Momentum in Presidential Primaries

Abstract

This paper investigates the role of momentum in sequential voting systems, such as the U.S. presidential primary. In particular, we develop and estimate a simple discrete choice econometric model with social learning, which generates momentum effects. In the model, voters are uncertain about candidate quality, and voters in late states attempt to infer the information held by voters in early states from aggregate voting returns. Candidates experience momentum effects when their performance in early states exceeds voter expectations. The magnitude of momentum effects depend upon prior beliefs about the quality of candidates held by voters, expectations about candidate performance, and the degree of variation in state-level preferences. The empirical application focuses on the 2004 Democratic primary. We find that Kerry benefitted substantially from surprising wins in early states and took votes away from Dean, who stumbled in early states after holding strong leads in polling data prior to the primary season. The estimated model demonstrates that social learning is strongest in early states and, by the end of the campaign, returns in other states are virtually ignored by voters in the latest states. Finally, we simulate the election under a counterfactual simultaneous primary and find that the primary would have been much closer under such a system due to the absence of momentum effects.

Brian Knight
Department of Economics
Brown University
Providence RI 02912
Brian_Knight@brown.edu

Nathan Schiff
Department of Economics
Brown University
Providence RI 02912
Nathan_Schiff@brown.edu

1 Introduction

While voting occurs simultaneously in many elections, voters choose sequentially in other cases, such as in roll-call voting in legislatures and in general elections for many federal offices prior to 1872. The most widely discussed example of a sequential election, however, is the Presidential primary. As shown in Figure 1, the 2004 Democratic primary season began with the Iowa caucus on January 19, followed by the New Hampshire primary on January 27 and mini-Super Tuesday on February 3 when voting occurred in the states of Arizona, Delaware, Missouri, New Mexico, North Dakota, Oklahoma, and South Carolina. Various primaries occurred during the months of March, April, and May with the process coming to an end with primaries in Montana and New Jersey on June 8. As shown in Figure 2, the 2008 schedule is expected to become increasingly front-loaded as Nevada has been scheduled between Iowa and New Hampshire, and, perhaps more importantly, many states, including California, have moved their primaries to February 5. While sequential aspects will likely remain important, this front loading in the 2008 schedule has led to February 5 being dubbed a "national primary."

When considering such changes in the primary schedule, one naturally wonders whether or not the order of voting matters. Relatedly, do sequential, relative to simultaneous, systems lead to different outcomes in terms of the selection of candidates? And, if so, why? In our view, as well as the view of others, the key distinction is that sequential, relative to simultaneous, elections provide late voters with an opportunity to learn about the desirability of the various candidates from the behavior of early voters. This opportunity for late voters to learn from early voting returns can in turn lead to momentum effects, defined as a positive effect of candidate performance in early states on candidate performance in later states.

While conventional wisdom holds that such momentum effects are important in sequential elections, any econometric attempt to identify their existence and measure their magnitude faces several challenges. First, what is the informational content of voting returns from early states? Do the absolute returns matter or should results be measured relative to voter expectations regarding candidate performance? If returns should be gauged relative to expectations, how can these expectations be measured? Second, how should researchers account for unobserved candidate quality? The fact that eventual winners tend to do well in early states has often been interpreted as evidence of momentum effects. But success in both early and late states could simply reflect the

underlying strength of these candidates. Said differently, winners in early states might have won the overall primary even with a simultaneous primary system under which momentum effects play no role. Third, how do voters weigh the voting returns from different states? For example, how should voters in states third in the sequence, such as those in South Carolina, weigh the returns from Iowa, the first state, relative to those from New Hampshire, the second state. Relatedly, how do voters account for the fact that voters in states earlier in the sequence might also condition on returns from even earlier states? More concretely, when attempting to learn about the desirability of candidates from voting returns in Iowa and New Hampshire, how do voters in South Carolina account for the fact that, before casting their ballots, voters in New Hampshire may have also attempted to learn about candidates from voting returns in Iowa?

In this paper, we attempt to overcome these econometric challenges through the development a simple discrete choice econometric model of voting and social learning. In the model, voters are uncertain about candidate quality, which is valued by all voters regardless of their ideology and can be interpreted, for example, as competence or integrity. Voters gather information about quality during the campaign, and voters in late states attempt to uncover the information of early voters from voting returns in these states. In the context of this model, we show that candidates benefit from momentum effects when their performance in early states exceeds voter expectations. Momentum is thus not exclusive to winners, who may actually experience reverse momentum effects if their margin of victory is smaller than expected. The degree of such momentum effects depends upon a variety of factors, including voters prior beliefs about the quality of candidates, expectations about candidate performance, and the degree of variation in state-level preferences.

In order to estimate the degree of social learning in sequential elections, we examine voting in the 2004 Democratic primaries. In particular, we examine reactions of respondents from late states in daily polling data to the revelation of aggregate voting outcomes in early states. To the extent that social learning is important, unexpected strength in voting returns from early states should lead to improved candidate evaluation by voters in late states in the daily polling data. The parameters of interest are those governing the social learning process and are chosen to reflect the dynamics in the polling data. Our estimates demonstrate substantial social learning and momentum effects. Finally, we use the model to simulate primary voting under a counterfactual simultaneous election.

The paper proceeds as follows. Section 2 provides an overview of the relevant theoretical and

empirical literature. Section 3 lays out the basic theoretical and econometric model of momentum in primaries. Section 4 describes our empirical application, section 5 describes the counterfactual simulations, and section 6 describes possible extensions and summarizes our key findings.

2   Literature Review

Welch (1992), Bikhchandani, Hirshleifer, and Welch (1992), and Banerjee (1992) provide the first formal analysis of social learning. Agents choose actions sequentially and are uncertain about the correct action, which depends upon the state of the world. Payoffs are thus correlated and agents may attempt to learn the correct action from the behavior of others. If agents are sufficiently unsure about the true state of the world, then they may ignore their private signals and simply follow the actions of others. Such behavior has become known alternatively as informational cascades or herding. Such cascades are fragile in the sense that small changes in early signals can lead to large changes in subsequent behavior. Also, cascades can lead to inefficient outcomes if realized early signals are outliers and thus not representative of the true state of the world.[1]

These ideas have also been applied in the context of sequential elections, the key distinction being that voters make a social choice, rather than a private choice as in the literature described above. Social choices depend upon the voting behavior of all agents, and rational agents recognize that their action only matters if they are pivotal, defined as situations in which their vote changes the voting outcome. Thus, incentives are quite different in these models, and it is not clear that social learning and herding generalize to this environment of strategic voting. Feddersen and Pesendorfer (1997) first address this issue in the context of a model with a binary, symmetric, and simultaneous election. Given that pivotal voters are choosing candidates based upon their private signal, the selected candidate is the same regardless of whether voters observe only their private information or whether all information is public.[2] Dekel and Piccione (2000) extend this result to sequential elections and show that every equilibrium of the simultaneous game is also an

---

[1] This social learning framework has been applied in a variety of settings. Welch (2000), for example, studies herding among security analysts. For a general overview of social learning in finance, see Devenow and Welch (1996). In development economics, social learning has been shown to play a key role in the choice of technology, such as in Foster and Rosenzweig (1995) and Munshi (2004). For a more comprehensive overview of the social learning literature, see the survey by Sushil, Hirshleifer, and Welch (1998) and the book by Chamley (2004).

[2] Specifically, they state that as the size of the electorate goes to infinity the percentage of voters basing their choice on their own private signal approaches zero. At the same time, the number of voters who vote based on their private signal goes to infinity so that in large elections most privately held information is revealed.

equilibrium of the sequential game, regardless of the sequence. Strategic voters know that their vote only matters if they are pivotal and hence they behave as if they know that all other voters are evenly divided between the two candidates. Thus, there is a symmetry between early and late voters, and it does not matter which candidate is supported by the early voters. It is important to note, however, that this result does not demonstrate an equivalence between simultaneous and sequential elections; due to multiplicity, there are equilibria of the sequential game that are not equilibrium of the simultaneous game. In particular, Ali and Kartik (2006) construct an equilibrium in posterior-based voting in the context of a sequential election. In this equilibrium, if other voters play history dependent strategies, then it is individually optimal for each and every voter to do so as well *even under strategic voting*. Intuitively, if all other voters condition on history, then early votes are more informative than later votes, breaking the symmetry underlying the Dekel and Piccione (2000) result.

In addition to social learning, several authors have suggested alternative models for momentum, both at the voter and candidate level. Callander (2004) proposes a model where every voter gains utility from both conforming, defined as supporting the eventual winner, and voting informatively, defined as supporting the best candidate based on their belief about the true state of the world. As the number of voters increases, the conforming component of utility dominates the information-based component and herding results, propelling the leading candidate to victory. On the candidate side, Klumpp and Polborn (2006) specify a model in which an early primary victory increases the likelihood of victory for one candidate and creates an asymmetry in campaign spending that furthers this advantage. Starting with two symmetric candidates, if one candidate randomly wins the first election, this winner will have a greater incentive to spend in subsequent elections while the loser will have a diminished incentive. Through this asymmetry of campaign spending, momentum is generated and can propel an early winner to overall victory. Finally, Strumpf (2002) discusses a countervailing force to momentum. In particular, a candidate who is expected to win several of the last elections can credibly commit to not dropping out of the race even if he is trailing early. From the perspective of opposing candidates, this commitment both increases the costs of running and decreases the probability of winning. This effect, which favors later winners, thus moves in the opposite direction of momentum, which favors early winners, and may make measurement of either effect more difficult.

Most of the empirical work on momentum has come from the political science literature. Bartels

(1987 and 1988) uses data from the National Election Study (NES) to predict the dynamics of the 1984 Democratic Primary. He shows that simple ratings of candidates do not fit the dynamics as well as do models that include measures of candidate viability. Bartels suggests that candidate Gary Hart's surprising early victories convinced later voters of his viability, an example of momentum. Adkins and Dowdle (2001) use cross-primary variation to measure the importance of wins in the first two elections by regressing overall primary shares on measures of primary outcomes in Iowa and New Hampshire.[3] While we find these papers to be both interesting and suggestive of momentum effects, they do not fully overcome the econometric challenges described in the introduction. In order to better address these challenges associated with measuring momentum effects, we believe that it is desirable to build an empirical model from microfoundations, and the next section provides such a framework for measurement.

3 Theoretical framework

This section lays out our basic theoretical and econometric framework for measuring momentum effects in sequential elections, and the notation here follows Chamley (2002). Given our empirical motivations, we keep things simple and make the assumptions necessary to generate a tractable empirical model. Many of these assumptions, however, will be discussed and relaxed in the empirical section to follow.

3.1 Setup

Consider a set of states (s) choosing between candidates $(c = 0, 1, ..., C)$ in a sequential election, where the order of voting is taken as given. We allow for the possibility that multiple states may vote on the same day; in particular, let Ω_t be the set of states voting on date t and let $N_t \geq 1$ be the size of this set.

Voter i residing in state s is assumed to receive the following payoff from candidate c winning

---

[3] There have also been experimental tests for momentum effects. Morton and Williams (1999) consider a model with three candidates, liberal, moderate, and conservative. Voters do not observe candidate ideology but can potentially learn about ideology from past voting. Partisan voters (liberal or conservative) are risk averse and thus would rather vote for the moderate if they believe that only the moderate and the opposing candidate have a chance of winning. The authors test this hypotheses in a laboratory setting and find that later voters do use the early results and that a sequential election increases the likelihood of victory for moderate, unknown candidates. In addition, Battaglini, Morton, and Palfrey (2005) test predictions of the sequential voting model of Battaglini (2005), which incorporates costly voting and endogenous turnout.

the election:

$$u_{cis} = q_c + \eta_{cs} + \nu_{cis} \tag{1}$$

where q_c represents the quality of candidate c, η_{cs} represents a state-specific preference for candidate c, and ν_{cis} represents an individual preference for candidate c and is assumed to be distributed type-I extreme value and independently across both candidates and voters. We normalize utility from the baseline candidate to be zero for all voters ($u_{0is} = 0$). While underlying preferences are assumed to be stable, or time-independent, there is uncertainty and expectations may evolve during the election, as described below.

