} = y_{2r} = m_2 - \frac{z(1-q)}{g_2}$. In these last two cases, the informed generate the profile RD and the uninformed the profile DR, that is $y_{1k} < y_{2k}$ and $y_{10} > y_{20}$.

B.2. R is the incumbent: $m_1 < m_2$ and $g_1 > g_2$

The only states that can generate all the profiles are $HHHH$ and $LHHH$. In the first case, $HHHH$, the cut-off points of informed and uninformed are the following: $y_{1d} = m_1 + \frac{z(1-q)}{g_1}$; $y_{1r} = m_1 - \frac{z(1-q)}{g_1}$; $y_{1i} = y_{10} = m_1$; $y_{2d} = y_{2i} = m_2$; $y_{20} = y_{2r} = m_2 - \frac{z(1-q)}{g_2}$. In the second case, $LHHH$, the cut-off points are: $y_{1d} = m_1 - \frac{zq}{g_1}$; $y_{1r} = m_1 - \frac{z(1-q)}{g_1}$; $y_{1i} = m_1 - \frac{z}{g_1}$; $y_{10} = m_1$; $y_{2d} = y_{2i} = m_2$; $y_{20} = y_{2r} = m_2 - \frac{z(1-q)}{g_2}$. In both cases, if we want to generate all the profiles, the informed must generate the profile RD and the uninformed the profile DR, that is $y_{1k} < y_{2k}$ and $y_{10} > y_{20}$.

---

[58]Remember that the incumbent's type and position are fixed so that for each case there are 8 possible states to analyze.

B.3. D is the incumbent: $m_1 < m_2$ and $g_1 < g_2$

The only states that potentially can generate the 4 voting profiles are: $HLHL$, $HLHH$, $LHLL$, $LHLH$. In the first case, $HLHL$, the cut-off points of informed and uninformed are the following: $y_{1d} = m_1 + \frac{z(1-q)}{g_1}$; $y_{1r} = m_1 + \frac{zq}{g_1}$; $y_{1i} = m_1 + \frac{z}{g_1}$; $y_{10} = m_1$; $y_{2r} = y_{2i} = m_2 + \frac{z}{g_2}$; $y_{20} = y_{2d} = m_2 + \frac{z(1-q)}{g_2}$. In the second case, $HLHH$, the cut-off points of informed and uninformed are the following: $y_{1d} = m_1 + \frac{z(1-q)}{g_1}$; $y_{1r} = m_1 + \frac{zq}{g_1}$; $y_{1i} = m_1 + \frac{z}{g_1}$; $y_{10} = m_1$; $y_{2r} = y_{2i} = m_2$; $y_{20} = y_{2d} = m_2 + \frac{z(1-q)}{g_2}$. In these first two cases, the informed must generate the profile DR and the uninformed the profile RD, that is $y_{1k} > y_{2k}$ and $y_{10} < y_{20}$.

In the third case, $LHLL$, the cut-off points of informed and uninformed are the following: $y_{1d} = m_1 - \frac{zq}{g_1}$; $y_{1r} = m_1 - \frac{z(1-q)}{g_1}$; $y_{1i} = m_1 - \frac{z}{g_1}$; $y_{10} = m_1$; $y_{2r} = y_{2i} = m_2$; $y_{20} = y_{2d} = m_2 - \frac{zq}{g_2}$. In the fourth case, $LHLH$, the cut-off points of informed and uninformed are the following: $y_{1d} = m_1 - \frac{zq}{g_1}$; $y_{1r} = m_1 - \frac{z(1-q)}{g_1}$; $y_{1i} = m_1 - \frac{z}{g_1}$; $y_{10} = m_1$; $y_{2r} = y_{2i} = m_2 - \frac{z}{g_2}$; $y_{20} = y_{2d} = m_2 - \frac{zq}{g_2}$. In these last two cases, the informed must generate the profile RD and the uninformed the profile DR, that is $y_{1k} < y_{2k}$ and $y_{10} > y_{20}$.

