Turnout and Learning in Sequential Election: The Case of U.S. Presidential Primaries*

Kei Kawai[†]
Northwestern University

Yasutora Watanabe[‡]
Northwestern University

January 2010
(Very Preliminary)

Abstract

Voter turnout has been one of the most fundamental research question in the study of political economy. Most study has focused on turnout in simultaneous election in which pivot probabilities are same for all voters. As pivot probabilities are the same in a simultaneous election, the only way to investigate empirically how pivot probabilities affect turnout decision is to compare outcomes of different elections controlling for other factors.

In sequential election, to the contrary, pivot probabilities evolve over time within an election and voters socially learn from preceding votes. We propose a dynamic strategic model of turnout and social learning, and estimate the model using the data of U.S. Presidential Primaries. We find a large effect of social learning and quantify its effects on turnout. Finally, we conduct counterfactual experiment on alternative voting schemes.

*This work is very preliminary. Comments are welcome.

[†]Department of Economics, Northwestern University, 2003 Sheridan Road, Evanston, IL 60208. Email: k-kawai@northwestern.edu

[‡]Department of Management and Strategy, Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208. Email: y-watanabe@kellogg.northwestern.edu

1 Introduction

Why people vote in large elections has been one of the most fundamental research questions in the political economy literature. Many papers are written to explain voter turnout and abstention both theoretically and empirically. The question is a natural one given that the probability of changing an election outcome is very small for each voter while the cost of voting can be significant.

Most studies in the literature considers simultaneous elections in which the pivot probability is the same across all voters. In a simultaneous election, whether a voter voted in a particular voting booth at a particular timing does not affect the pivot probability of the vote because the votes are counted only after all votes are cast. Thus, the only way to investigate empirically how pivot probabilities affect turnout decision is to compare outcomes of different elections controlling for other factors.

We depart from the existing literature by investigating voter turnout in a sequential election. In a sequential election, voters cast votes after observing the decision of preceding voters. Hence, pivot probabilities evolve over time within one election depending on the difference in vote counts across candidates at the time of voting. Furthermore, if voters do not have perfect information about candidates, they can learn from the outcomes of preceding votes. This social learning aspect also contributes to the evolution of pivot probabilities within a election in addition to the simple arithmetic difference in cumulative votes at the time of voting. As a result of evolving pivot probability, voter turnout could vary across the timing of voting. In fact, Table 1 reports a significant correlation of timing of voting on "voter turnout" from a simple regression. The results shows that voter turnout decreases significantly over time.

In order to accommodate these features of sequential election, we propose a dynamic strategic model of voter turnout and social learning in sequential election. We then estimate the model using the county-level data of U.S. Presidential Primaries, and conduct counterfactual policy experiment with alternative electoral system.We propose a dynamic model of sequential voting in U.S. Presidential Primaries. We depart from a typical model by incorporating the institutional feature that there is a sequence of large simultaneous voting (which corresponds to primaries at each state)

[1] Controls include demographic characteristics, presence of other elections, and formats of primaries.

Turnout Regression	(1)	(2)	(3)	(4)
Timing	-0.075	-0.064	-0.318	-0.253
	(0.004)	(0.005)	(0.014)	(0.017)
$(\text{Timing})^2$	–	–	0.018	0.013
			(0.001)	(0.001)
Controls	–	Included	–	Included
R^2	0.211	0.363	0.374	0.422
# of Obs	1,269	1,269	1,269	1,269

Table 1: Regression of "turnout" (# of Votes / # of Registered Democrats) onto Constant, Timing, $(\text{Timing})^2$, and Controls.[1] Unit of observation is municipalities. Timing is a number between 1 (New Hampshire) and 14 (New Jersey, Montana), which we define as the order of voting.

instead of the typical theoretical formulation that each single vote is cast sequentially. We assume a group-utilitarian model of voter turnout similar to Coate and Conlin (2004) and Feddersen and Sandroni (2006) in which voters not only maximize their own self-interest but also makes ethical considerations. In the model, quality of candidates are not perfectly know, and voters accumulate information on candidate quality from their own signal as well as from the history of preceding vote counts. Because subsequent voters infers quality of candidates from preceding votes, early voters strategically considers the effect of their votes on the subsequent voters.

We estimate the model in two steps using county-level data of 2004 Presidential Primary for the Democratic Party. First, we recover the cost of voting, voting threshold, and perceived quality of candidates using the data of election outcome and demographic characteristics. These objects are identified by relating the variation of election outcomes and the variation of county-level demographic characteristics. Second, we use the perceived quality of candidates estimated in the first stage to recover the distribution of signal for each candidates. The distribution is identified as the parpametrization of the distribution allows us to compute the evolution of perceived quality over time.

We find a large effect of social learning. One of the candidate who could not win the election (Howard Dean) on average has a better quality than the candidates who actually won the election (John Kelly). In terms of our model, we can interpret this result as negative shocks to Howard Dean on earlier stages had persistent effect over the race due to a large effect of social learning.