We assume the following information structure. Voters know their own state-level preference (η_{cs}) but not those in other states. Voters do, however, know the distribution from which these state-level preferences are drawn. In particular, we assume that state-level preferences are normally distributed [$\eta_{cs} \sim N(0, \sigma_\eta^2)$] and independently across states. We further assume that voters are uncertain over candidate quality and are Bayesian. In particular, initial ($t = 1$) priors over candidate quality (q_c) are assumed to be normally distributed with a candidate-specific mean μ_{c1} and a variance σ_1^2 that is common across candidates. Under the assumptions to follow, the posterior distribution will be normal as well. Before going to the polls, all voters in state s receive a noisy signal (θ_{cs}) over the quality of candidate c :

$$\theta_{cs} = q_c + \varepsilon_{cs} \tag{2}$$

where the noise in the signal is assumed to be normally distributed [$\varepsilon_{cs} \sim N(0, \sigma_\varepsilon^2)$] and independently across states. These signals can be interpreted in a variety of ways, including personal meetings with candidates, media coverage of candidate debates within the state, local newspaper endorsements of candidates, political advertising on local television channels, media coverage of candidate appearances in the state, etc. We assume that this signal is common within a state but is unobserved by voters in other states.[4]

---

[4] We feel that this assumption of a common signal within states is reasonable given the role of the mass media in modern elections. However, some campaign messages, such as mailings, can be targeted to individual voters, suggesting an alternative formulation that would allow for voters within the same state to receive independent signals. This formulation implies that, in the absence of heterogeneity in state-level preferences ($\sigma_\eta^2 = 0$), quality is perfectly revealed by voting returns from states with large populations. Thus, voters will learn only from returns in the first state and will ignore both their private signals and voting returns from other states thereafter. We view this feature of a model with individual-level signals as both unattractive and unrealistic and thus focus on the case of state-level signals. One could also consider a hybrid model with both individual-level and state-level signals. While this formulation would overcome the problem of perfect revelation of quality after voting in the first state, as described above, it is not clear how the variance in these two signals, which is a key parameter of interest in the empirical analysis to follow, would be separately identified.

Given the state-level signal (θ_{cs}), expected utility for voter i in state s from candidate c winning can be written as follows:

$$E(u_{cis}|\theta_{cs}) = E(q_c|\theta_{cs}) + \eta_{cs} + \nu_{cis} \tag{3}$$

Finally, regarding voter behavior, we assume sincere voting. That is, given the information available to voter i in state s at time t, voters support the candidate who maximizes their expected utility. We thus abstract from several forms of strategic voting under which optimal voter behavior may depend upon the behavior of other voters. This issue of sincere versus strategic voting will be discussed more completely below in the empirical application.

3.2 Voting behavior

Then, for voters in state s observing a signal over quality (θ_{cs}) and with a prior given by (μ_{ct}, σ_t^2), private updating over quality is given by:

$$E(q_c|\theta_{cs}) = \alpha_t \theta_{cs} + (1 - \alpha_t)\mu_{ct} \tag{4}$$

where the weight on the signal is given by:

$$\alpha_t = \frac{\sigma_t^2}{\sigma_t^2 + \sigma_\varepsilon^2} \tag{5}$$

Reflecting well-known results in the literature on Bayesian updating, voters thus place more weight on their private signal the higher is the variance in the prior over quality (σ_t^2) and the lower is the degree of noise in the signal (σ_ε^2).

Given this updating rule, aggregate vote shares in state s voting at time t can be described as follows:

$$\ln(v_{cst}/v_{0st}) = \eta_{cs} + \alpha_t \theta_{cs} + (1 - \alpha_t)\mu_{ct}. \tag{6}$$

where v_{cst} is the vote share for candidate c and v_{0st} is the vote share for the baseline candidate. Thus, the log-odds ratio can be expressed as a linear combination of state-level preferences (η_{cs}), the signal (θ_{cs}) received by voters in state s, and the mean of the quality distribution (μ_{ct}) prior to the realization of the signal, where the relative weight on the latter two terms depends upon the parameter α_t. As will be seen below, this expression for aggregate voting returns provides the key link between the individual-level voting data and the aggregate returns in the econometric

7

formulation, and the linearity will be a particularly attractive feature in the analysis of social learning from early voting returns.

3.3 Social learning and momentum

From the perspective of measuring momentum, the key question is then how voters in late states update their beliefs upon observing vote shares in early states. Given that state preferences (η_{cs}) are unobserved by voters in other states, signals (θ_{cs}) cannot be inferred directly from vote shares in equation (6). Using the fact that $\theta_{cs} = q_c + \varepsilon_{cs}$, however, we can say that transformed vote shares provide a noisy signal of quality:

$$\frac{\ln(v_{cst}/v_{0st}) - (1 - \alpha_t)\mu_{ct}}{\alpha_t} = q_c + \frac{\eta_{ct}}{\alpha_t} + \varepsilon_{cs} \tag{7}$$

where the variance of the noise equals $(\sigma_\eta^2/\alpha_t^2) + \sigma_\varepsilon^2$. If voters in early states place little weight (α_t) on their private signal and, under our assumption that state-level preferences are unobserved, then voting returns from these early states are a very noisy signal of quality from the perspective of late voters. Given $N_t \geq 1$ such signals, the posterior distribution is also normal and can thus be characterized by its first two moments:

$$\mu_{ct+1} = \beta_t \left[\frac{1}{N_t} \sum_{s \in \Omega_t} \frac{\ln(v_{cst}/v_{0st}) - (1 - \alpha_t)\mu_{ct}}{\alpha_t} \right] + (1 - \beta_t)\mu_{ct} \tag{8}$$

$$\frac{1}{\sigma_{t+1}^2} = \frac{1}{\sigma_t^2} + \frac{N_t}{(\sigma_\eta^2/\alpha_t^2) + \sigma_\varepsilon^2} \tag{9}$$

where the weight on the public information is given by:

$$\beta_t = \frac{N_t \sigma_t^2}{N_t \sigma_t^2 + (\sigma_\eta^2/\alpha_t^2) + \sigma_\varepsilon^2} \tag{10}$$

Before describing the evolution of the mean of the belief distribution, we note that the precision of the posterior, defined as the inverse of the variance $(1/\sigma_{t+1}^2)$, is increasing in the number of states (N_t) voting at time t along with the degree of precision in these voting returns $[(\sigma_\eta^2/\alpha_t^2) + \sigma_\varepsilon^2]^{-1}$. To provide further interpretation of this social learning rule, it is useful to re-write equation (8) as follows:

$$\mu_{ct+1} = \mu_{ct} + \frac{\beta_t/N_t}{\alpha_t} \sum_{s \in \Omega_t} [\ln(v_{cst}/v_{0st}) - \mu_{ct}] \tag{11}$$

Social learning $(\mu_{ct+1} - \mu_{ct})$ thus depends upon the surprises in voting returns, defined as the deviations in vote shares from expectations over candidate performance. Interestingly, this learning

rule implies that candidates who do not win the primary in state s can still benefit from momentum effects so long as they perform well relative to expectations. At the same time, candidates who win primaries may actually experience reverse momentum effects in the event that their margin of victory is smaller than expected.

To provide a sense of the degree of social learning, note that the effect of an increase in vote shares on the mean of the posterior distribution of candidate quality can be expressed as follows:

$$\frac{\partial \mu_{ct+1}}{\partial \ln(v_{cst}/v_{0st})} = \frac{\beta_t/N_t}{\alpha_t} = \frac{\sigma_t^2 + \sigma_\varepsilon^2}{N_t \sigma_t^2 + (\sigma_\eta^2/\alpha_t^2) + \sigma_\varepsilon^2} \tag{12}$$

Interestingly, this parameter is less than one, reflecting the inability of voters in late states to perfectly infer signals from vote shares in early states due to their inability to observe state-level preferences of voters in other states. Relatedly, the social learning parameter is decreasing in the degree of heterogeneity in state-level preferences (σ_η^2). Moreover, for the special case of single-state primaries at time t $(N_t = 1)$, such as in Iowa and New Hampshire, we can say that the degree of social learning is decreasing in the degree of noise in the signal (σ_ε^2) and is increasing in the variance of the prior (σ_t^2).

4   Empirical Application

Our empirical application focuses on the 2004 Democratic primary. During the months leading up to the primary season, Howard Dean, governor of Vermont, held a substantial lead in opinion polls. After his third place finish in the Iowa caucuses, however, Dean soon lost that lead in opinion polls to the Iowa winner, John Kerry, a senator from Massachusetts, and was forced to withdraw after a disappointing performance in Wisconsin. Kerry continued his success in Iowa with a win in New Hampshire and with strong performances in all of the subsequent states. The only serious challenge to Kerry after Iowa came from John Edwards, a senator from North Carolina, who came in a surprisingly strong second in Iowa and proceeded to win in South Carolina as well. Edwards was forced to withdraw, however, on March 3, the day after a string of second-place finishes to Kerry on Super Tuesday.

4.1   Data

To measure the degree of social learning in the 2004 primaries, we examine reactions of voters in daily opinion polls to candidate performance in primaries. Individual-level data are taken from the

National Annenberg Election Survey 2004, which conducted interviews on a daily basis beginning on October 7, 2003 and continuing through the general election in November 2004. Given our focus on the primary season, we use voting intentions for 4,084 respondents who identify as likely Democratic primary voters between October 7, 2003 and March 2, 2004, the day before Edwards withdrew from the race. To be clear, these are respondents living in states that have not yet held their primaries. Voters living in states that have already voted are not asked their voting intentions in the survey and are thus excluded from our analysis. We focus on the campaigns of the three major candidates: Dean, Edwards, and Kerry, where Kerry is considered the baseline candidate.[5] Finally, as will be described below, we aim to estimate the state-specific preference parameters (η_{cs}) and we thus also delete respondents from the District of Columbia and seven small states. These individual-level data are then merged with state-aggregate vote shares from the 2004 primary season as reported on the website www.cnn.com.

Our identification strategy is illustrated in Figures 3-5. As shown, however, Dean had a substantial and stable lead over Kerry and Edwards during the month preceding the Iowa primary. As shown, Dean under-performed in Iowa relative to expectations, as captured by pre-Iowa polling levels, and voters in the Annenberg survey updated appropriately. Kerry, by contrast, outperformed expectations in Iowa, and survey respondents updated accordingly. Edwards also outperformed his pre-Iowa polling numbers and his polling numbers did increase following Iowa. After a few days, however, his support fell back to pre-Iowa levels.

4.2   Empirical Model

As noted above, our empirical strategy for identifying momentum effects involves measuring reactions of voting intentions of likely voters in polling data to aggregate voting returns in state primaries. In our econometric specification, we assume that these voters have not yet observed their private signals and their voting intentions can thus be summarized as follows:

$$\Pr(ist \text{ prefers } c) = \frac{\exp(\eta_{cs} + \mu_{ct})}{\sum_d \exp(\eta_{ds} + \mu_{dt})} \tag{13}$$

---

[5] Another candidate, Wesley Clark, was considered viable in the months leading up to the primary season. He chose, however, to not participate in the Iowa caucuses and subsequently fell out of serious contention. Given that we do not have a model of candidate campaign strategies and the possible negative signals sent by non-participation, we felt it best to exclude him from the analysis. Another candidate, Richard Gephardt, polled well prior to Iowa but withdrew from the race shortly thereafter.

To better understand our empirical strategy for estimating the parameters governing the learning process, it is useful to first note that voter updating over quality can be summarized by the weight on private signals, the weight on public signals, updating over the mean, and updating over the variance as follows:

$$\alpha_t = \frac{\sigma_t^2}{\sigma_t^2 + \sigma_\varepsilon^2} \tag{14}$$

$$\beta_t = \frac{N_t \sigma_t^2}{N_t \sigma_t^2 + (\sigma_\eta^2/\alpha_t^2) + \sigma_\varepsilon^2} \tag{15}$$

$$\mu_{ct+1} = \mu_{ct} + \frac{\beta_t/N_t}{\alpha_t} \sum_{s \in \Omega_t} [\ln(v_{cst}/v_{0st}) - \mu_{ct}] \tag{16}$$

$$\frac{1}{\sigma_{t+1}^2} = \frac{1}{\sigma_t^2} + \frac{N_t}{(\sigma_\eta^2/\alpha_t^2) + \sigma_\varepsilon^2} \tag{17}$$

As seen, with information regarding the initial priors (μ_{c1}, σ_1^2) along with the parameters σ_ε^2 and σ_η^2, one can compute the weight on the private signal in the first period (α_1) and, with this weight in hand, one can then compute the weight placed upon the public voting signals in the first period (β_1). Then, with the entire set of first-period values $(\mu_{c1}, \sigma_1^2, \alpha_1, \beta_1)$, along with information on first-period voting returns, we can successively compute the second-period values $(\mu_{c2}, \sigma_2^2, \alpha_2, \beta_2)$. With these second-period values, along with information on second-period voting returns, we can then successively compute the third-period values $(\mu_{c3}, \sigma_3^2, \alpha_3, \beta_3)$, etc.

