November 2007

Momentum and Social Learning in Presidential Primaries *

Abstract

This paper provides an investigation of the role of momentum and social learning in sequential voting systems. In the econometric model, voters are uncertain over candidate quality, and voters in late states attempt to infer the information held by those in early states from voting returns. Candidates experience momentum effects when their performance in early states exceeds expectations. The empirical application focuses on the responses of daily polling data to the release of voting returns in the 2004 presidential primary. We find that Kerry benefited from surprising wins in early states and took votes away from Dean, who held a strong lead prior to the beginning of the primary season. The voting weights implied by the estimated model demonstrate that early voters have up to 20 times the influence of late voters in the selection of candidates, demonstrating a significant departure from the ideal of "one person, one vote." We then address several alternative, non-learning explanations for our results. Finally, we run simulations under different electoral structures and find that a simultaneous election would have been more competitive due to the absence of herding and that alternative sequential structures would have yielded different outcomes.

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

*The authors received helpful comments from Nageeb Ali, Andy Foster, Alan Gerber, Rob McMillan and seminar participants at Yale, Johns Hopkins, University of Pennsylvania (Wharton), Northwestern, Toronto, Brown, and the NBER political economy meetings.

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 19th, followed by the New Hampshire primary on January 27th, and then mini-Super Tuesday on February 3rd, when voting occurred in the states of Arizona, Delaware, Missouri, New Mexico, North Dakota, Oklahoma, and South Carolina. The primary season continued with various elections in March, April, and May before concluding with the Montana and New Jersey primaries on June 8th. As shown in Figure 2, the 2008 schedule is becoming increasingly front-loaded with Nevada scheduled between Iowa and New Hampshire and, perhaps more importantly, many states moving their primaries to February 5th. 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. That is, do outcomes of primaries depend upon the sequencing of states? 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 unmeasured candidate characteristics? 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 underlying candidate strength, which is often unobserved by the econometrician. 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. A similar question is 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 conditioned their decisions on voting returns in Iowa?

In this paper we attempt to overcome these econometric challenges through the development of 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 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 returns 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 momentum effects. Using the estimated model, we examine how implied voting weights differ depending upon location in the voting sequence. We next address several alternative, non-learning explanations for our momentum results. Finally, we use the model to simulate electoral outcomes under a counterfactual simultaneous election and also alternative primary calendars.

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]

A key question is whether these social learning results extend to the context of sequential elections, with the main distinction being that voters make a social choice, and individual payoffs thus depend upon the actions of all agents. Under strategic voting, rational agents recognize that their individual action only matters if they are pivotal, defined as situations in which their vote changes the voting outcome. Feddersen and Pesendorfer (1997) first address this issue in the context of a model with a binary, symmetric, and simultaneous election. Given that pivotal

[1] This social learning framework has been applied in a variety of empirical 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). Cai, Chen, and Fang (2007) conduct a field experiment in which the top selling dishes were posted in restaurant menus and find that these postings are influential for orders and especially so for infrequent customers. Finally, Glaeser and Sacerdote (2007) provide a social learning explanation for aggregation reversals, where an individual relationship, such as income and ideology, is reversed at some level of aggregation, such as the state-level. For a more comprehensive overview of the social learning literature, see the survey by Sushil, Hirshleifer, and Welch (1998) and Chamley (2004).

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 under binary and symmetric environments and show that every equilibrium of the simultaneous game is also an equilibrium of the sequential game, regardless of the sequence. Strategic voters condition on being pivotal and hence 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 equilibria 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 late votes, breaking the symmetry underlying the Dekel and Piccione (2000) result. These posterior-based strategies can be interpreted as sequential analogues to sincere voting under simultaneous elections, providing support for our sincerity assumption to be described below.[3]

In addition to social learning, several authors have suggested alternative models for momentum, both at the voter and candidate level. Callander (2007) 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

[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.