We discuss the related literature below, and describes our model in Section 2. Section 3 explains our data, and Section 4 provides details on our estimation. We discuss our results of estimation and of our counterfactual experiments in Section 5. We conclude in Section 6.

1.1 Related Literature

This paper adds to two strands of literature; the literature on voter turnout and the literature on sequential election. In the voter turnout literature, pivotal voter model As voter turnout literature is vast, we only discuss the papers that are directly related to our paper (See, e.g., Dhillon and Peralta, 2002, Feddersen, 2004, for surveys of the literature.). The paper closest to ours in the turnout literature is that of Coate and Conlin (2004, Hereafter denoted as CC).

Literature on sequential election:

To the best of our knowledge, Battaglini (2005) is the only paper that considers turnout in sequential election.

2 Model

We consider a model of sequential election in which voter makes two decisions; whether to vote or not and whom to vote if they do. We denote the timing of voting by $t \in \{1, ..., T\}$, $T < \infty$, and the candidates by $k \in \{1, ..., K\}$. In regular models of sequential voting, one voter cast a vote in each period. Our model differs from regular models in the sense that all voters in a State cast votes simultaneously and timing of elections differ across states. For simplicity of notation, we assume voters in one State is voting at time t (and hence we also use t to index the State voting at date t). We allow more than one states to vote on the same date in the empirical section of the paper. A State has a continuum of eligible voters with mass $M_t \in \mathbb{R}_+$.

Voter i in State t receives a noisy signal $s_{it} \in \mathbb{R}^K$ about the quality of each candidate, where each element of vector $s_{it} = [s_{it1}, ..., s_{itK}]$ is a random variable that can be decomposed to three components as

$$s_{itk} = \chi_k + \eta_{tk} + \xi_{ik}$$

where χ_k denotes the common-value component of the quality of candidate k, η_{tk} is a State specific noise on the signal that follows a Normal distribution with mean zero and precision ρ_η, $N(0, 1/\rho_\eta)$, and ξ_{ik} is an individual voter level noise on the quality of the candidate. Voter i observes signal s_{itk}, but not each of its components, χ_k, η_{tk}, and ξ_{ik}. The voter only observes his own signal, and not the signal of others.

Voter i receives utility u_{ik} from candidate k winning the election, which we write as

$$u_{ik} = \chi_k + \upsilon_k(X_i) + \varepsilon_{ik}$$

where $\upsilon_k(X_i)$ is a component of the utility which is a function of voter i's demographic characteristics X_i, and ε_{ik} is a voter-level idiosyncratic preference shock. We assume that ε_{ik} follows a type-I extremum value distribution. The function υ_k is known to all voters. At the time of voting, voter i does not know the realization χ_k. Thus the expected utility conditional on the information the voter has at the time of voting is,

$$E[u_{ik}|\Omega_i] = E[\chi_k|\Omega_i] + \upsilon_k(X_i) + \varepsilon_{ik}$$

where Ω_i is the information set for voter i. The information set Ω_i includes voter i's signal, the voting history leading up to State t. We will describe below what belongs in Ω_i more precisely as well as how voters form and update beliefs.

Voter Turnout Our model of turnout is an adaptation of Coate and Conlin (2004) who consider groups of voters who receive warm glow payoff from "doing their part" within the group in the spirit of Harsanyi (1980) and Feddersen and Sandroni (2006). We consider a *group* to be a collection of voters whose expected utility from a particular candidate is the highest in each State. We can write the size of group k in State t as

$$\begin{aligned} M_{tk} &= M_t \iint \mathbf{1}\left\{E[u_{ik}(X_i)|\Omega_i] \geq E[u_{il}(X_i)|\Omega_i]\ \forall l\right\} dF_\varepsilon dF_X \\ &= M_t \int M_k(X) dF_X \end{aligned} \quad (1)$$

where F_X denotes the distribution of individual demographic characteristics X_i in State t.

Voters choose whether to vote or not, and if they do, they incur a cost of voting c_i drawn from the cost distribution denoted by $F_c(\cdot|X_i)$ which has non-negative support. Conditional on voting, we assume that voter from group k votes sincerely for candidate k. Moreover, as in Coate and Conlin (2004), we assume that voters decide to vote or abstain based on the rule which maximizes the aggregate expected utility of the group if all voters in the group act according to the rule. Then it is without loss of generality to consider a cutoff rule in the cost of voting whereby voters with cost above a certain threshold, λ, vote and below which abstain. This cutoff λ is chosen as a function of h_t, the voting history leading upt to State t, so that it maximizes the aggregate utility of the group.