Thus, it should be clear that the key parameters to be estimated are the distribution of the initial priors (μ_{c1}, σ_1^2) along with the variance in state-level preferences (σ_η^2) and the degree of noise in the signal (σ_ε^2). These key parameters are estimated via a two-step approach. In the first step, we use the pre-Iowa polls to estimate the initial conditions. In particular, for the case of $t = 1$, we have that:

$$\Pr(is1 \text{ prefers } c) = \frac{\exp(\eta_{cs} + \mu_{c1})}{\sum_d \exp(\eta_{ds} + \mu_{d1})} \tag{18}$$

We estimate the state-level preference parameters (η_{cs}), which are normalized to sum to zero and which can be used to calculate σ_η^2, along with a constant term, which provides an estimate of μ_{c1}. In the second step, we use reactions of voters in post-Iowa opinion polls to the revelation of voting returns in other states in order to estimate the key parameters $(\sigma_\varepsilon^2, \sigma_1^2)$ governing the social learning process. Given the two-stage estimation approach, conventional confidence intervals will

not reflect the uncertainty associated with using generated regressors in the second stage. We address this issue by computing bootstrap confidence intervals.[6]

The key social learning parameters are identified by voter responses to the release of voting returns in others states. If voters are unresponsive to the release of such information, this suggests an absence of social learning, and the variance in the initial prior (σ_1^2) will have a small estimate or the variance in the degree of noise in the signal (σ_ε^2) will have a large estimate. If voters are responsive to voting returns, by contrast, then the variance in the initial prior (σ_1^2) will have a large estimate or the variance in the degree of noise in the signal (σ_ε^2) will have a small estimate. In terms of separately identifying these two social learning parameters, note that changes in the variance of the initial prior (σ_1^2) will have strong effects on social learning in early stages and smaller effects in later stages of the primary. Changes in the variance of the noise in the signal (σ_ε^2), by contrast, have effects on social learning that are, roughly speaking, time-independent. Thus, the dynamics of social learning helps to separately identify these two key social learning parameters.

4.3   Baseline Results

Table 1 provides the results from the first-step of the estimation procedure. As shown in columns 1 and 2, the coefficient on the candidate-specific constant term demonstrates Dean's substantial lead over Kerry and Kerry's lead over Edwards prior to the commencement of the primary season. As noted above, this coefficient can be interpreted as the mean of the initial prior (μ_{c1}), and this variable will play a key role in the updating rule given by equation (16). The significant degree of variation in the state specific coefficients demonstrates the significant diversity in preferences towards the candidates across states. As shown, there are strong regional effects with Kerry holding a substantial advantage in his home state of Massachusetts, and Edwards enjoying a corresponding strong advantage in the South, with statistically significant advantages over Kerry in North Carolina and South Carolina. This advantage likely reflects the fact that Edwards was the only candidate of the three from the South. This issue of regional advantages will be considered below in an alternative specification, which relaxes the assumption that such advantages are unobserved by voters in later states.

---

[6] In particular, we draw 100 samples with replacement from the underlying sample. In some replications, an insufficient number of cases were drawn to allow for identification of the specific state fixed effects, and we thus exclude such states from the analysis in these bootstrap samples.

Table 2 provides estimates of the other key parameters. The degree of heterogeneity in state-level preferences (σ_η^2) is calculated by taking the cross-state and cross-candidate variance in the coefficients on the state dummy variables as reported in Table 1. As described above, the variance in the initial prior (σ_1^2) and the degree of noise in the signal (σ_ε^2), by contrast, are identified by gauging the reactions of voters in the daily polling data to the revelation of aggregate voting returns from state primaries. As shown, both of these parameters are positive and statistically significant.

Given the difficulties in providing a direct social learning interpretation of these parameters, we instead present in Figures 6-9 the key dynamics of the model as implied by these estimated parameters. As shown in Figure 6, for example, the degree of variance in the beliefs over candidate quality (σ_t^2) falls substantially over the primary season. Prior to the Iowa caucus, the variance in this distribution was just above 4.0, reflecting the estimated parameter in Table 2, but falls to around 0.5 by March 2, or Super Tuesday. Thus, voters learn a substantial amount over the course of the campaign about candidate quality purely from the release of voting returns in other states.

At the same time as the degree of uncertainty over candidate quality fell, voters learned about the quality of the candidates relative to one another. As shown in Figure 7, prior to the primary season, voters viewed Dean as the highest quality followed by Kerry and Edwards, reflecting the pattern of coefficients on the candidate indicator variables in Table 1. Following Kerry's win in Iowa, Kerry pulled ahead of Dean in terms of mean quality ratings. Although Kerry defeated Edwards in Iowa, voters updated favorably over Edwards relative to Kerry, reflecting the fact candidates can benefit, even relative to first place finishers, from surprisingly strong second place finishers. On the other hand, although Edwards defeated Dean in Iowa, voters still evaluated Dean and Edwards roughly equally. This in turn reflects the fact that voters also placed some weight on their beliefs prior to voting in Iowa, and these priors were strongly in favor of Dean relative to Edwards. Following New Hampshire and mini-Super Tuesday, Kerry held a strong advantage, and Dean never recovered from his weak performances in Iowa.

To provide further interpretation of these results, Figure 8 plots the implied weights on the private signals observed by voters (α_t) as well as the weight place upon aggregate vote shares after scaling by the number of primaries (β_t/N_t). As shown, voters place less weight on their prior than on the private signal at the beginning of the sample period. This in turn reflects the fact that the

estimated degree of noise in the signal is less than the estimated degree of variance in the initial prior ($\sigma_\varepsilon^2 < \sigma_1^2$) and that the weight on the private signal can be shown to be inversely related to the ratio of these parameters (i.e. $\alpha_t = (1 + \sigma_\varepsilon^2/\sigma_t^2)^{-1}$). As implied by the model, the weight placed upon the private signal falls during the primary season and, by Super Tuesday, voters place roughly 80 percent weight on their priors and only 20 percent on their private signals.

Figure 8 also plots the weight placed upon the revelation of aggregate voting returns in other states during the primary season. As shown, voters initially place roughly equal weight on these signals and their priors; the fact that this weight on public information is lower than the weight placed on the private signal reflects the inability of voters in late states to perfectly uncover the signals in early states from voting returns due to the observation noise associated with unobserved state preferences. As more and more primary returns come in, voters place less weight on voting returns and more weight on their prior. By the end of the sample, voters place almost all of the weight on their prior and are largely unmoved by developments in primaries held in other states.

While the weights on private and public signals seem to fall in a similar parallel manner in Figure 8, the weight on the public signal is quickly approaching zero, and hence the ratio of these two weights ($\beta_t/\alpha_t N_t$), which is the key social learning parameter, also falls quickly to zero. This pattern in social learning is reflected in Figure 9, where voters in late states initially learn substantially from returns in early states. The initial weight on the public signal is roughly 75 percent of the weight on the private signal. This social learning, however, falls off quickly and the weight on the public signal is around 10 percent of the weight on the private signal by the end of the sample period.

In summary, our estimated model demonstrates that voters in late states placed significant weight on Kerry's early victories. It is the deviations from expectations that matters, however, and Edwards benefitted relative to Kerry from a surprisingly strong second-place finish in Iowa. While Dean came in third place in Iowa, he benefitted from strong voter beliefs regarding his quality prior to Iowa and was able to remain viable. At the same time that voters shifted their relative evaluations of candidate quality, they became increasingly confident in these evaluations, and voters in late states thus placed less weight on both their private signals as well from returns in other states. Taken together, these results demonstrate significant social learning and momentum effects as reflected in the disproportionate influence of voters in early states.

14

4.4 Additional specifications

As noted above, the baseline model assumes that voters observe their own state-level preferences but not those in other states. What is key to the social learning result is that some component of state-level preferences is unobserved by voters in other states, and thus voters in late states cannot perfectly infer signals from voting returns in early states. If preferences are perfectly observed, then, in the case of a single primary ($N_t = 1$), public and private learning are equivalent ($\alpha_t = \beta_t$) and momentum effects are effectively assumed, rather than measured. As an alternative to this assumption of perfect observability, we consider and estimate an alternative specification in which state-level preferences consist of both an unobserved component (η_{cs}) and an observed component (X_{cs}), such as geography, which could capture advantages enjoyed by politicians campaigning in their home states. Then, aggregate voting returns can be written as follows:

$$\ln(v_{cst}/v_{0st}) = \eta_{cs} + \gamma X_{cs} + \alpha_t \theta_{cs} + (1 - \alpha_t)\mu_{ct}. \tag{19}$$

where γ is a weight, or vector of weights, on observed preferences that will be estimated. It is then straightforward to show that the social learning rule is adjusted for these observed characteristics as follows:

$$\mu_{ct+1} = \mu_{ct} + \frac{\beta_t/N_t}{\alpha_t} \sum_{s \in \Omega_t} [\ln(v_{cst}/v_{0st}) - \gamma X_{cs} - \mu_{ct}] \tag{20}$$

Thus, voters in late states incorporate these observed characteristics into their expectations of candidate performance, and, in our example of geography, returns showing that a candidate performed well in his home state, even relative to national expectations over candidate performance (μ_{ct}), do not necessarily lead to momentum effects.

To operationalize this specification, we incorporate into X_{cs} a measure of the distance between state s and the home state of candidate c, where the measure is relative to the distance between state s and Kerry's home state of Massachusetts. After the first step, or pre-Iowa, analysis, we regress the estimated fixed effects on this distance measure and use the residuals from this regression as an estimate of unobserved preferences (η_{cs}). As shown at the bottom of columns 3 and 4 of Table 1, distance has a negative and statistically significant effect on voting decisions, as reflected in polling data. After accounting for this observed dimension of preferences, the regional advantages enjoyed by candidates are diminished although the home state advantage enjoyed by Kerry and Edwards remains. As shown in Table 2, the estimated variance of unobserved

15

preferences (σ_η^2) is reduced in this model, reflecting the assumption that some component of preferences are observed by voters in other states. The other key parameters are qualitatively similar to those in column 1.

The second specification relaxes the assumption that underlying voter preferences are stable over the campaign. Trends in candidate-specific preferences could of course confound the estimation of social learning effects. To address this issue, we estimate a model with a candidate-specific trend (γ_c) in preferences. Then, aggregate voting returns are adjusted as follows:

$$\ln(v_{cst}/v_{0st}) = \eta_{cs} + \gamma_c t + \alpha_t \theta_{cs} + (1 - \alpha_t)\mu_{ct}. \tag{21}$$

where γ is a weight, or vector of weights, on observed preferences that will be estimated and t is normalized to equal zero on the date of the Iowa primary. It is then straightforward to show that the social learning rule is adjusted as follows:

$$\mu_{ct+1} = \mu_{ct} + \frac{\beta_t/N_t}{\alpha_t} \sum_{s \in \Omega_t} [\ln(v_{cst}/v_{0st}) - \gamma_c t - \mu_{ct}] \tag{22}$$

Thus, voters in late states incorporate these trends into their expectations of candidate performance. As shown in columns 5 and 6 of Table 1, the pre-Iowa trends tended to favor Dean and Edwards, while Kerry was disadvantaged. Thus, at the time of the Iowa primary, the mean evaluation of Dean and Edwards are higher than are those in the baseline specification. This is reflected in the first row of Table 1. As shown in Table 2, however, the key social learning parameters here are similar to those in the baseline specification.

4.5   Discussion of Sincere Voting Assumption

In the baseline model, we have assumed sincere voting, under which voters support the candidate that provides the highest utility level. We have thus abstracted from four possible departures from sincerity. First, given our empirical focus on presidential primaries, we abstract from electability considerations associated with the general election. That is, a voter who prefers Dean over Kerry may nonetheless vote for Kerry over Dean if she believes that Kerry has a better chance than Dean of beating Bush in the general election. To further explore this issue, consider a framework with both a primary election and a general election. Republican voters choose a candidate in the Republican primary and Democrats choose a candidate in the Democratic primary. In the subsequent general election, partisan voters are assumed to support the candidate of their party,

16

and independent voters choose the candidate that maximizes their utility, which includes a quality dimension. While a full exploration of this model is beyond the scope of this paper, we conjecture that the probability of the Democrat winning the general election is increasing (decreasing) in the perceived quality of the Democratic (Republican) candidate. Therefore, forward-looking primary voters have an incentive to support candidates with the highest perceived quality level. Thus, social learning over candidate quality continues to play an important role when primary voters consider electability although the exact mechanism is somewhat different than that in the baseline model.