B.4. D is the incumbent: $m_1 > m_2$ and $g_1 > g_2$

The only states that can generate all the profile are $HHHH$ and $HLHH$. In the first case, $HHHH$, the cut-off points of informed and uninformed are the following: $y_{1d} = m_1 + \frac{(1-q)z}{g_1}$; $y_{1r} = m_1 - \frac{z(1-q)}{g_1}$ $y_{1i} = m_1$; $y_{10} = m_1$; $y_{2d} = y_{20} = m_2 + \frac{z(1-q)}{g_2}$; $y_{2i} = y_{2r} = m_2$. In the second case, $HLHH$, the cut-off points of informed and uninformed are the following: $y_{1d} = m_1 + \frac{(1-q)z}{g_1}$; $y_{1r} = m_1 + \frac{zq}{g_1}$ $y_{1i} = m_1 + \frac{z}{b_1}$; ; $y_{10} = m_1$; $y_{2d} = y_{20} = m_2 + \frac{z(1-q)}{g_1}$; $y_{2i} = y_{2r} = m_2$. In both cases, the informed must generate the profile DR and the uninformed the profile RD. We need that $y_{1k} > y_{2k}$ and $y_{10} < y_{20}$.

Notice that when the model generates all the four voting profiles, it must be the case that informed and the uninformed generate opposite switching behaviors. Table B.1 shows the probability of the voting profiles when the uninformed generate the profile RD and the informed the profile DR. Table B.2 shows the probability of the voting profiles when the uninformed generate the profile RD an the informed the profile DR.

Table B.1

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

Table B.2

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

The conditions that must be satisfied by the parameters in order to generate the four possible profiles can be derived from the conditions on the cut-off-off points. Such conditions, in the cases represented in table B.1 are: $y_{1k} > -1$ and $y_{2k} < 1$ (conditions on the profile RD); $y_{20} < 1$ and $y_{10} > -1$ (conditioncs on the profile DR); $max\{y_{2k}, y_{10}\} > -1$ (condition on the profile DD); $min\{y_{2k}, y_{10}\} < 1$ (condition on the profile RR). Equivalently, in the cases represented in table B.2: $y_{1k} > -1$ and $y_{2k} < 1$; $y_{20} < 1$ and $y_{10} > -1$; $max\{y_{2k}, y_{10}\} > -1$; $min\{y_{2k}, y_{10}\} < 1$.

Figure 1. Aggregate distribution of ideal points
by race
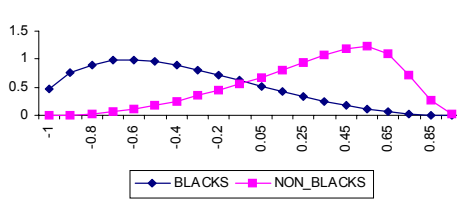
Figure 2. Aggregate distribution of ideal points
by education
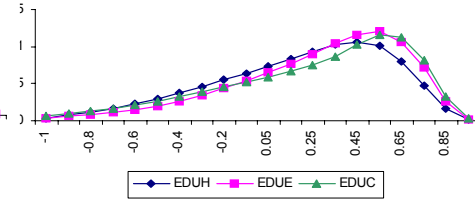
Figure 3. Aggregate distribution of ideal points
by region
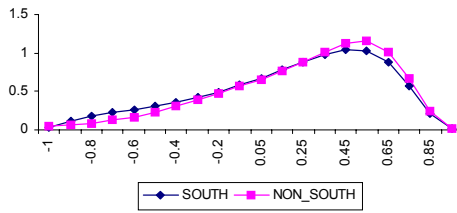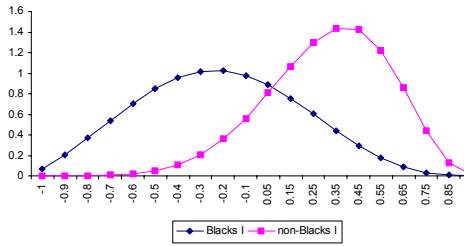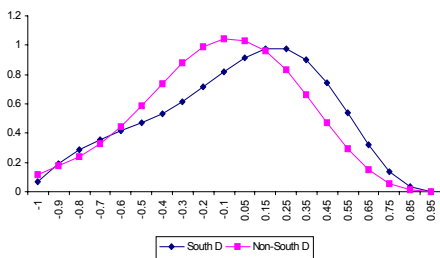