[3] In related work, McKelvey and Ordeshook (1985) show that momentum effects can be generated even under simultaneous elections if polling data or endorsements are released during the period leading up to the election.

elections while the loser will have a diminished incentive. Through this asymmetry of campaign spending, momentum is generated and can lead 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. He also suggests that candidate Gary Hart's surprising early victories convinced later voters of his viability. 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.[4] 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 the 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

[4] 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.

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 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, endorsements of candidates by either local media outlets or local politicians, 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.[5]

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.[6] We thus abstract from several forms of strategic voting under which optimal voter behavior may depend upon the behavior of other voters. Importantly, these forms of strategic voting can also generate momentum effects that are unrelated to social learning, and this issue of alternative explanations for any measured momentum effects 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:

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

[5] 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.

[6] As noted above, these strategies are similar to those used in Ali and Kartik (2007).

Reflecting well-known results in the literature on Bayesian learning, 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 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 over quality upon observing vote shares in early states (i.e. $E(q_c|v_{cst}, v_{0st})$). Given that state-level 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:

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

where the noise in the voting signal includes the noise in the quality signal (ε_{cs}) but also the noise due to the unobserved state preferences (η_{ct}/α_t); the combined variance of the noise in the voting signal thus equals $(\sigma_\eta^2/\alpha_t^2) + \sigma_\varepsilon^2$. 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 voting signals 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).

An important implicit assumption in the above formulation is that expectations over electoral outcomes, as captured by μ_{ct} in equation 11, depend upon national, but not state-specific, factors. We make this assumption for two reasons. First, national polls reveal national preferences (μ_{ct}), while state-specific polls reveal both national and state-level preferences $(\mu_{ct} + \eta_{cs})$. Thus, with both types of polls, voters can uncover state-specific preferences (η_{cs}), and this inference would violate our assumption that voters cannot observe state-level preferences in other states. If voters can learn state-specific preferences, then the signal can be uncovered from voting returns in equation 6, and the key social learning parameter in this case would be given by $(\sigma_t^2+\sigma_\varepsilon^2)/(N_t\sigma_t^2+\sigma_\varepsilon^2)$ and equals one for single-state primary days ($N_t=1$), such as Iowa and New Hampshire. Thus, social learning would be assumed, rather than measured, in our empirical application to follow. Second, while state-specific polling data was often reported in the media for Iowa and New Hampshire, polls in other states were reported far less frequently, if at all, and it is thus far from clear that voters had this information for all states. National polls, by contrast, were readily available on a high-frequency basis. Finally, we should note that this assumption (expectations depend purely on national factors) is much stronger than we need, and we relax this assumption in one of the alternative specifications, which assumes that some state-specific factors are observed and that others are unobserved. What is crucial to our result is that state-level preferences are not perfectly observed by voters in other states.

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 and Oklahoma. 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.[7] Finally, as will be described below, we aim to estimate the state-specific preference parameters (η_{cs}) and thus insufficient data forces us to 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 for the case of Iowa. As shown, Dean had a substantial and stable lead over Kerry and Edwards during the month preceding the Iowa primary. Dean under-performed in Iowa relative to expectations, as captured by pre-Iowa polling levels, and voters in the Annenberg survey updated appropriately.[8] 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.[9]

[7] 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.

[8] This reaction in polling data to the Iowa outcome is somewhat confounded by Dean's reaction, which was dubbed the "Dean scream" and was televised extensively in the days following the Iowa outcome. It is important to note, however, that this media coverage would probably not have occurred had Dean fared better in Iowa. That is, the Iowa outcome and Dean's reaction to that outcome are not necessarily independent events. In addition, the votes that Dean lost went to Kerry and Dean, the winners in Iowa, rather than the losers, including Gephardt and Lieberman. Thus, even if Dean's vote loss was due to Dean's reaction, the reallocation of those votes is consistent with our story of momentum associated with social learning from aggregate returns.