Second, as described in the literature review, conditioning on being pivotal is an important consideration as it has been the focus of the theoretical literature on sequential elections. These strategies in turn can support equilibria in which voting is history-independent. We abstract from this issue given problems of multiplicity. It is comforting to note, however, that, as described above, Ali and Kartik (2006) have shown that an equilibrium in posterior-based voting exists even with strategic voting, and the voting strategies supporting this equilibrium are similar in spirit to those used here. Moreover, our empirical finding in favor of momentum documents that voters are playing history-dependent strategies.

The third form of strategic voting involves concerns over wasted votes if there are more than two candidates. That is, supporters of minor candidates might vote for another candidate if their first choice is perceived as having little or no chance of winning. This behavior may mimic social learning even if preferences over candidates are uncorrelated across voters and thus provides an alternative explanation for our empirical results. That is, even though voters do not learn about candidate quality from voting returns, they may learn about the prospects of winning for the various candidates and may thus switch their allegiance away from minor candidates who fare poorly in early primaries. Given the problems associated with multiplicity in this case, we feel that sincere voting is a natural starting point and leave this issue for future researchers. We do, however, address this issue below by examining voter responses to questions about the favorability of candidates. If voters only learn about the prospects of candidates winning from early elections, then their evaluation of candidate favorability should not respond to the results from early elections.

The final departure from sincere voting involves bandwagon effects, under which voters have conforming preferences and thus vote for the candidate expected to win. Similarly, to concerns over

wasted votes, bandwagon effects may mimic social learning even if preferences are uncorrelated across voters and thus provides an alternative explanation for our empirical results. That is, even though candidates do not learn about candidate quality from voting returns, they may update their beliefs about the eventual winner of the primary and may thus switch their allegiance to winners in early states. We also address this issue below by examining voter responses to questions about the favorability of candidates. Under bandwagon voting, voter evaluation of candidate favorability should not respond to the results from early elections.

To shed further light on our assumption of sincere voting, we examine the link between measures of individual evaluation of candidate quality, as reflected in favorability measures, and voting behavior, as reflected by the implied mean candidate quality at that time (μ_{ct}). More concretely, we examine voter responses in the polling data to candidate evaluation questions, in which voters were asked to evaluate the favorability of each candidate on a 1-10 scale. Using these responses, we then estimate the following regression model:

$$\text{favorability}_{itc} = \delta_c + \kappa \times \mu_{ct} \tag{23}$$

where favorability is measured relative to Kerry and δ_c represents a candidate specific constant. Under our assumption of sincere voting, we would expect $\kappa > 0$, whereas, as argued above, the parameter κ should equal zero under the two models, wasted votes and bandwagon effects, that provide alternative explanations for our results. As shown in column 1 of Table 3, there is a strong link between these two factors, providing support for our sincere voting assumption. The additional columns provide responses to various measures of candidate quality. As shown, the first additional six measures have the expected positive coefficients, whereas the coefficient associated with the measure of "reckless" has the expected negative sign. The final two, however, are statistically insignificant, likely reflecting the reduced sample sizes.[7]

5 National primary simulation

In order to further highlight the importance of momentum and social learning in sequential elections, we next provide a counterfactual simulation in which every state votes in a simultaneous national primary on January 19 and thus held the same prior as did voters in Iowa when they

---

[7] All respondents were queried as to candidate favorability but were then randomly queried as to the additional traits.

went to the poll. Accordingly, behavior in later states may be altered for two reasons. First, voter priors favored Howard Dean in the days leading up to Iowa and thus it is reasonable to expect that he would have performed better in a simultaneous national primary. Second, at the time Iowa voted, voters were less certain in their evaluations of candidate quality and thus placed more weight on their private signals. Thus, if their signals were in favor of Dean, then that will be amplified in a simultaneous primary. This second effect, however, could benefit any of the three candidates depending upon the distribution of the realized signals.

In the baseline model, vote shares in state s under a national primary at time $t = 1$ can be summarized as follows:

$$\ln(v_{cs1}/v_{0s1}) = \eta_{cs} + \alpha_1 \theta_{cs} + (1 - \alpha_1)\mu_{c1} \tag{24}$$

In order to calculate vote shares under this counterfactual simulation, note that equations 6 and 24 can be combined as follows:

$$\ln(v_{cs1}/v_{0s1}) = \eta_{cs} + \alpha_1 \left[\frac{\ln(v_{cst}/v_{0st}) - (1 - \alpha_t)\mu_{ct} - \eta_{cs}}{\alpha_t} \right] + (1 - \alpha_1)\mu_{c1} \tag{25}$$

To highlight the two effects of a move to the national primary discussed above, note that the change in vote shares from a movement to simultaneous elections can be written as follows:

$$\ln(v_{cs1}/v_{0s1}) - \ln(v_{cst}/v_{0st}) = (\alpha_1 - \alpha_t)\theta_{cs} + (1 - \alpha_1)\mu_{c1} - (1 - \alpha_t)\mu_{ct} \tag{26}$$

As shown, the first effect described above, changes in voting patterns based upon changes in the mean of the prior, is captured by the latter two terms. The second effect, changes in vote shares based upon changes in the weight placed upon the signal, is captured by the first term.

Table 4 provides the key results from the actual sequential primary and the counterfactual simultaneous primary based upon the baseline estimated coefficients in tables 1 and 2. As noted above, Dean dropped out of the race following the Wisconsin primary and we thus calculate two-candidate vote shares for states thereafter.[8]   Reflecting the importance of the momentum effects documented above, we find that both Dean and Edwards would have fared substantially

---

[8] Of course, under a national primary, he would have been on the ballot in every state. But we abstract from that issue given our inability to observe actual returns $\ln(v_{cst}/v_{0st})$ in equation 25. Moreover, without a model of candidate exit, it is difficult to predict how Dean would have performed in subsequent states following his decision to drop out of the race.

better under a simultaneous national primary with Dean winning in Michigan, Washington, Maine, and Nevada and Edwards winning in many of the southern states and also performing well on Super Tuesday. We do not wish to over-emphasize the predictive nature of the results from this simulation for specific states, such as the surprising finding that Edwards would have won Massachusetts under a national simultaneous primary.[9] Rather, we hope that the general pattern of voting under this simulation helps to further reinforce our findings of significant momentum and social learning effects in favor of Kerry in the sequential primary system.

6   Conclusion

Given our goal to develop a tractable empirical framework, we have kept the model simple and have thus abstracted from many relevant features of electoral politics in the United States. We thus view this model as a first step in a larger research agenda and plan to extend the environment in a variety of ways in subsequent work. A first possible extension involves the media. While the process through which voters observe signals was taken as given here, one could introduce a media outlet that reports election results in early states to voters in late states. Then, possible applications include how social learning depends upon the intensity of media coverage both of the general campaign and of the specific candidates. One could also examine the interaction between social learning and possible media bias towards specific candidates and how this interaction depends upon media credibility from the perspective of voters. Second, one could model candidate entry and exit, which we have taken as given in this paper. Candidate exit would presumably depend upon the degree of social learning, which may reduce the ability of trailing candidates to make up lost ground in late states. Third, one could model the allocation of campaign resources, such as political advertising and candidate visits to specific states, as it is well-known that candidates channel such resources into early states. Relatedly, candidates may alter their platforms towards issues that are most important to voter in early states. Whether or not such strategies are effective presumably depends upon whether or not voters in later states condition on such candidate behavior when analyzing voting returns from early states. Finally, one could conduct a welfare

---

[9] This prediction of a win by Edwards in Massachusetts under a national primary reflects the fact that Edwards did better than expected from the perspective of the econometrician given Kerry's home state advantage and the state of the race going into Super Tuesday. This in turn implies that voters in Massachusetts received a positive signal regarding Edwards relative to Kerry and this signal is amplified when considering the significantly higher weight placed upon signals at the beginning of the primary season, relative to the end of the primary season.

analysis of simultaneous versus sequential elections. On the one hand, voters in later states have more information under a sequential system and thus presumably make better choices. On the other hand, signals in early states are over-weighted and, in the event that early signals are outliers and thus not representative of candidate quality, then undesirable herding may occur.

In summary, we have developed and estimated a simple model of voter behavior under sequential elections. In the model, voters are uncertain about candidate quality, and voters in late states attempt to infer private information held by early voters from voting returns in early states. Candidates experience momentum effects when their performance in early states exceeds voter expectations. The magnitude of momentum effects depend upon prior beliefs about the quality of candidates held by voters, expectations about candidate performance, and the degree of variation in state-level preferences. Our empirical application focuses on the 2004 Democratic primary. We find that Kerry benefitted substantially from surprising wins in early states and took votes away from Dean, who stumbled in early states after holding strong leads in polling data prior to the primary season. The estimated model demonstrates that social learning is strongest in early states and that, by the end of the campaign, returns in other states are virtually ignored by voters in the latest states. Finally, we simulate the election under a counterfactual simultaneous primary and show that the race would have been much tighter under such a system.

References

R.E. Adkins and A.J. Dowdle. How Important Are Iowa and New Hampshire to Winning Post-Reform Presidential Nominations? *Political Research Quarterly*, 54(2):431–444, 2001.

S. Ali and N. Kartik. A Theory of Momentum in Sequential Voting. *working paper, UC-San Diego*, 2006.

A.V. Banerjee. A Simple Model of Herd Behavior. *The Quarterly Journal of Economics*, 107(3):797–817, 1992.

L.M. Bartels. Candidate Choice and the Dynamics of the Presidential Nominating Process. *American Journal of Political Science*, 31(1):1–30, 1987.

L.M. Bartels. *Presidential Primaries and the Dynamics of Public Choice*. Princeton University Press, 1988.

M. Battaglini. Sequential Voting with Abstention. *Games and Economic Behavior*, 51(2):445–63, 2005.

M. Battaglini, R. Morton, and T.R. Palfrey. *Efficiency, Equity and Timing in Voting Mechanisms*. Centre for Economic Policy Research, 2005.

S. Bikhchandani, D. Hirshleifer, and I. Welch. A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *Journal of Political Economy*, 100(5):992, 1992.

S. Callander. Bandwagons and Momentum in Sequential Voting. *working paper, Northwestern University*, 2002.

C. Chamley. *Rational herds*. Cambridge University Press Cambridge, 2004.

E. Dekel and M. Piccione. Sequential Voting Procedures in Symmetric Binary Elections. *Journal of Political Economy*, 108(1):34, 2000.

A. Devenow and I. Welch. Rational Herding in Financial Economics. *European Economic Review*, 40(3-5):603–615, 1996.

T. Feddersen and W. Pesendorfer. Voting Behavior and Information Aggregation in Elections With Private Information. *Econometrica*, 65(5):1029–1058, 1997.

A.D. Foster and M.R. Rosenzweig. Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture. *The Journal of Political Economy*, 103(6):1176–1209, 1995.

T. Klumpp and M.K. Polborn. Primaries and the New Hampshire Effect. *Journal of Public Economics*, 90(6-7):1073–1114, 2006.

R.B. Morton and K.C. Williams. Information Asymmetries and Simultaneous versus Sequential Voting. *The American Political Science Review*, 93(1):51–67, 1999.

K. Munshi. Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution. *Journal of Development Economics*, 73(1):185–215, 2004.

K.S. Strumpf. Strategic Competition in Sequential Election Contests. *Public Choice*, 111(3):377–397, 2002.

B. Sushil, D. Hirshleifer, and I. Welch. Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades. *Journal of Economic Perspectives*, pages 151–170, 1998.