The aggregate expected utility of the group is the sum of two parts, the first of which comes from having the preferred candidate win, and the second of which is the sum of the cost of voting. Let h_t denote the history of voting up to State t, and p_k be the probability candidate k will win, given history h_t and group tk's choice of λ_{tk} and the choices of all other groups, $(\lambda_{t',-k})_{t'\geq t}$. Then we can express the first term of the aggregate expected utility as

$$U(h_t : \lambda_{tk}, (\lambda_{t',-k})_{t'\geq t}) = \sum_k p_k(h_t : \lambda_{tk}, (\lambda_{t',-k})_{t'\geq t}) \left(E[\chi_k|\Omega_i,\, k \text{ wins}] + \int \upsilon_k(X_i) dF_X \right).$$

The second term is the aggregate sum of voting costs,

$$C(\lambda_{tk}) = -\int\int^{\lambda_{tk}} M_k(X)\ c\ dF_{c|X} dF_X.$$

Voters in each group chooses the threshold in their cost, λ, in such a way to maximize $U(h_t : \lambda_{tk}, (\lambda_{t',-k})_{t'\geq t}) - C(\lambda_{tk})$. If the voters in group tk choose the threshold λ_{tk}, voters in group tk vote for candidate k if his cost is below λ_{tk} and abstains if his cost is above the threshold.

Voters' Information We now specify the voters' information set Ω_i. We assume that h_t, and $M_{t'k}$ ($t' \leq t$) are in Ω_i. That is, voters form beliefs based on the voting history in previous States as well as the size of the supporters $M_{t'k}$. Note that the voter observes the size of the supporters in his own State M_{tk} as well. This is implicit in the turnout rule where the voter must compute the aggregate utility of the group. (Alternatively, we can assume that the voters know the cost distribution $F_{|X}$ to have the same information environment - following the same argument as our identification.)

Equilibrium Now, we define the equilibrium of the voting game. The equilibrium is defined by a profile of strategies $\{\lambda_{tk}\}$, sizes of groups $\{M_{tk}\}$, and the beliefs of the voters regarding candidate quality χ. As described above, for all t and k, λ_{tk} maximizes aggregate group utility given $\lambda_{t',-k}$ and voting history h_t, M_{tk} is determined by equation (1) and belief updating follows Bayes' Rule. We can show that an equilibrium of the model always exists, but we cannot show uniqueness of the equilibrium.

Note that the size of groups $\{M_{tk}\}$ is part of the equilibrium. This is because the voters in State t choose to join one of the groups depending on the information they have at time t. As the information set Ω_i itself contains M_{tk}, Ω_i and M_{tk} must jointly be determined: That is, once voters decide on which group to join, they observe the size of M_{tk}, and the voters should not have any incentive to change groups after observing this. Moreover, regarding the information they have at the time of joining groups, Ω_i, recall that the signal of the voter is $s_{itk} = \chi_k + \eta_{tk} + \xi_{ik}$. Because ξ_{ik} is independently drawn for each individual, knowledge of M_{tk} reveals the components of the signal that is common to the State, i.e. $\chi_k + \eta_{tk}$ $(\equiv s_{tk})$. Thus $M_{tk} \in \Omega_i$ means that we can assume as if the voters knew the individual specific component of the signal ξ_{ik}.

Belief Updating We describe the updating of voters' beliefs for the State voting in time t. At $t = 1$, only information available for voter i is signal s_{itk}. He uses this signal $s_{itk}$to update his prior, which follows Normal distribution with mean $\mu_{\chi k}$ with precision $\rho_{\chi k}$, $N(\mu_{\chi k}, 1/\rho_{\chi k})$. Then the posterior distribution B_{1k} is written as

$$B_{1k} \sim N\left(\frac{\rho_{\chi_k}\mu_{\chi k} + \rho_\eta s_{1k}}{\rho_{\chi_k} + \rho_\eta}, \frac{1}{\rho_{\chi_k} + \rho_\eta}\right).$$

Using this posterior, supporter population for candidate k is written as

$$M_{1k} = M_1 \iint \mathbf{1}\left\{E[u_{ik}(X_i)|\Omega_i] \geq E[u_{il}(X_i)|\Omega_i]\ \forall l\right\} dF_\varepsilon dF_X$$

Noting that $E[u_{ik}(X_i)|\Omega_i] = \frac{\rho_{\chi_k}\mu_{\chi k}+\rho_\eta s_{1k}}{\rho_{\chi_k}+\rho_\eta} + \upsilon_k(X_i) + \varepsilon_{ik}$ and that ε_{ik} follows type-I extremum distribution, the above expression simplifies to

$$M_{1k} = M_1 \int \frac{\exp(V_{1k} + \upsilon_k(X_i))}{\sum_{k'} \exp(V_{1k'} + \upsilon_{k'}(X_i))} dF_X$$

where $V_{1k} = \frac{\rho_{\chi_k}\mu_{\chi k}+\rho_\eta s_{1k}}{\rho_{\chi_k}+\rho_\eta}$.