[9] These patterns are similar to those in prices from the Iowa Electronic Market, in which market participants purchased contracts that pay $1 in the event that Kerry, for example, is the party's nominee in the general election, and the price of this contract can thus be interpreted as the probability that a given candidate wins the nomination (Wolfers and Zitzewitz, 2004). We choose to focus on polling data, rather than these prices from prediction markets, for two reasons. First, the mapping from voting in primaries to the probability of nomination, as provided by the prediction market prices, is confounded by the presence of superdelegates, which are not pledged to the winning candidate, as well as the possibility that no candidate wins a majority of the delegates, in which case the nominee is chosen through a bargaining process at the party convention. Second, the daily polling data, but not prediction market data, include additional measures of candidate quality, and we will make use of these in our discussion of alternative explanations in section 4.6.

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:

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

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.[10]

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.

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 for 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

[10] 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.

in an alternative specification, which relaxes the assumption that such advantages are unobserved by voters in later states.

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 and the aggregate returns. 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 around 3.5, 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 weights placed 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 75 percent weight on their priors and only 25 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 60 percent weight on these signals and 40 percent on 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 momentum effects as reflected in the effect of early returns on the choices of late voters.

4.4 Implied Voting Weights

Due to these documented momentum effects, early voters have a disproportionate influence over the selection of candidates. This over-weighting of early voters associated with sequential voting thus leads to potential deviations from the democratic ideal of "one person, one vote." While this property of sequential voting has been frequently discussed in policy debates over the design of the primary system, there is little evidence on the degree of this disproportionate influence. Interestingly, we can use the estimated model to explicitly calculate the voting weights associated with sequential voting in the 2004 primary.

Our first measure of voting weights is based upon the effect of changes in state-level preferences (η_{cs}) on candidate vote shares:

$$\frac{\Delta v_c}{\Delta \eta_{cs}} \tag{19}$$

where v_c is the vote-share for candidate c averaged across all states. In the absence of momentum effects, the impact of a change in voter preferences should be equal across all states. With momentum effects, by contrast, changes in voter preferences in early states will lead to changes in vote shares in that period but in all subsequent states as well. We compute this derivative for one state voting on each of the primary dates ($t = 1, 2, .., 10$) and, normalizing the influence of the final period to 1, this derivative provides a measure of the weight placed upon state-level preferences.[11] As shown in Figure 10, preferences of voters from the state of Iowa, the first state to vote, have almost six times the influence of the states voting on Super Tuesday ($t = 10$).[12]

Our second measure is based upon the effect of changes in state-level information (θ_{cs}) on candidate vote shares:

[11] In particular, we increase state-level preferences by one unit and re-compute vote shares for that state, as expressed in equation 6. In order to predict vote shares for subsequent states, we re-compute the posterior mean quality, as expressed in equation 8, and ultimately vote shares, as expressed in equation 6. Note that simulating these vote shares requires explicit measures of the voting signal (θ_{cs}), which can be backed out of equation 6 with information over state-level preferences (η_{cs}).

[12] A shock to a state in period t+1 may have a larger impact on overall vote share than a shock to a state in period t, despite having a shorter duration. This is a result of the non-linear relationship between votes shares and shocks, making the overall impact of a shock depend upon the pre-shock vote share. If we calculate weights using the log vote ratio instead of vote share we find that the weight is monotonically decreasing in t.

$$\frac{\Delta v_c}{\Delta \theta_{cs}} \tag{20}$$

As shown in Figure 11, the information held by Iowa voters has roughly 20 times the influence of information held by Super Tuesday voters. These information-based weights are substanially larger than the preference-based weights, as described above, given that voters in late states place less weight on their own signal. This under-weighting of signals has a direct effect in the calculation of voting weights but also has an indirect effect as late vote shares are thus a noisier signal of quality. Taken together, these results confirm the often-held notion that early states have a disproportionate influence over the selection of candidates in sequential primary systems and thus represents a significant departure from “one person, one vote.”