I. Welch. Sequential Sales, Learning, and Cascades. *The Journal of Finance*, 47(2):695–732, 1992.

I. Welch. Herding among security analysts. *Journal of Financial Economics*, 58(3):369–396, 2000.

Figure 1: Number of Primaries by Date
2004 Democratic Primary Season
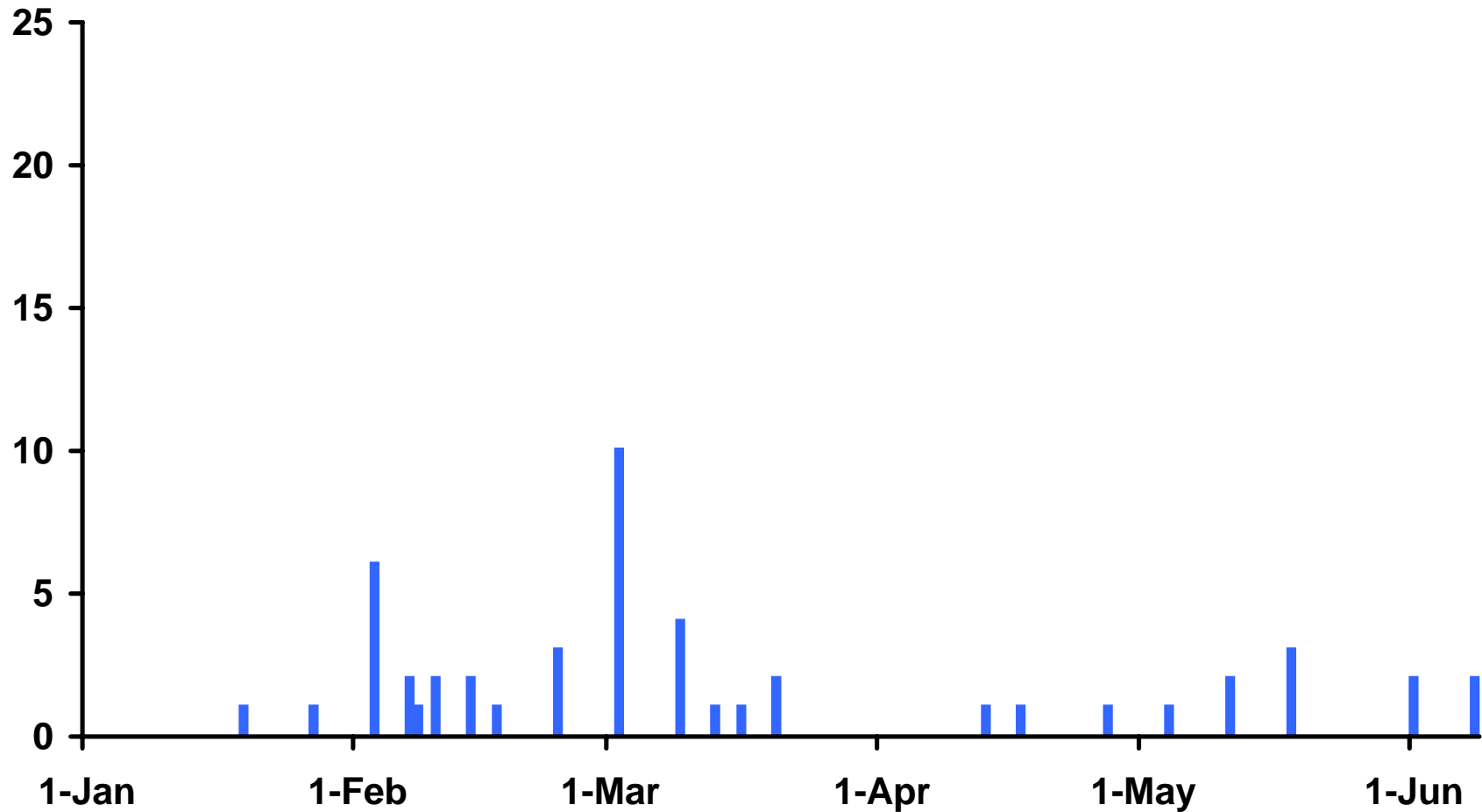
**Figure 2: Number of Primaries by Date
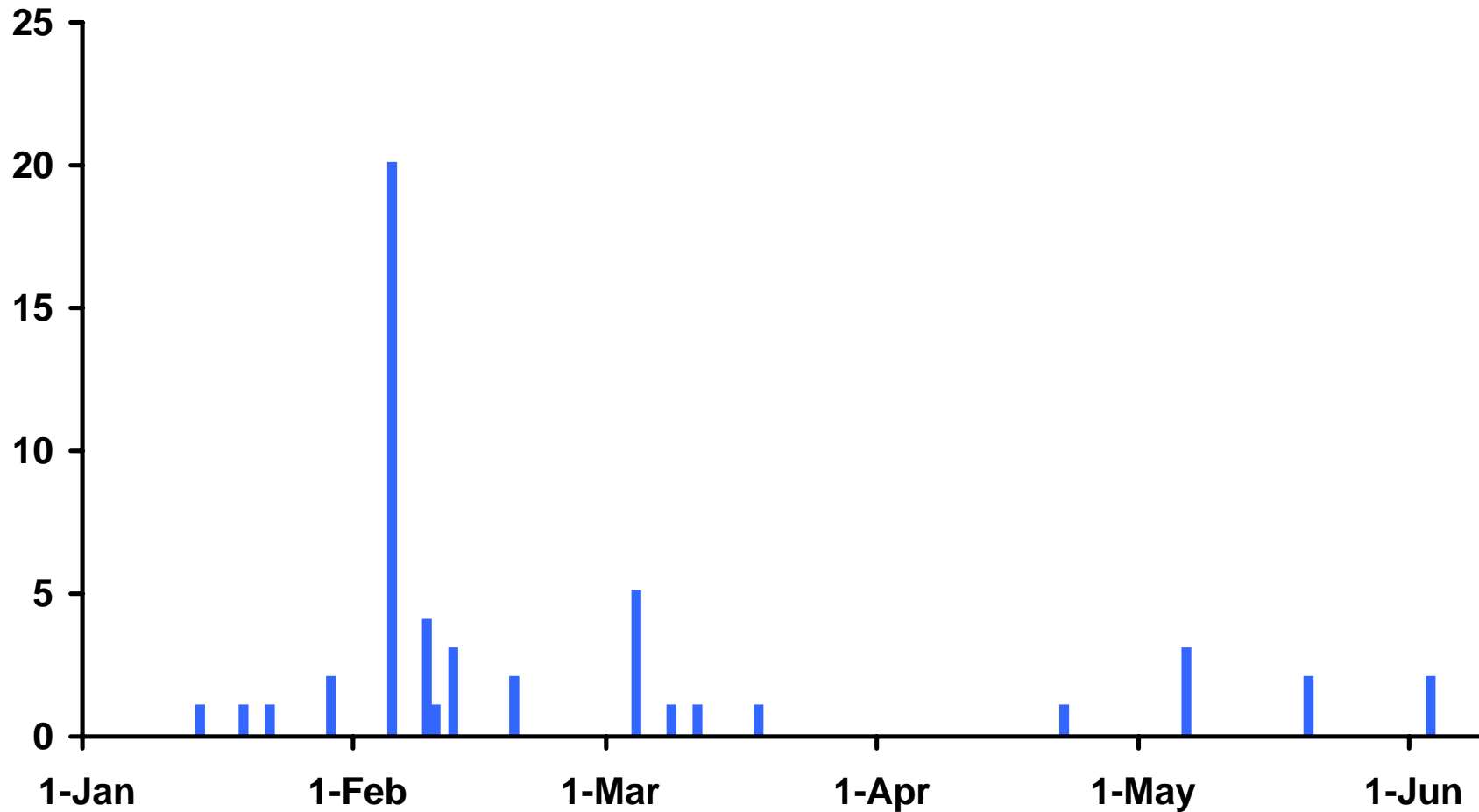2008 Democratic Primary Season**

Figure 3: Dean before and after the Iowa primary

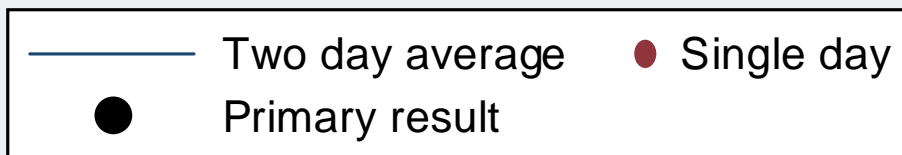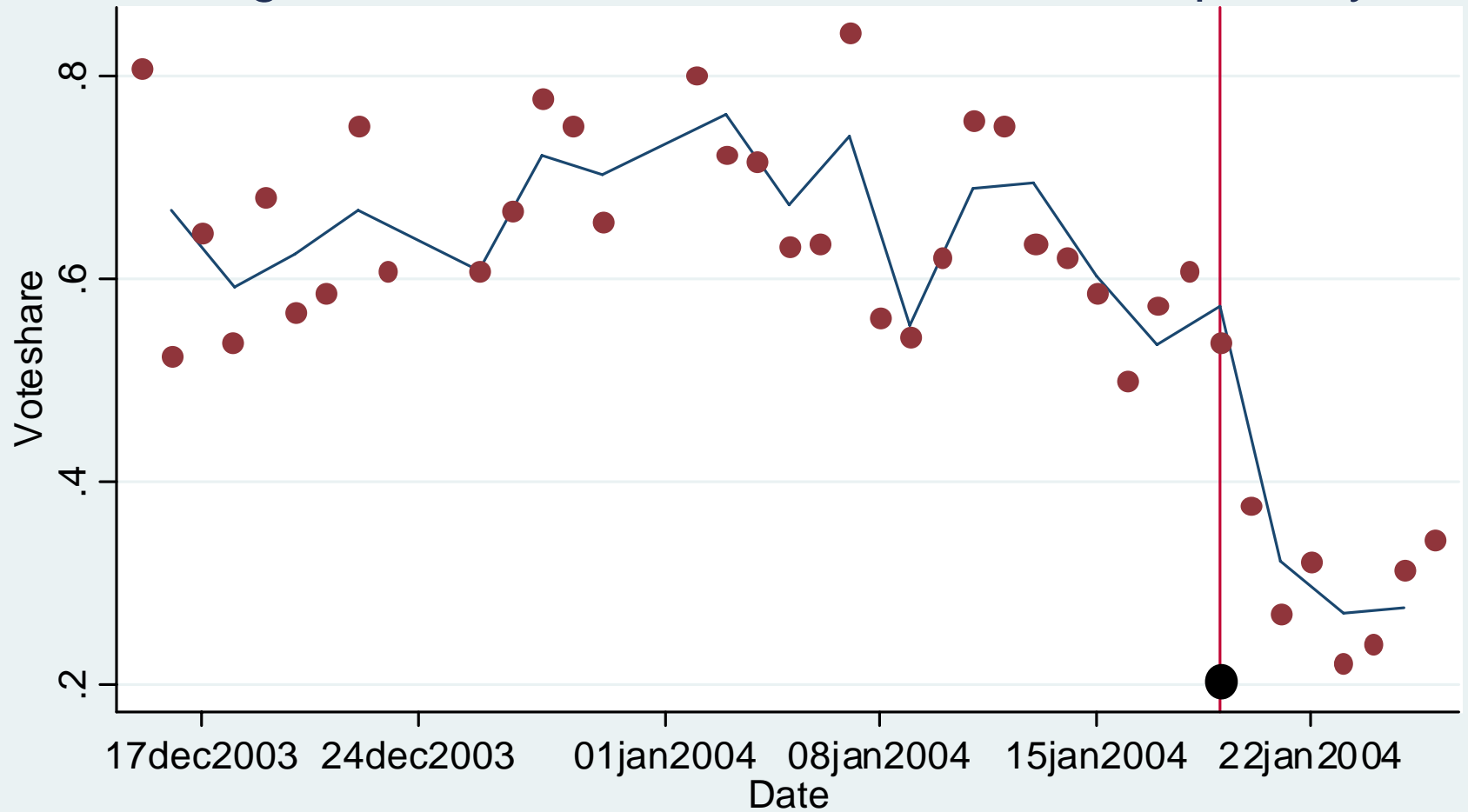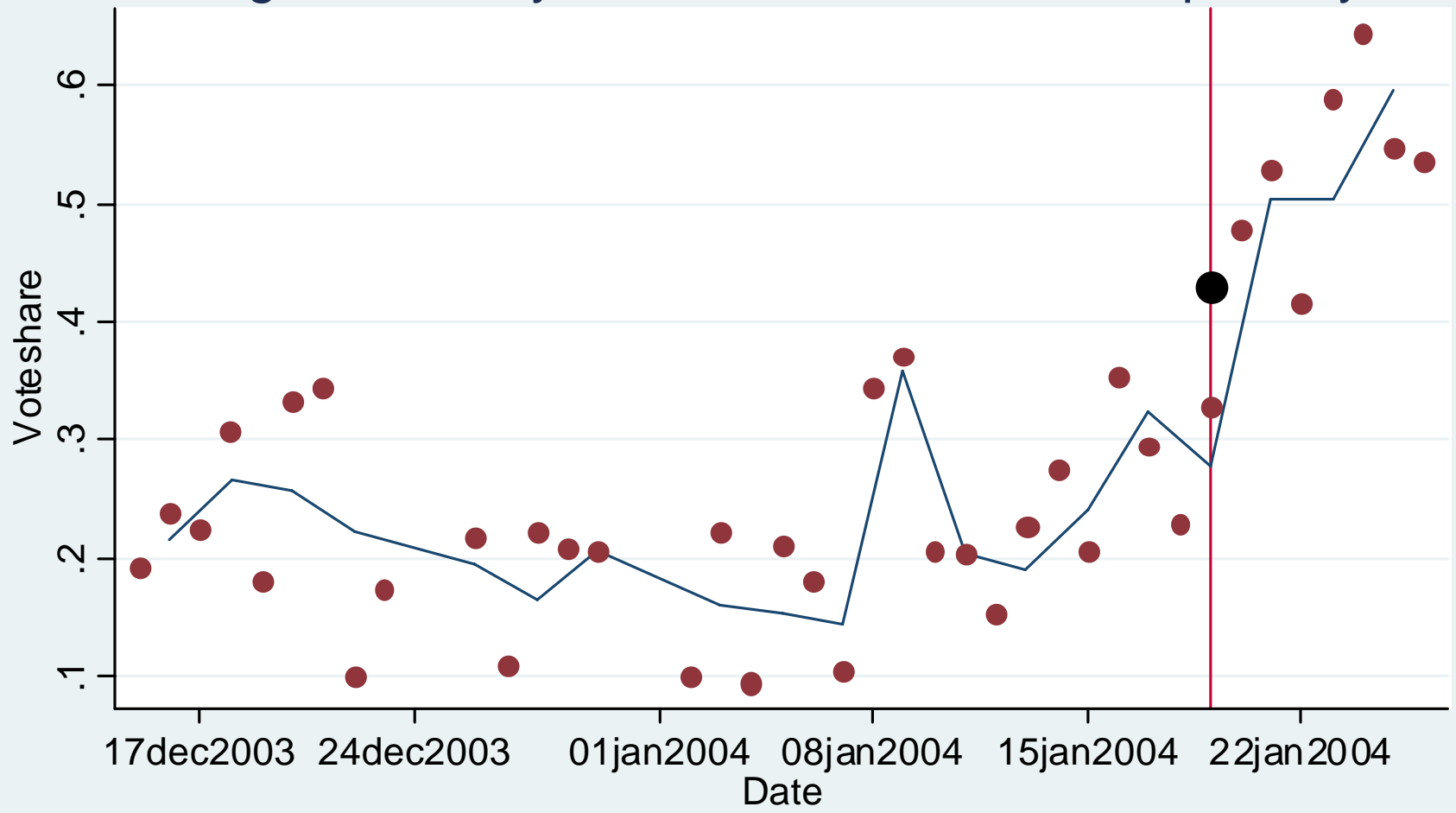Figure 4: Kerry before and after the Iowa primary