At period t, voter i uses two sources of information, his own signal s_{itk} and the voting record for all $t' < t$, h_t. Since voter i knows the cost distribution of voters and observes the turnout in preceding States, he can infer the cutoff $\lambda_{kt'}$ and hence the realization of $M_{kt'}$ for all $t' < t$. Thus, the voter's posterior on χ_k is given by

$$B_{tk} \sim N\left(\frac{\rho_{\chi_k}\mu_{\chi k} + \rho_\eta \sum_{\tau=1}^{t} s_{\tau k}}{\rho_{\chi_k} + t\rho_\eta}, \frac{1}{\rho_{\chi_k} + t\rho_\eta}\right)$$

and the supporter population is given by

$$M_{tk} = M_t \int \frac{\exp(V_{tk} + \upsilon_k(X_i))}{\sum_{k'} \exp(V_{tk'} + \upsilon_{k'}(X_i))} dF_X,$$

where $V_{tk} = \frac{\rho_{\chi_k}\mu_{\chi k}+\rho_\eta \sum_{\tau=1}^{t} s_{\tau k}}{\rho_{\chi_k}+t\rho_\eta}$.

Finally, the equilibrium voter turnout in State t for candidate k, denote by G_{tk}, is written as

$$G_{tk} = \int F_c(\lambda_{tk}|X_i) \frac{\exp(V_{tk} + \upsilon_k(X_i))}{\sum_{k'} \exp(V_{tk'} + \upsilon_{k'}(X_i))} dF_X. \tag{2}$$

3 Data

We use county and township breakdown (as opposed to State-level observation) of the number of registered voters and their party affiliation on the day of the 2004 Democratic Presidential Primary and the vote counts of each of the candidates.[1] (we refer to both counties and townships as municipalities for the rest of the paper

[1] Township level voting data was available only for some States. We used township level data for Conneticut, Delaware, New Hampshire, Rhode Island, and Vermont: The size of a township in these States were roughly equivalent to the size of counties in other States in terms of population. Therefore we deemed the use of township level data in these States appropriate.

Clark	Dean	Edwards	Kerry	Other	# of Obs.
5.73%	11.69%	21.41%	52.97%	8.19%	2,467
(0.071)	(0.156)	(0.120)	(0.156)	(0.05)	
–	9.01%	16.46%	69.46%	5.06%	186
–	(0.033)	(0.075)	(0.085)	(0.027)	
4.96%	–	10.35%	70.00%	14.69%	56
(0.018)	–	(0.034)	(0.596)	(0.061)	
–	13.71%	–	79.35%	13.72%	66
–	(0.035)	–	(0.044)	(0.035)	
–	–	–	73.55%	26.45%	199
–	–	–	(0.116)	(0.116)	

Table 2: Mean of Municipality-level Vote Shares. Standard errors are reported in paranthesis. The first row reports the vote shares when all of the four major candidates are on the ballot. The second to the fifth row reports numbers when some of the candidate(s) are not on the ballot.

and distinguish them from States.). The municipality-level data on vote counts was obtained through *David Leip's Atlas of U.S. Presidential Elections*, a website which collects election data. The data on the number of registered voters and their party affiliation (Democrat, Republican, Independent, etc.) was collected from the websites of each State (usually the Board of Election of each State). While the majority of States require voters to report their party affiliation at the time of voter registration, there are some States that do not. For municipalities in these States, we imputed a value by using a regression we ran on the municipalities for which this information was available.[2]

Table 2 reports the municipality level summary statistics of the vote shares of the candidates. As some candidates drop out and disappear from the ballot during the course of the Primary, we report the numbers depending on who is on the ballot. First

[2]We regressed the number of registered Democrats on (the means and quantiles of) the distribution of the demographic characteristics of the municipalities (such as income, race, and education), as well as other factors (such as dummies for the presence of other elections and the timing of the Primary.).

	Mean	Std. Err.	Min	Max	# of Obs.
Open primary	0.331	0.383	0.179	9.246	1,453
Closed primary	0.232	0.131	0.039	0.711	808
Modified-Open primary (excl NH)	0.314	0.195	0.015	2.552	485
New Hampshire (Modified-Open)	1.401	1.085	0.079	13.571	237
Total	0.387	0.516	0.015	13.571	2,983

Table 3: Descriptive Statistics of Number of Total Votes Cast Divided by Number of Registered Democrats: This variable is equivalent to turnout only for Closed primaries. It could be larger than one in Open and Modified-Open primaries.

row reports the vote shares when all four major candidates (Clark, Dean, Edwards, and Kerry) are present in the election. The other rows report the vote shares when some of the major candidates are not present on the ballot. Kerry was the eventual winner of the Primary, and he has the highest vote share in all five rows, with more than 50% on average.