4.5 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, as noted above, 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 a 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:

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

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:

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

Thus, voters in late states incorporate these observed state-level 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 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:

$$\ln(v_{cst}/v_{0st}) = \eta_{cs} + \gamma_c t + \alpha_t \theta_{cs} + (1-\alpha_t)\mu_{ct}. \tag{23}$$

where 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:

$$\mu_{ct+1} - \mu_{ct} = \frac{\beta_t/N_t}{\alpha_t} \sum_{s\in\Omega_t} [\ln(v_{cst}/v_{0st}) - \gamma_c t - \mu_{ct}] \tag{24}$$

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.

4.6 Alternative Explanations

In the baseline model, we have assumed sincere voting, under which voters support the candidate that provides the highest expected utility level. We have thus abstracted from strategic voting, several forms of which provide alternative, non-learning explanations for our documented momentum results. The first form of strategic voting involves electability considerations associated with the general election. For example, consider a voter who prefers Dean over Kerry as president but prefers either over Bush. This voter may learn that Kerry is more popular among other voters after Iowa and thus has a better chance of defeating Bush in the general election; this voter may thus switch to Kerry after Iowa, and this electabilty-driven switching provides an alternative explanation for our results.[13] The second form of strategic voting involves coordination and concerns over wasted votes with more than two candidates. Consider a voter, for example, who ranks the candidates as Edwards first, Dean second, and Kerry last. This voter may view Edwards as not viable before Iowa and support Dean instead. After Iowa, Edwards becomes viable and this voter may shift from Dean to Edwards. Again, this viability-driven switching provides an alternative explanation for our results. The third form of strategic voting involves bandwagon effects, under which voters have conforming preferences and thus vote for the candidate expected to win. Again, bandwagon effects may mimic social learning and thus provide an alternative explanation for our results.

Importantly, all three of these alternative explanations for our measured momentum effects involve learning about the *preferences of other voters* rather than underlying candidate quality, as is emphasized by our social learning model. Thus, to distinguish between strategic voting explanations and social learning explanations, we examine the dynamics of measures of candidate quality, which we proxy by auxiliary questions in which voters evaluated candidates on a 1-10 scale for the following candidate characteristics: favorability, cares about people like me, inspiring, strong leader, trustworthy, shares my values, knowledgeable, and reckless. These characteristics can be interpreted as measures of candidate quality given that they are arguably traits that would be valued by all voters regardless of ideology.

[13] Note that electability considerations are not necessarily inconsistent with our results if primary voters believe that the probability of winning in the general election is increasing in candidate quality. This could be the case, for example, if independent voters are pivotal in the general election and place significant weight on quality. Then primary voters driven by electability considerations will still value quality although the rationale for this preference is somewhat different.

More concretely, for each of these quality proxies, we run the following regression:

$$\text{quality}_{itc} = \delta_c + \kappa \times \mu_{ct} + \xi_{itc} \tag{25}$$

where μ_{ct} is the mean candidate quality at time t as implied by our estimated model and reflected in figure 7 and quality_{itc} is measured relative to Kerry. Under our assumption of sincere voting, we would expect $\kappa > 0$, whereas, as argued above, there should be no link ($\kappa = 0$) under strategic voting explanations given that voters only learn about the preferences of other voters from early returns.

As shown in Table 3, there is a strong link between these factors, providing support for our social learning story. The first six measures have the expected positive coefficients under a social learning story, whereas the coefficient associated with the measure of "reckless" has the expected negative sign. The final two measures, however, are statistically insignificant, likely reflecting the reduced sample sizes.[14] One important caveat of this analysis is that it does not rule out these alternative explanations if some voters act in a sincere manner whereas others act strategically.[15]

Even if other motives are present, however, this analysis does provide strong support for the presence of social learning among at least a subset of voters.