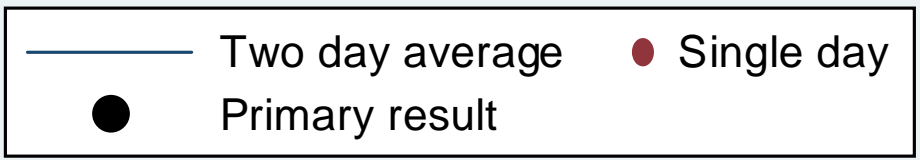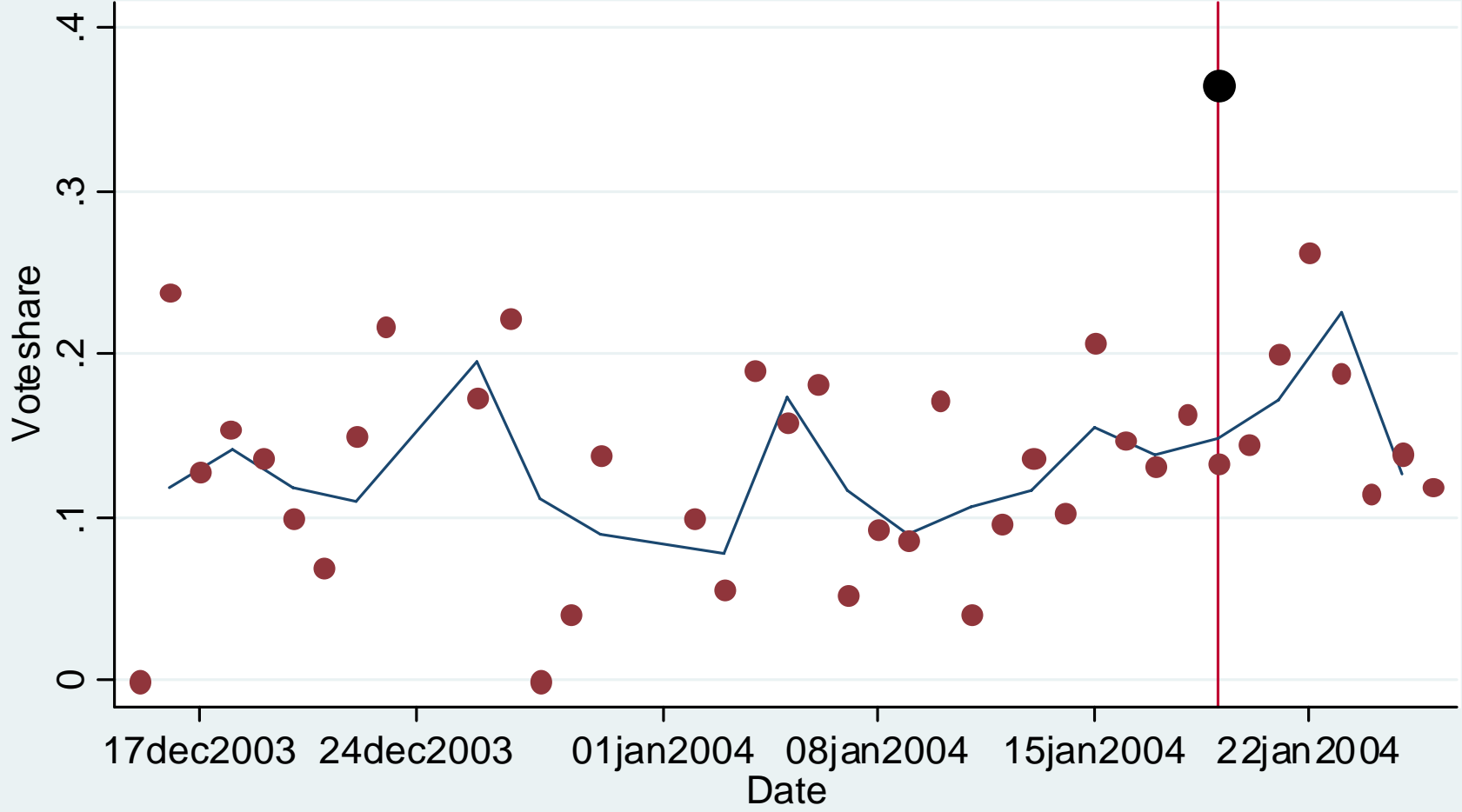Figure 5: Edwards before and after the Iowa primary

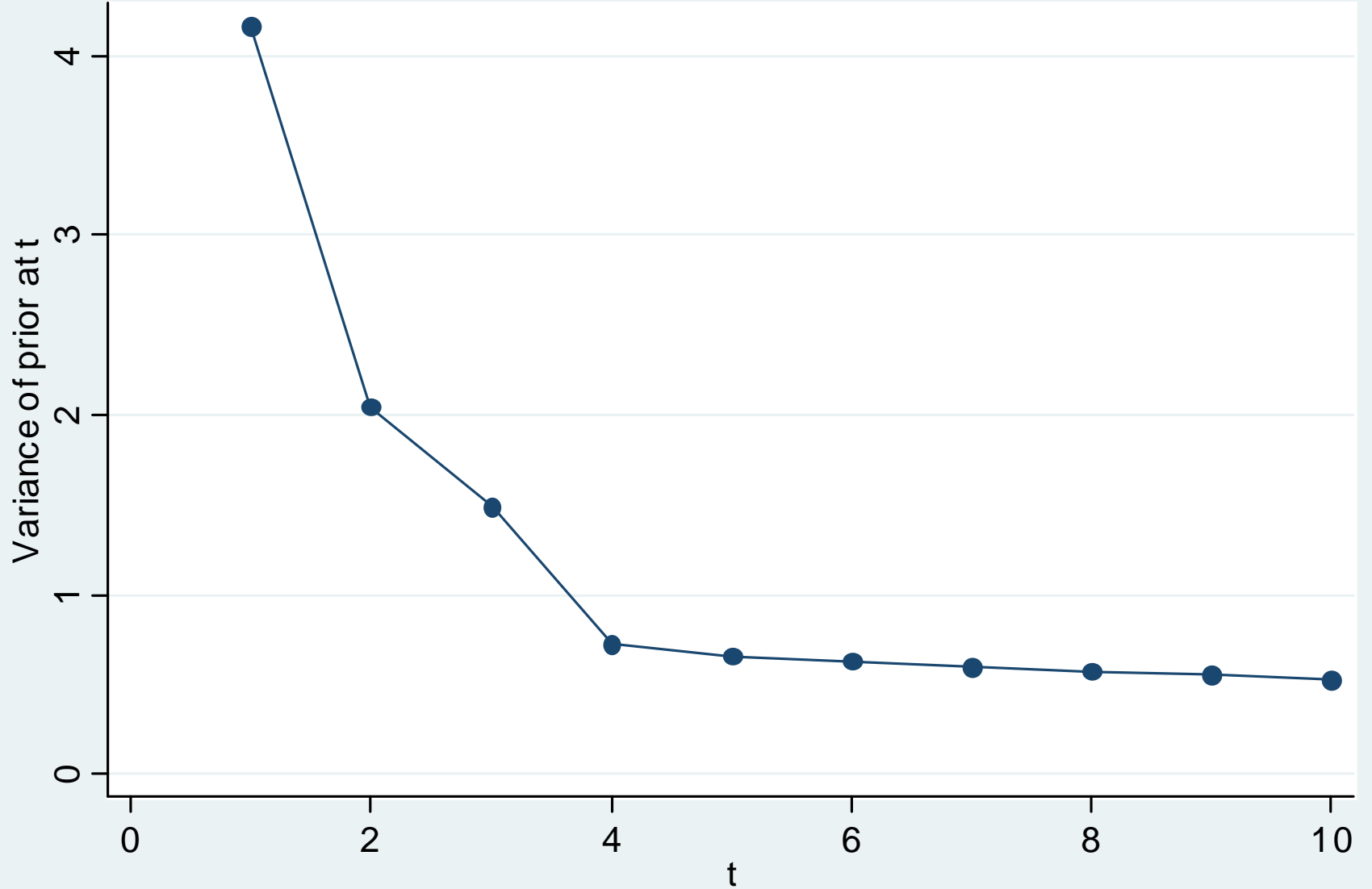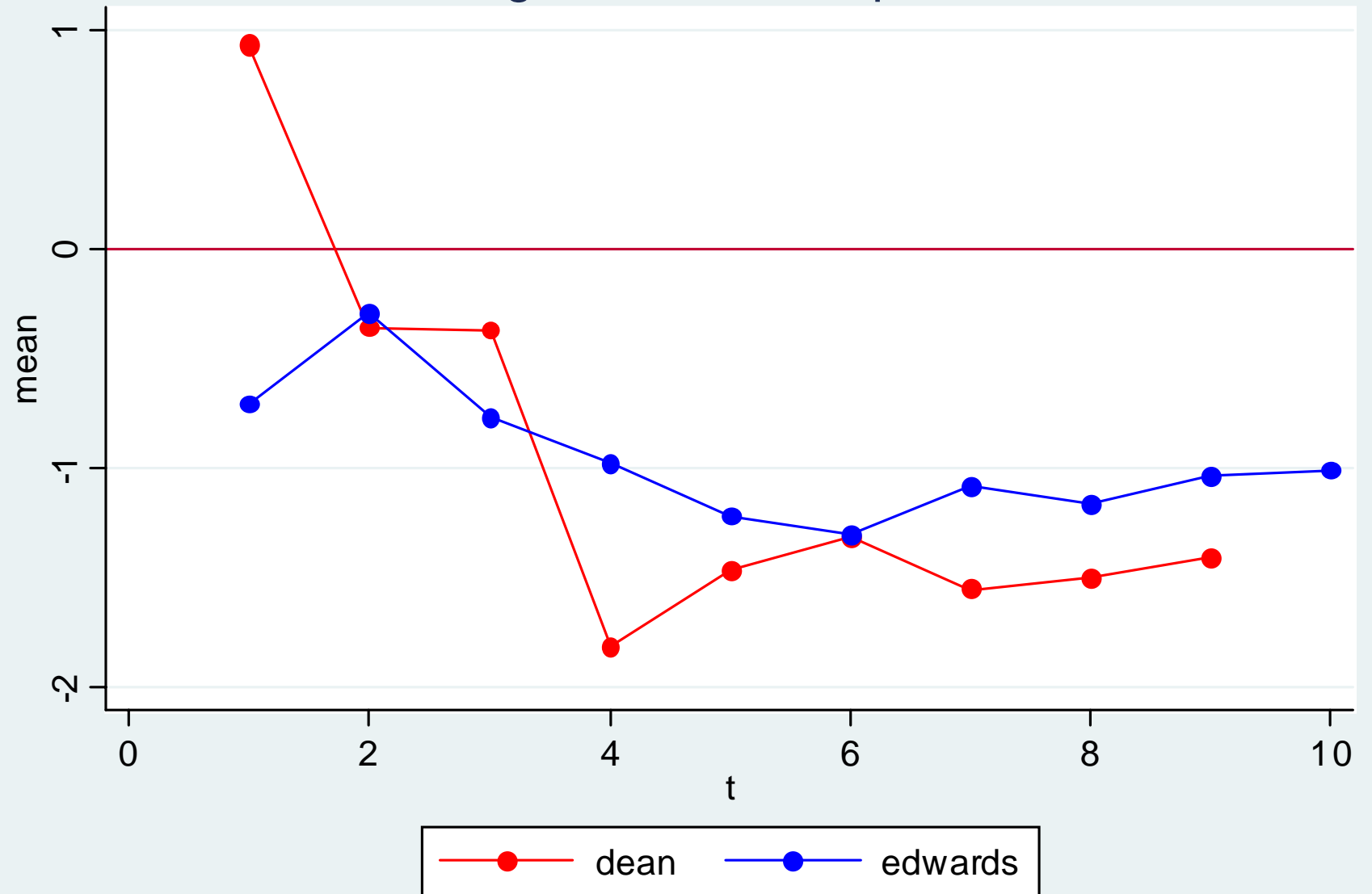Figure 6: Uncertainty over quality falls