There are roughly three types of Presidential Primaries depending on who the eligible voters are: Open, Modified-Open, and Closed Primaries. The format differs across States but it is constant across all municipalities within a State. In Closed Primaries, only registered Democrats are eligible to vote. In Modified-Open Primaries, registered Democrats as well as the Independents are eligible voters. In Open Primaries, any registered voter can vote. Note that we modify our definition of voter turnout accordingly. For Closed Primaries, we define voter turnout as the number of total votes cast divided by the number of registered Democrats. For Modified-Open Primaries, we use the number of registered Democrats and Independents in the denominator. For Open primaries, we use in the denominator, the number of registered Democrats, Independents and some fraction Δ of the rest of the voters, where Δ is a parameter to be estimated. This is because it is not realistic to assume that all voters who register as a Republican, for example, to consider voting on the Democratic Primary. Δ accounts for the fraction of voters who register as being Republican, Green, etc. and consider voting on the Democratic Primary.

Table 3 reports the number of total votes divided by the number of registered

	Mean		Mean
Race		Population/Registration/Votes	
% Black	8.19% (0.141)	Population	76,903 (290,244)
% White	87.55% (0.157)	Registered Democrats	20,062 (74,086)
% Other	4.26% (0.077)	Total Votes	5,299 (23,986)
Education		Income	
B.A. and higher	18.98% (0.109)	< 25K	33.47% (0.116)
Some College	25.04% (0.053)	25K–45K	26.19% (0.043)
Completed H.S.	34.33% (0.073)	45K–100K	32.04% (0.085)
Less Than H.S.	21.65% (0.095)	100K <	8.30% (0.074)
# of Obs.	2,974	# of Obs.	2,974

Table 4: Descriptive Statistics of Electoral Districts – Candidate Characteristics. The mean of each variable is reported. Standard errors are in parenthesis.

Democrats for Open, Modified-Open and Closed primaries respectively. For Closed primaries, the number is same as the turnout, and the mean turnout is 23.2%. For Open and Modified-Open primaries, the number can be larger than one because registered non-Democrats can also vote. Accordingly, the mean of this ratio is larger than Closed primaries at 33.1% and 31.4% for Open and Modified-Open primaries respectively. The ratio for New Hampshire, which is the first Primary and adopts Modified-Open format, is beyond one on average implying that many non-Democrats are voting in the Democratic Primary.

The demographic characteristics of each municipality (counties and townships) were collected from the 2000 Census. As demographic data for very small municipalities are not reported in the Census due to privacy concerns, we only use municipalities with a total population of more than 100. Descriptive statistics of the composition of demographic characteristics in municipalities are presented in Table 4.

Table 5 reports the information on whether there were additional races (Senatorial

		House				Governer				Governer	
		No	Yes			No	Yes			No	Yes
Senate	No	2,264	309	Senate	No	2,462	111	House	No	2,458	93
	Yes	287	123		Yes	410	0		Yes	414	18

Table 5: Presence of Other Democratic Primaries on the Same Day. Each entry corresponds to the number of municipalities with ("Yes") and without ("No") House, Senatorial, and Gubernatorial Primaries on the same day as the Presidential Primary.

primary, House Primary[3], Gubernatorial primary) that took place at the same time as the Presidential Primary. We use this information to control for the effects of other races on the turnout of the Presidential Primary. This information was obtained from the webpages of the Secretary of State and Board of Elections of each State. In 410 (13.8%) municipalities, there was a Democratic U.S. Senatorial Primary on the same day as the Presidential Primary. The same for House and Gubernatorial Primaries were 432 (14.53%) and 111 (3.73%), respectively.

4 Estimation

We estimate the model in two stages. First, we recover λ_{tk}, the cutoff values for voter turnout, $\upsilon_k(\cdot)$, the preference parameters, $F_c(\cdot|X)$, the distribution of voting costs, and V_i, the mean of the voters' posterior beliefs on candidate quality. This is done by using variation in the distribution of demographics, F_X, among counties within a State. In the second stage we recover the parameters related to the signals and learning using the estimated mean of voter beliefs, V_{tk}. Though the model may exhibit multiple equilibria, our estimation strategy allows us to recover the primitives of the model.

The model we estimate is the same as the one we have presented in Section 2 except for the modification that multiple states may vote on the same date to reflect the institutional feature of the Presidential Primaries. We denote the number of states voting at time t by $S(t)$.

[3]The number for Congressional Primaries includes municipalities in South Dakota, where a Special Election was held on the same day as the Democratic Presidential Primary to replace a resigned House incumbent, Bill Janklow.

First Stage We use within county level variation in turnout to estimate λ_{tk}, the cutoff values for voter turnout, υ_k, the preference parameters, $F_c(\cdot|X)$, the distribution of voting costs, and V_{tk}, the mean of the voters' posterior beliefs on candidate quality.

We use two sources of variation for identification in the first stage. From the model, we know that the cost threshold λ_{tk} determining turnout is common across all counties in a given State. Thus, we can use the variation in turnout across counties within a State to identify the cost distribution $F_c(\cdot|X)$ as a function of demographic characteristics. The variation in the vote share for the candidates (after accounting for turnout variations) identifies the preference parameters υ_k as function of demographic characteristics. On the other hand, λ_{tk} and V_{tk} can be identified from intertemporal variation in turnout and vote shares. By tracing out the turnout and vote shares for counties with similar demographics over time, we can identify λ_{tk} and V_{tk}.