5 Counterfactual simulations

In order to further highlight the importance of momentum and social learning in sequential elections, we next provide two counterfactual simulations: simultaneous voting and alternative ordering of states under a sequential system.

5.1 Simultaneous primary

We first consider an election in which every state votes in a simultaneous national primary on January 19, 2004. In this case, vote shares in state s can be summarized as follows:

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

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

[15] Similarly, it could be that a given voter has a mix of motives and may care about both candidate quality and, for example, bandwagon effects.

Accordingly, behavior in states voting after Iowa may be altered, relative to the sequential voting returns, for two reasons. First, all voters use the pre-Iowa prior (μ_{c1}). Given that voter priors favored Howard Dean in the days leading up to Iowa, we thus 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 ($\alpha_1 > \alpha_t$ for $t > 1$). Thus, signals will be amplified in a simultaneous primary, and this second effect could benefit any of the three candidates depending upon the distribution of the realized signals.

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 cannot calculate counterfactual Dean vote shares for states thereafter.[16] We thus run two counterfactual simultaneous primaries: one in which Dean is included but in which states after Wisconsin do not vote and one in which Dean is excluded but all states vote. As shown in the Table 4, the results from the counterfactual three-candidate simultaneous election demonstrate that the election would have been much closer, with Dean winning in Michigan, Washington, Maine, and Nevada and Edwards winning in Tennessee and Wisconsin. While Kerry would have won a plurality with 39 percent of delegates, he did not win a majority and thus the eventual nominee would have been decided at the convention. Similarly, the two-candidate simultaneous primary would have been much closer with Edwards winning four states 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 defeated Kerry in Massachusetts.[17] Rather, we hope that the heightened competition under a counterfactual simultaneous election helps to further reinforce our findings of herding in favor of Kerry under the sequential primary system.

5.2 Alternative sequential schedules

[16] 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 measure the signals for these states. 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.

[17] 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.

Our second counterfactual election involves changes in the voting order under a sequential schedule. To the extent that herding occurs under a sequential election, then the voting outcome may be fragile, or sensitive to the order of voting. To investigate this issue, we randomly generated alternative voting sequences, holding constant the number of states voting on each date. Again, we consider a two-candidate election, in which all states are included, and a three-candidate election, in which only states voting prior to and including Wisconsin are included. As shown in Table 5, Kerry continues to win a plurality of states in most cases under both the sequential two-candidate and sequential three-candidate elections. Using delegate weights, however, the counterfactual sequential elections are somewhat more competitive, with Edwards winning 11 percent of the two-candidate sequential election schedules. Again, while Kerry still wins most of these counterfactual elections, there are a sizeable number of cases in which Edwards would have won. This is surprising given the wide margin by which Kerry won the actual election and highlights the sensitivity of electoral outcomes to the sequencing of states.

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 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 depends 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 largely ignored by voters in the latest states. Finally, we simulate the election under a number of counterfactual primary systems and show that the race would have been much tighter under a simultaneous system and that electoral outcomes are sensitive to the order of voting. While these results are specific to the 2004 primary, we feel that they are informative more generally in the debate over the design of electoral systems in the United States and elsewhere.

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.

H. Cai, Y. Chen, and H. Fang. Observational Learning: Evidence from a Randomized Natural Field Experiment. *working paper, Yale University*, 2007.

S. Callander. Bandwagons and Momentum in Sequential Voting. *Review of Economic Studies*, 74(3):653–684, 2007.

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.

E.L. Glaeser and B. Sacerdote. Aggregation Reversals and the Social Formation of Beliefs. *NBER Working Paper*, 2007.

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

R.D. McKelvey and P.C. Ordeshook. Elections with limited information: a fulfilled expectations model using contemporaneous poll and endorsement data as information sources. *Journal of Economic Theory*, 36(1):55–85, 1985.

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.