Figure 7: Mean of priors
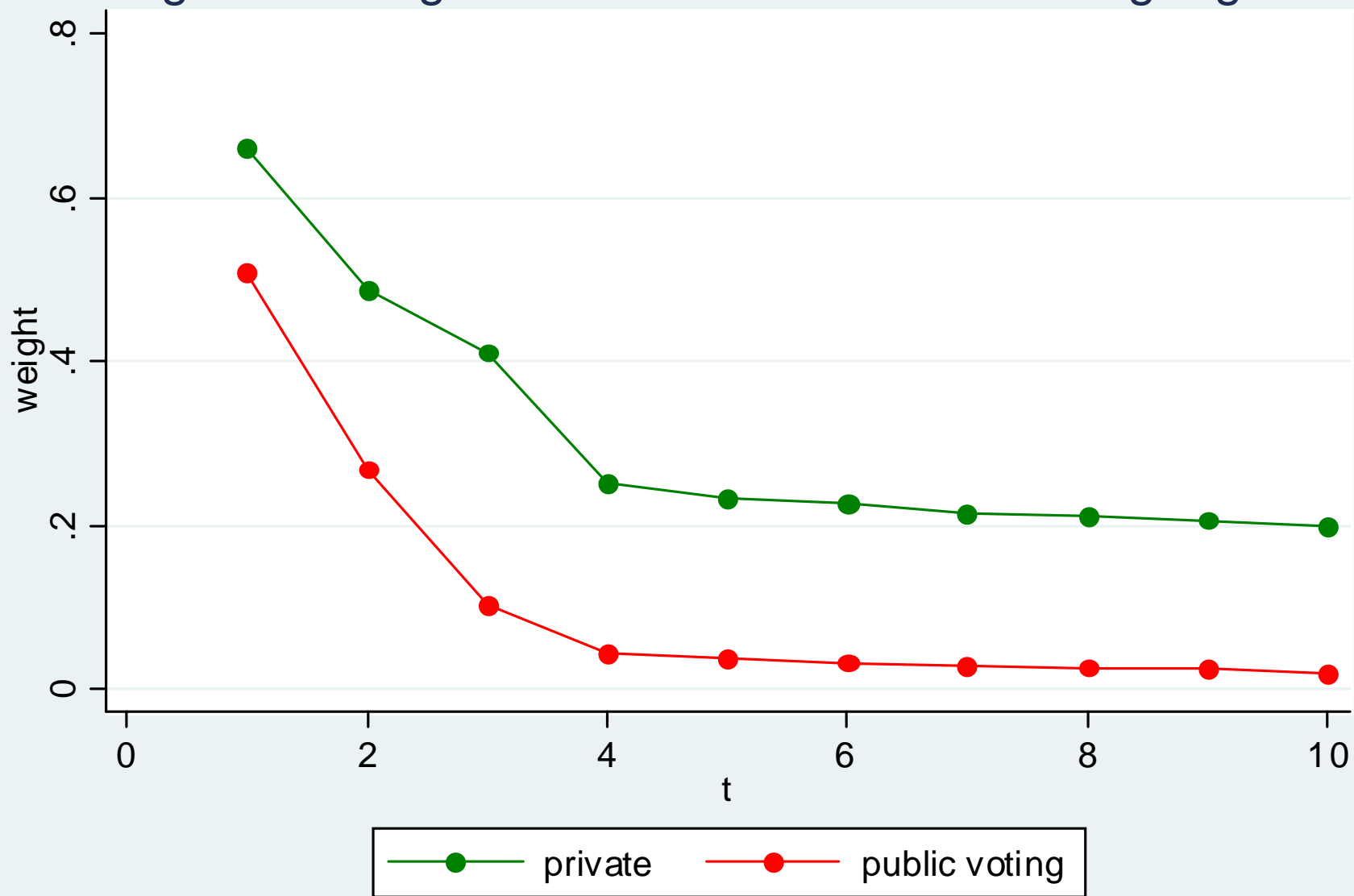
Figure 8: Weights on Private and Public Voting Signals

Figure 9: Social learning falls over time
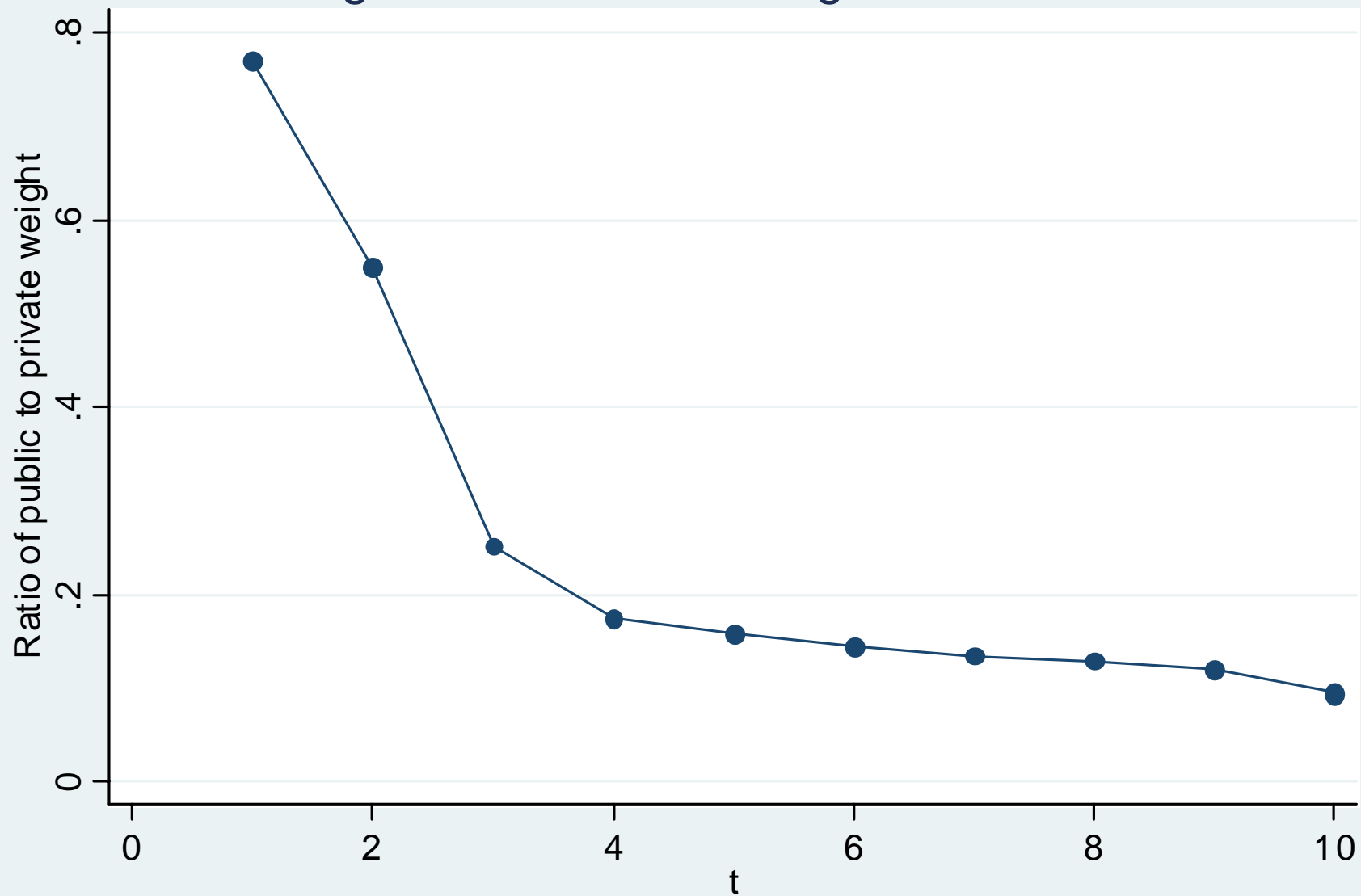
	Base Specification		Includes Distance		Includes Time Trend	
TABLE 1: First Stage Multinomial Logit						
	Dean	**Edwards**	**Dean**	**Edwards**	**Dean**	**Edwards**
Constant	0.938**	-0.701**	0.938**	-0.701**	1.404**	-0.32
	[0.773, 1.14]	[-0.913, -0.433]	[0.773, 1.14]	[-0.913, -0.433]	[1.108, 1.691]	[-0.761, 0.125]
AL	0.64	1.114	0.676	0.75	0.642	1.119
	[-0.427, 1.849]	[-0.784, 2.697]	[-0.38, 1.886]	[-1.272, 2.301]	[-0.382, 1.878]	[-0.826, 2.662]
AZ	0.169	-0.316	0.123	-0.496	0.188	-0.313
	[-0.716, 1.86]	[-1.665, 0.973]	[-0.742, 1.825]	[-1.849, 0.703]	[-0.716, 1.866]	[-1.695, 1.105]
CA	0.071	-0.235	0.015	-0.377	0.134	-0.195
	[-0.24, 0.425]	[-0.813, 0.316]	[-0.292, 0.376]	[-0.946, 0.194]	[-0.193, 0.495]	[-0.775, 0.37]
CO	-0.53	-0.737	-0.592	-0.881	-0.469	-0.744
	[-1.219, 0.611]	[-2.128, 0.559]	[-1.246, 0.53]	[-2.271, 0.432]	[-1.113, 0.612]	[-2.163, 0.591]
CT	-0.103	0.028	-0.03	0.654	0.052	0.152
	[-1.641, 1.345]	[-1.088, 2.113]	[-1.52, 1.414]	[-0.214, 2.47]	[-1.48, 1.534]	[-1.047, 2.162]
DE	-1.371*	0.749	-1.308*	1.075	-1.277	0.813
	[-2.832, 0.239]	[-0.863, 2.352]	[-2.777, 0.305]	[-0.667, 2.703]	[-2.804, 0.685]	[-0.742, 2.268]
FL	0.116	-0.141	0.195	-0.559	0.091	-0.177
	[-0.32, 0.871]	[-1.293, 0.718]	[-0.205, 0.939]	[-1.67, 0.36]	[-0.342, 0.861]	[-1.32, 0.696]
GA	0.332	0.835	0.405	0.444	0.315	0.797
	[-0.578, 1.475]	[-0.515, 1.971]	[-0.528, 1.538]	[-0.864, 1.525]	[-0.524, 1.47]	[-0.536, 1.954]
IA	0.014	0.348	-0.037	0.302	-0.054	0.313
	[-0.674, 1.146]	[-0.861, 1.606]	[-0.713, 1.095]	[-0.911, 1.581]	[-0.78, 1.16]	[-0.917, 1.572]
IL	0.23	-0.701	0.213	-0.837	0.258	-0.671
	[-0.286, 0.975]	[-2.29, 0.27]	[-0.297, 0.944]	[-2.395, 0.141]	[-0.254, 0.961]	[-2.295, 0.33]
IN	-0.325	-1.202*	-0.332	-1.31*	-0.266	-1.145*
	[-1.079, 0.446]	[-2.349, 0.169]	[-1.094, 0.433]	[-2.408, 0.021]	[-1.085, 0.542]	[-2.282, 0.213]
KY	-0.122	0.426	-0.111	0.165	-0.119	0.394
	[-1.332, 1.006]	[-1.049, 1.803]	[-1.321, 1.018]	[-1.215, 1.53]	[-1.342, 1.024]	[-1.095, 1.698]
LA	-0.095	0.449	-0.062	0.046	0.046	0.573
	[-1.255, 0.982]	[-0.873, 1.621]	[-1.228, 1.006]	[-1.268, 1.284]	[-1.136, 1.104]	[-0.678, 1.74]
MA	-1.346**	-2.1**	-1.195**	-1.184**	-1.27**	-2.066**
	[-1.885, -0.819]	[-3.299, -1.194]	[-1.706, -0.724]	[-2.181, -0.22]	[-1.799, -0.759]	[-3.267, -1.091]
MD	-0.195	0.066	-0.139	0.234	-0.158	0.11
	[-0.938, 0.789]	[-1.518, 0.95]	[-0.836, 0.838]	[-1.382, 1.086]	[-0.844, 0.965]	[-1.498, 0.994]
ME	0.25	0.004	0.197	0.604	0.233	-0.017
	[-0.803, 1.524]	[-1.18, 1.865]	[-0.861, 1.476]	[-0.367, 2.504]	[-0.869, 1.545]	[-1.167, 1.891]
MI	0.278	-0.244	0.204	-0.029	0.3	-0.285
	[-0.339, 1.363]	[-1.581, 0.975]	[-0.404, 1.288]	[-1.464, 1.142]	[-0.38, 1.335]	[-1.664, 0.985]
MN	-0.038	-0.188	-0.138	-0.06	-0.039	-0.19
	[-0.772, 0.803]	[-1.951, 0.795]	[-0.837, 0.72]	[-1.806, 0.945]	[-0.756, 0.807]	[-2.015, 0.859]
MO	0.421	-0.007	0.399	-0.197	0.419	0.027
	[-0.809, 2.074]	[-1.555, 1.986]	[-0.817, 2.055]	[-1.703, 1.811]	[-0.784, 2.107]	[-1.552, 2.058]
MS	0.588	0.697	0.619	0.245	0.498	0.687
	[-0.743, 1.594]	[-0.367, 1.763]	[-0.712, 1.622]	[-0.701, 1.409]	[-0.86, 1.669]	[-0.454, 1.736]
MT	-0.922	0.724	-0.995	0.742	-1.064	0.608
	[-2.347, 0.47]	[-0.706, 2.035]	[-2.43, 0.374]	[-0.627, 2.067]	[-2.526, 0.319]	[-0.787, 1.973]
NC	0.639	3.13**	0.714	2.717**	0.639	3.143**
	[-0.51, 2.133]	[2.005, 4.578]	[-0.415, 2.184]	[1.55, 4.179]	[-0.523, 2.154]	[2.017, 4.588]
NE	-0.764	-0.445	-0.82	-0.522	-0.839	-0.474
	[-2.817, 0.734]	[-1.667, 1.094]	[-2.868, 0.681]	[-1.762, 1.035]	[-2.73, 0.62]	[-1.578, 1.043]
NJ	-0.252	-0.491	-0.188	0.004	-0.208	-0.464
	[-0.838, 0.73]	[-1.688, 0.601]	[-0.747, 0.819]	[-1.352, 1.003]	[-0.832, 0.828]	[-1.689, 0.614]
NM	-0.018	-0.458	-0.047	-0.667	-0.042	-0.442
	[-0.921, 1.491]	[-1.246, 1.888]	[-0.95, 1.481]	[-1.465, 1.644]	[-1.096, 1.54]	[-1.291, 1.866]
NV	-0.368	-0.05	-0.422	-0.168	-0.377	-0.055
	[-2.069, 1.411]	[-1.189, 1.402]	[-2.146, 1.375]	[-1.286, 1.257]	[-1.976, 1.224]	[-1.11, 1.375]