We write the individual voter-level cost for voter i, c_i, as a function of the voter's demographic characteristics X_i as

$$c_i = \alpha_0 + \alpha_1 X_i + \alpha_2 Z_m + w_i + v_m.$$

Z_m is a vector of indicator variables that indicate whether there were other elections on the same day (e.g. Primary for the U.S. Senate, U.S. House, etc.). w_i and v_m are random shocks: w_i is a voter-level shock with $w_i \sim N(0, \sigma_w^2)$, and v_m is a municipality-level shock with $v_m \sim N(0, \sigma_v^2)$. The random shock w_i captures unobservable variation in voting costs which may arise from the variation in the distance between a voter's home and his voting station, for example. The municipality random shock v_m captures municipality wide variation in voting costs, such as whether local weather and traffic conditions. We do not restrict c_i to take positive values, allowing for the presence of expressive voters who receives consumption value from voting.

Note that the probability of voter i in State t to join group tk can be written as $\exp(V_{tk} + \upsilon_k(X_i))/\sum_j \exp(V_{tj} + \upsilon_j(X_i))$ following the model section. We assume $\upsilon_k(X_i)$ to be linear in X_i, and parameterize it as $\beta_k X_i$. Then, using equation (2), the turnout of group tk in municipality m can be written as

$$
\begin{aligned}
G_{tk,m} &= \int\int 1\{\alpha X_i + w_k + v_m < \lambda_{tk}\} \frac{\exp(V_{tk} + \beta_k X_i)}{\sum_j \exp(V_{tj} + \beta_j X_i)} dF_{w_k} dF_{m,X} \\
&= \int \Phi(\frac{\lambda_{tk} - v_m - \alpha X_i}{\sigma_w}) \frac{\exp(V_{tk} + \beta_k X_i)}{\sum_j \exp(V_{tj} + \beta_j X_i)} dF_{m,X} \\
&\equiv G(\lambda_{tk} - v_m : \{V_{tk}\}, \alpha, \beta, \sigma_v),
\end{aligned}
$$

where Φ is the CDF of the standard Normal distribution. Note that we do not have identification of σ_w: We set $\sigma_w = 1$ for normalization. Note that we use $\lambda_{tk} = 0$ for any States voting after the period t^* at which the number of committed delegates of a candidate is larger than the number of uncommitted ones, which was March 9th in this case. By rearranging, we have

$$
v_m = \lambda_{tk} - G^{-1}\left(G_{tk,m} : \{V_{tk}\}, \alpha, \beta, \sigma_v\right)
$$

Thus, we can use MLE to recover $\{V_{tk} - V_{tl}\}$, $\{\beta_k - \beta_l\}$, α, σ_v and λ_{tk}. The mean of the voters' posterior beliefs on candidate quality, $\{V_{tk}\}$ for each candidate takes different value in different States. We use the recovered value of $\{V_{tk} - V_{tl}\}$ in the second stage.

Second Stage In this stage, we recover belief and learning parameters which are, $\mu_{\chi k}$, the mean of the prior distribution of candidate k, $\rho_{\chi k}$, the precision of the prior, and ρ_η, the precision of the State level signal. From the first stage, we have recovered the differences in the mean of the voters' posterior beliefs on candidate quality, $V_{tk} - V_{tl}$. Note that in our estimation we also take account of the fact that more than one State may vote on the same date. We have described the specification with this modification in the Appendix. The distribution of posterior beliefs derived in the model section can now be written as

$$
N\left(\frac{\rho_{\chi k}\mu_{\chi k} + \rho_\eta \sum_{\tau=1}^{t} s_{\tau k}}{\rho_{\chi_k} + \rho_\eta t}, \frac{1}{\rho_{\chi_k} + \rho_\eta t}\right)
$$

Therefore, the expected common value of candidate quality V_{tk} can be written as

$$
V_{tk} = \frac{\rho_{\chi k}\mu_{\chi k} + \rho_\eta \sum_{\tau=1}^{t} s_{\tau k}}{\rho_{\chi_k} + \rho_\eta t}.
$$

If we assume that the precision of the priors to be equal across candidates, and impose $\rho_{\chi k} = \rho_{\chi l}$ for any k and l, then

$$V_{tk} - V_{tl} = \frac{\rho_\chi(\mu_{\chi k} - \mu_{\chi l}) + \rho_\eta \sum_{\tau=1}^{t}(s_{\tau k} - s_{\tau l})}{\rho_\chi + \rho_\eta t}.$$

The expression above has a straight-forward interpretation. The difference in the posterior mean of the candidate quality, $V_{tk} - V_{tl}$, is a weighted sum of the difference in the prior $(\mu_{\chi k} - \mu_{\chi l})$ and cumulated differences in the signals, $\sum_{\tau=1}^{t}(s_{\tau k} - s_{\tau l})$, where the weights are given by the relative size of the precision parameters ρ_χ and ρ_η. A large value of ρ_χ relative to ρ_η implies that voters place relatively little weight on the signals, and learning takes place slowly. Hence it is possible to interpret ρ_η/ρ_χ as the speed of learning. Note that the assumption of common precision, $\rho_{\chi k} = \rho_{\chi l}$, allows us to obtain a simple expression that is easy to interpret. Although $\rho_{\chi k} = \rho_{\chi l}$ is not necessary for identification nor estimation, we will maintain this assumption in order to simplify the estimation and keep the number of parameters small.