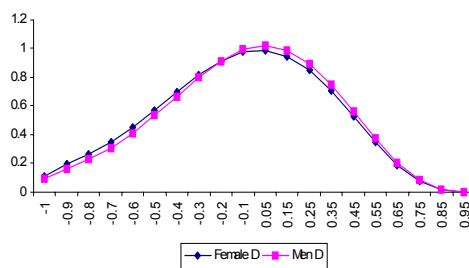
Figure 4. Aggregate distribution of ideal points
by gender
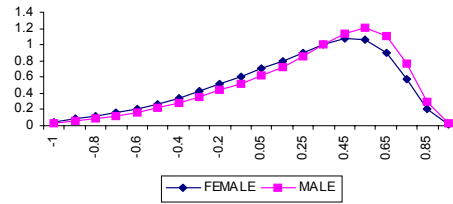
Figure 5. Aggregate distribution of ideal points
by income level
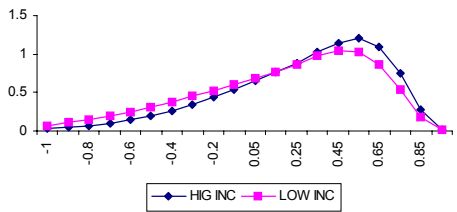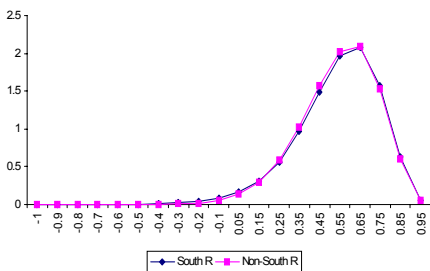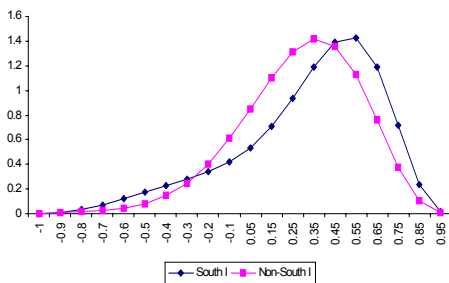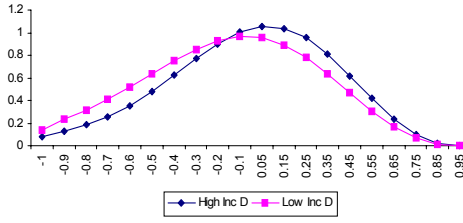

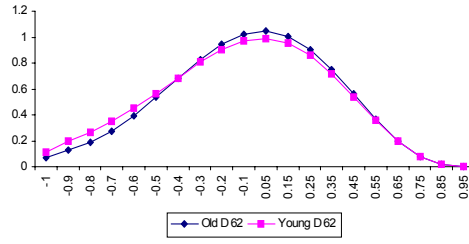Figure 6. Aggregate distribution of ideal points
by age

42
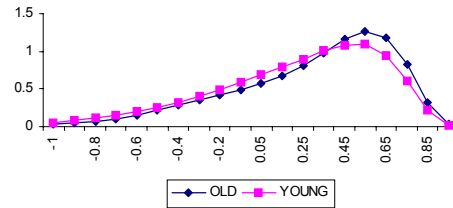
Figure 7. Democrats: distribution of ideal points by race
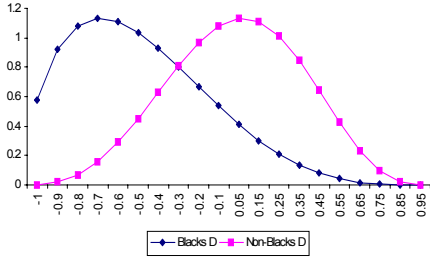
Figure 10. Democrats: distribution of ideal points by education
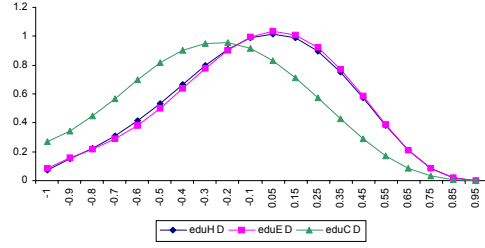
Figure 8. Republicans: distribution of ideal points by race
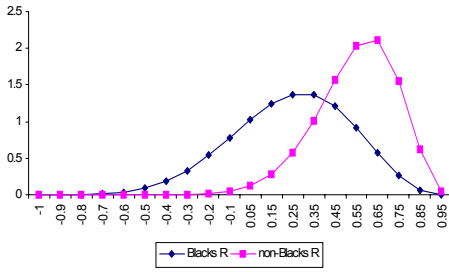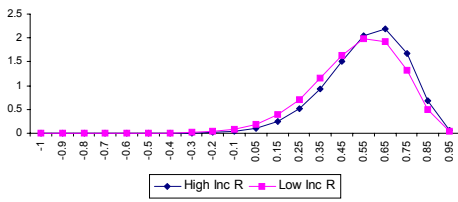