	TABLE 1: First Stage Multinomial Logit					
	Base Specification		*Includes Distance*		*Includes Time Trend*	
	Dean	**Edwards**	**Dean**	**Edwards**	**Dean**	**Edwards**
NY	0.35	-0.911**	0.319	-0.391	0.397*	-0.868*
	[-0.078, 0.793]	[-2.593, -0.01]	[-0.116, 0.763]	[-1.84, 0.476]	[-0.025, 0.845]	[-2.553, 0.022]
OH	0.124	-0.094	0.122	-0.115	0.154	-0.065
	[-0.465, 0.955]	[-1.175, 0.885]	[-0.469, 0.951]	[-1.18, 0.876]	[-0.419, 0.989]	[-1.163, 0.89]
OK	-0.533	0.744	-0.541	0.431	-0.601	0.748
	[-2.321, 1.008]	[-0.656, 2.116]	[-2.328, 1.004]	[-0.888, 1.993]	[-2.44, 0.983]	[-0.63, 2.127]
OR	-0.127	-0.669	-0.194	-0.67	-0.2	-0.741
	[-0.963, 0.61]	[-2.099, 0.644]	[-1.055, 0.537]	[-2.099, 0.653]	[-1.021, 0.546]	[-2.102, 0.619]
PA	-0.231	-1.116**	-0.213	-0.776*	-0.204	-1.085**
	[-0.649, 0.419]	[-2.393, -0.117]	[-0.638, 0.426]	[-2.125, 0.164]	[-0.671, 0.458]	[-2.392, -0.096]
RI	-0.527	-0.438	-0.393	0.409	-0.577	-0.458
	[-1.642, 1.2]	[-1.536, 1.127]	[-1.494, 1.314]	[-0.715, 1.713]	[-1.751, 1.317]	[-1.699, 1.187]
SC	0.908	2.031**	0.971	1.527**	0.898	2.05**
	[-0.493, 1.986]	[0.705, 3.389]	[-0.434, 2.084]	[0.244, 2.873]	[-0.444, 2.015]	[0.805, 3.357]
TN	-0.115	-0.074	-0.073	-0.368	-0.069	-0.042
	[-0.833, 1.216]	[-1.434, 1.773]	[-0.81, 1.257]	[-1.942, 1.273]	[-0.806, 1.297]	[-1.416, 1.83]
TX	0.034	0.493	0.035	0.152	0.019	0.462
	[-0.408, 0.663]	[-0.319, 1.372]	[-0.399, 0.665]	[-0.645, 0.987]	[-0.43, 0.673]	[-0.333, 1.309]
UT	0.433	0.249	0.398	0.118	0.325	0.172
	[-0.815, 1.35]	[-0.866, 1.388]	[-0.876, 1.309]	[-0.981, 1.322]	[-0.979, 1.331]	[-0.871, 1.394]
VA	0.387	0.221	0.456	0.144	0.476	0.245
	[-0.576, 1.175]	[-1.397, 1.439]	[-0.537, 1.214]	[-1.493, 1.343]	[-0.495, 1.208]	[-1.313, 1.486]
WA	0.177	-0.287	0.082	-0.226	0.187	-0.284
	[-0.408, 0.618]	[-1.343, 0.681]	[-0.51, 0.533]	[-1.29, 0.734]	[-0.383, 0.657]	[-1.302, 0.726]
WI	0.207	-0.028	0.146	0.109	0.262	0.031
	[-0.769, 1.083]	[-1.35, 1.122]	[-0.838, 0.975]	[-1.21, 1.23]	[-0.719, 1.091]	[-1.295, 1.102]
WV	0.078	0.008	0.102	-0.048	0.038	-0.063
	[-1.133, 1.659]	[-1.322, 1.25]	[-1.108, 1.698]	[-1.336, 1.227]	[-1.138, 1.631]	[-1.333, 1.19]
Distance			-0.062**	-0.062**		
			[-0.103, -0.025]	[-0.103, -0.025]		
Trend					0.01**	0.008**
					[0.005, 0.015]	[0, 0.018]
	[bootstrap 95% confidence interval], ** significant at 5%, * significant at 10%					

	Table 2: Estimated Parameters		
	Base Specification	*Includes Distance*	*Includes Time Trend*
σ_η^2	0.815**	0.707**	0.829**
	[0.551, 1.194]	[0.402, 1.05]	[0.546, 1.192]
σ_1^2	4.162**	6.074**	5.924**
	[2.41, 8.295]	[2.85, 23.03]	[3.416, 16.789]
σ_ε^2	2.147**	2.653**	3.988**
	[0.491, 7.058]	[0.687, 9.818]	[1.619, 9.46]
	95% confidence interval in brackets		

Table 3: Favorability Regressions

	Favorability	Cares About People Like Me	Inspiring	Strong Leader	Trustworthy	Shares My Values	Knowledge-able	Reckless
Mean Candidate Quality	1.132**	0.470**	0.751**	0.674**	0.366**	0.605**	0.116	-0.135
	[0.052]	[0.136]	[0.164]	[0.153]	[0.133]	[0.136]	[0.171]	[0.196]
Dean	-0.588**	-0.583**	-0.598**	-0.429**	-0.494**	-0.706**	-0.194	1.017**
	[0.070]	[0.189]	[0.199]	[0.187]	[0.179]	[0.197]	[0.246]	[0.294]
Constant	0.154**	0.381**	0.488**	-0.195	0.025	0.304*	-0.801**	-0.246
	[0.057]	[0.160]	[0.188]	[0.175]	[0.155]	[0.161]	[0.208]	[0.236]
Observations	6374	965	991	972	962	954	488	479
R-squared	0.085	0.018	0.029	0.027	0.013	0.027	0.002	0.028

Robust standard errors in brackets
* significant at 10%; ** significant at 5%

			Sequential Primary			Simultaneous Primary		
Table 4: Counterfactual Primary								
State	Period	Date	Dean	Edwards	Kerry	Dean	Edwards	Kerry
IA	1	1/19/2004	21%	36%	43%	21%	36%	43%
NH	2	1/27/2004	34%	16%	50%	36%	14%	50%
AZ	3	2/3/2004	22%	11%	67%	25%	6%	69%
DE	3	2/3/2004	14%	15%	70%	29%	5%	66%
MO	3	2/3/2004	10%	29%	60%	7%	28%	65%
NM	3	2/3/2004	23%	16%	60%	31%	12%	57%
OK	3	2/3/2004	7%	49%	44%	8%	46%	46%
SC	3	2/3/2004	6%	56%	38%	4%	41%	55%
MI	4	2/7/2004	20%	16%	63%	66%	5%	29%
WA	4	2/7/2004	35%	8%	57%	94%	0%	6%
ME	5	2/8/2004	33%	10%	57%	86%	1%	13%
TN	6	2/10/2004	6%	37%	57%	2%	76%	22%
VA	6	2/10/2004	8%	31%	61%	3%	49%	48%
NV	7	2/14/2004	19%	12%	70%	76%	1%	24%
WI	8	2/17/2004	20%	37%	43%	38%	54%	8%
UT	9	2/24/2004		35%	65%		39%	61%
CA	10	3/2/2004		23%	77%		22%	78%
CT	10	3/2/2004		29%	71%		28%	72%
GA	10	3/2/2004		47%	53%		45%	55%
MA	10	3/2/2004		20%	80%		91%	9%
MD	10	3/2/2004		30%	70%		30%	70%
MN	10	3/2/2004		35%	65%		61%	39%
NY	10	3/2/2004		25%	75%		64%	36%
OH	10	3/2/2004		40%	60%		72%	28%
RI	10	3/2/2004		21%	79%		22%	78%