We can identify the parameters that govern beliefs and learning by exploiting the evolution of $V_{tk} - V_{tl}$ over time. For example, if $V_{tk} - V_{tl}$ is large and stays relatively large throughout, the speed of learning, ρ_η/ρ_χ, is low and difference in the mean of the prior, $\mu_{\chi_k} - \mu_{\chi l}$, is large. On the other hand, if $V_{tk} - V_{tl}$ is initially large, and the difference fluctuates, the speed of learning is high and the differences in prior is large. Thus the speed of learning is identified from how much $V_{tk} - V_{tl}$ fluctuates over time, whereas the difference in the prior is identified from the absolute magnitude of $V_{tk} - V_{tl}$. It should be clear from this discussion that we can only identify the ratio of ρ_χ and ρ_η.

In order to construct a likelihood, we rearrange the expression above as follows:

$$\sum\nolimits_{\tau=1}^{t}(s_{\tau k} - s_{\tau l}) = (\frac{\rho_\chi}{\rho_\eta} + t)(V_{tk} - V_{tl}) - \frac{\rho_\chi}{\rho_\eta}(\mu_{\chi k} - \mu_{\chi l}).$$

This implies that $s_{tk} - s_{tl}$ can be computed recursively from period $t = 1$ using the above expression from the estimated difference in the posterior mean, $V_{tk} - V_{tl}$, for any given parameter values. Hence we can construct a likelihood function in $s_{tk} - s_{tl}$ because $s_{tk} - s_{tl}$ follows a Normal distribution with mean $\mu_{\chi k} - \mu_{\chi l}$ and variance $2/\rho_\chi + 2/\rho_\eta$. The form of the likelihood can be found in the Appendix.

Finally, note that realizations of η in early dates contribute to the likelihood many

	Parameter Estimates
μ_{CLARK}	-8.46
μ_{DEAN}	22.46
$\mu_{EDWARDS}$	-17.46
$\sigma_{\chi^i}/\sigma_\eta$	4.92

Table 6: Second Stage Parameter Estimates. We use the normalization $\mu_{KERRY} = 0$

times while those in later dates do not because early shocks affect not only the beliefs of early voters but also the beliefs of all subsequent voters.

5 Results

The parameter estimates in the second stage is reported in Table 3. The ratio of the precision parameters, ρ_χ/ρ_η=4.92 implies that the voter's prior of voter beliefs are about five times larger than the precision of signal of candidate quality. In other words, the prior the voter had contained the information about five signals.

Regarding the candidate quality, μ_k, the result is surprising. We have normalized John Kerry's one to be zero, $\mu_{KERRY} = 0$. The result shows that the quality of Kerry is lower than that of the Howard Dean, $\mu_{DEAN} = 22.46$, and higher than that of John Edward's, $\mu_{EDWARDS}$ =-17.46 and Howard Clark's, μ_{CLARK} =-8.46. This implies that the Howard Dean has been hit with negative shocks in early stages, and the persistent impact of those shocks have affected the beliefs of all subsequent voters. In other words, if no such shocks have realized in early stages, Howard Dean would have been a winner in the race.

Include dummies for other elections.(on cost)

Change the mean we put for demographic characteristics in step1.m

6 Counterfactual Experiments

To be completed.

7 Conclusion

To be completed

References

[1] Battaglini, Marco (2005) "Sequential Voting with Abstention," *Games and Economic Behavior*, 51, pp. 445-463.

[2] Banerjee, Abhijit V. (1992) "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, 152(3), pp. 797-817

[3] Coate, Stephen, and Michael Conlin (2004) "A Group-Utilitarian Approach to Voter Turnout: Theory and Evidence," *American Economic Review*, 94(5), pp. 1476-1504.

[4] Dhillon, Amrita, and Susana Peralta (2002) "Economic Theories of Voter Turnout", *Economic Journal*, 112(480), pp. F332-F352.

[5] Feddersen, Timothy (2004) "Rational Choice Theory and the Paradox of not voting", *Journal of Economic Perspectives,* 18(1), pp. 99-112.