Figure 11. Republicans: distribution of ideal points by education
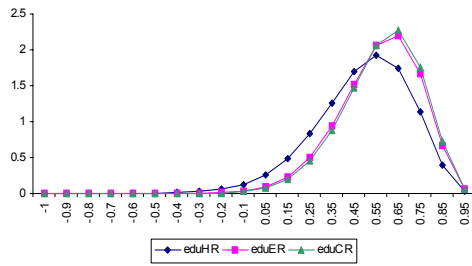
Figure 9. Independents: distribution of ideal points by race

Figure 12. Independents: distribution of ideal points by education

43
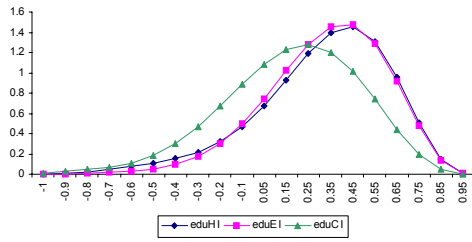
Figure 13. Democrats: distribution of ideal points by region

Figure 16. Democrats: distribution of ideal points by gender

Figure 14. Republicans: distribution of ideal points by region
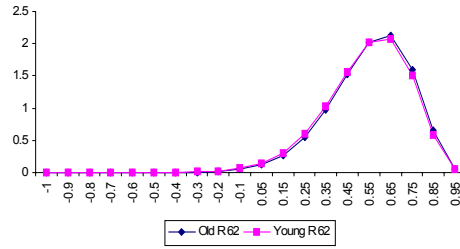
Figure 17. Republicans: distribution of ideal points by gender
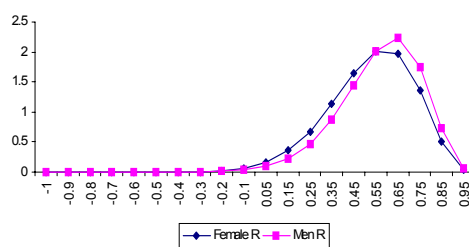
Figure 15. Independents: distribution of ideal points by region

Figure 18. Independents: distribution of ideal points by gender
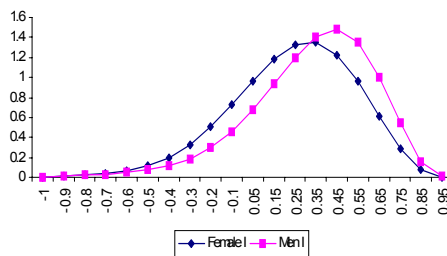
Figure 19. Democrats: distribution of idal points by income level

Figure 22. Democrats: distribution of ideal points by age

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

Figure 23. Republicans: distribution of ideal points by age

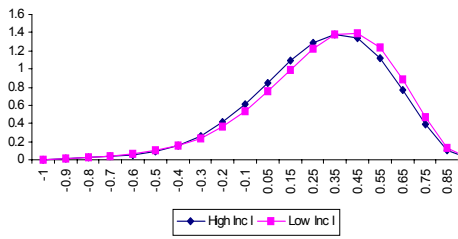Figure 21. Independents: distribution of ideal points by income level

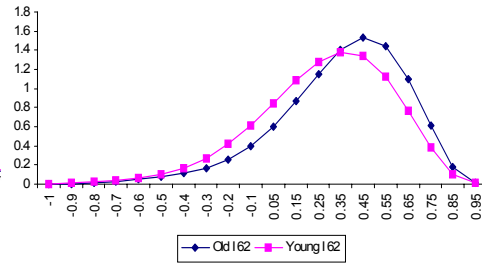Figure 24. Independents: distribution of ideal points by age
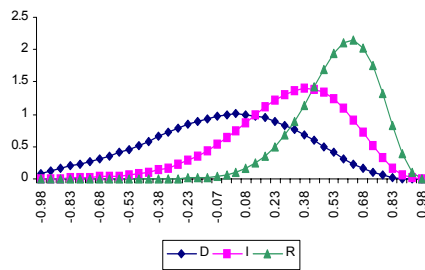
Figure 25: Distribution of ideal points by party identification

Figure 26: Agregate distribution of ideal points