[6] Knight, Brian and Nathan Schiff (2007) "Momentum and Social Learning in Presidential Primaries," mimeo, Brown Univeristy

8 Appendix

In this Appendix, we describe the construction of the likelihood for the second stage. Taking account of the fact that more than one State may vote on the same date, the distribution of posterior beliefs derived in the model section can now be written as

$$N\left(\frac{\rho_{\chi k}\mu_{\chi k}+\rho_{\eta}\left[s_{tk}+\sum_{\tau=1}^{t-1}\sum_{j=1}^{S(\tau)}s_{kj}\right]}{\rho_{\chi_k}+\rho_{\eta}\left[1+\sum_{\tau=1}^{t}S(\tau)\right]},\frac{1}{\rho_{\chi_k}+\rho_{\eta}\left[1+\sum_{\tau=1}^{t}S(\tau)\right]}\right)$$

where $S(t)$ is the number of States voting on period t. Therefore, the expected common value of candidate quality V_{tk} can be written as

$$V_{tk} = \frac{\rho_{\chi k}\mu_{\chi k} + \rho_\eta \left[s_{tk} + \sum_{\tau=1}^{t-1}\sum_{j=1}^{S(\tau)} s_{kj}\right]}{\rho_{\chi_k} + \rho_\eta \left[1 + \sum_{\tau=1}^{t} S(\tau)\right]}.$$

If we assume that precision of the priors to be equal across candidates, and impose $\rho_{\chi k} = \rho_{\chi l}$ for any k and l, then

$$V_{tk} - V_{tl} = \frac{\rho_\chi(\mu_{\chi k} - \mu_{\chi l}) + \rho_\eta \left[s_{tk} - s_{tl} + \sum_{\tau=1}^{t-1}\sum_{j=1}^{S(\tau)}(s_{kj} - s_{lj})\right]}{\rho_\chi + \rho_\eta \left[1 + \sum_{\tau=1}^{t} S(\tau)\right]}.$$

This expression can be rearranged as

$$\begin{aligned} s_{tk} - s_{tl} &= \left[\frac{\rho_\chi}{\rho_\eta} + (\sum_{\tau=1}^{t-1} S(\tau) + 1)\right](V_{tk} - V_{tl}) \\ &\quad -\frac{\rho_\chi}{\rho_\eta}(\mu_{\chi k} - \mu_{\chi l}) - \sum_{\tau=1}^{t-1}\sum_{j=1}^{S(\tau)}(s_{kj} - s_{lj}) \end{aligned}$$

Thus, $s_{tk} - s_{tl}$ can be computed recursively from period $t = 1$ using the above expression from the estimated difference in the posterior mean, $V_{tk} - V_{tl}$, for any given parameter values as follows.

$$\begin{aligned} s_{k1,j} - s_{l1,j} &= \left[\frac{\rho_\chi}{\rho_\eta} + 1\right](V_{k1,j} - V_{l1,j}) - \frac{\rho_\chi}{\rho_\eta}(\mu_{\chi k} - \mu_{\chi l}) \\ &\equiv H_{kl,j}(1) \end{aligned}$$

where index j indicates the State which votes on the same date. Using $H_{kl,j}(1)$ computed above, we can compute $s_{k2,l} - s_{l2,j}$ as

$$\begin{aligned} s_{k2,j} - s_{l2,j} &= \left[\frac{\rho_\chi}{\rho_\eta} + (S(1) + 1\right](V_{k2,j} - V_{l2,j}) \\ &\quad -\frac{\rho_\chi}{\rho_\eta}(\mu_{\chi k} - \mu_{\chi l}) - \sum_{\varphi=1}^{S(1)} H_{kl,\varphi}(1) \\ &\equiv H_{kl,j}(2). \end{aligned}$$

In the same way, we can compute

$$\begin{aligned} s_{tk,j} - s_{tl,j} &= \left[\frac{\rho_\chi}{\rho_\eta} + \sum\nolimits_{\tau=1}^{t-1} S(\tau) + 1\right](V_{tk,j} - V_{tl,j}) \\ &\quad -\frac{\rho_\chi}{\rho_\eta}(\mu_{\chi k} - \mu_{\chi l}) - \sum\nolimits_{\tau=1}^{t-1} \sum_{\varphi=1}^{S(\tau-1)} H_{kl,\varphi}(\tau - 1) \\ &\equiv H_{kl,j}(t). \end{aligned}$$

Hence we can construct a likelihood function in $s_{tk} - s_{tl}$ because $s_{tk} - s_{tl}$ follows a Normal distribution with mean $\mu_{\chi k} - \mu_{\chi l}$ and variance $2/\rho_\chi + 2/\rho_\eta$. Thus, we can write the likelihood as

$$\begin{aligned} L(\mu_{\chi 1}, ..., \mu_{\chi 4}, \rho_\eta/\rho_\chi) &= \prod_{t,k,l,j} \phi(H_{kl,j}(t)) \\ &= \prod_{t,k,l.j} \frac{1}{2\sqrt{\pi(2/\rho_\chi + 2/\rho_\eta)}} \exp\left(\frac{\left(H_{kl,j}(t) - (\mu_{\chi_k} - \mu_{\chi l})\right)^2}{-2(2/\rho_\chi + 2/\rho_\eta)^2}\right). \end{aligned}$$