J. Wolfers and E. Zitzewitz. Prediction Markets. *The Journal of Economic Perspectives*, 18(2):107–126, 2004.

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

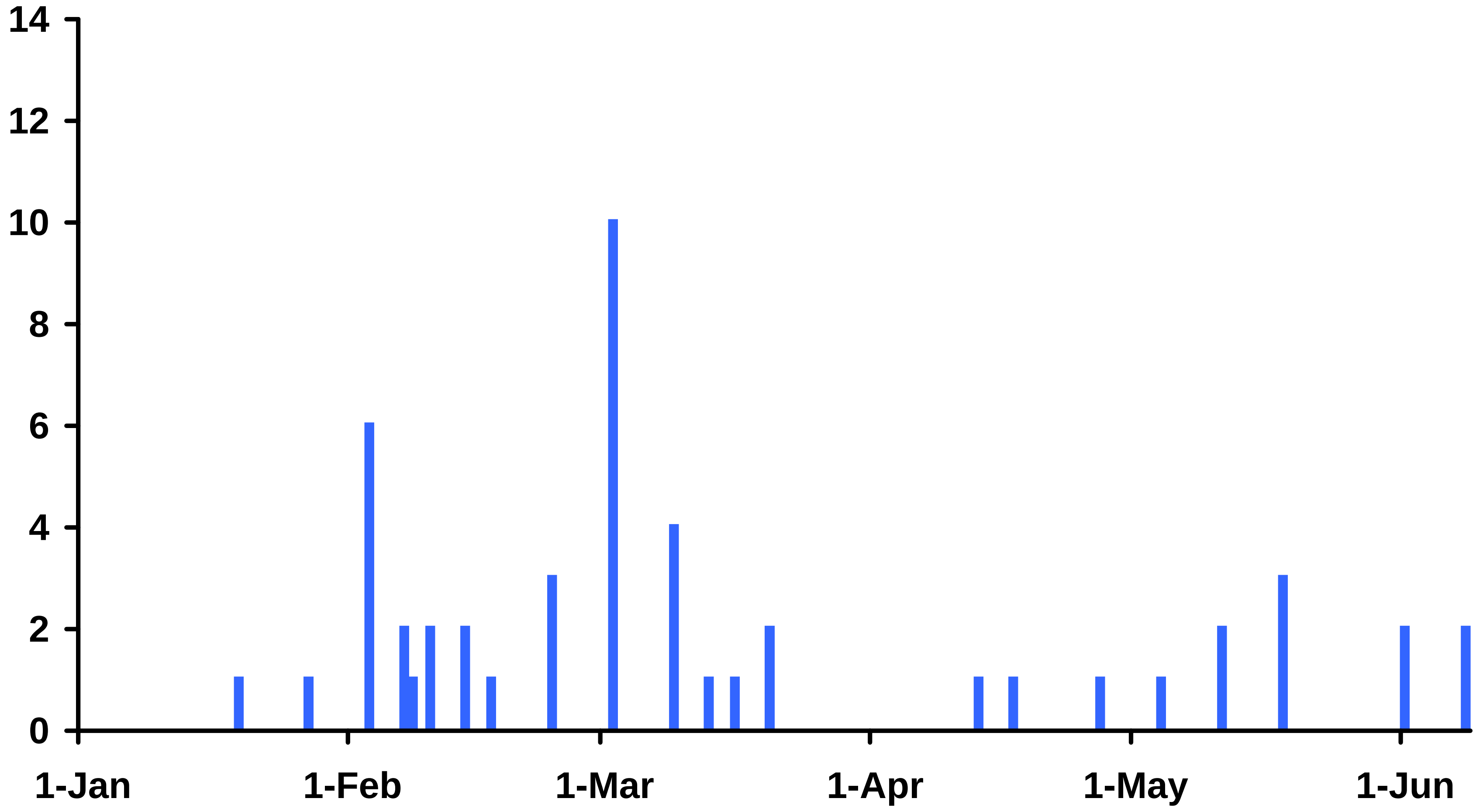
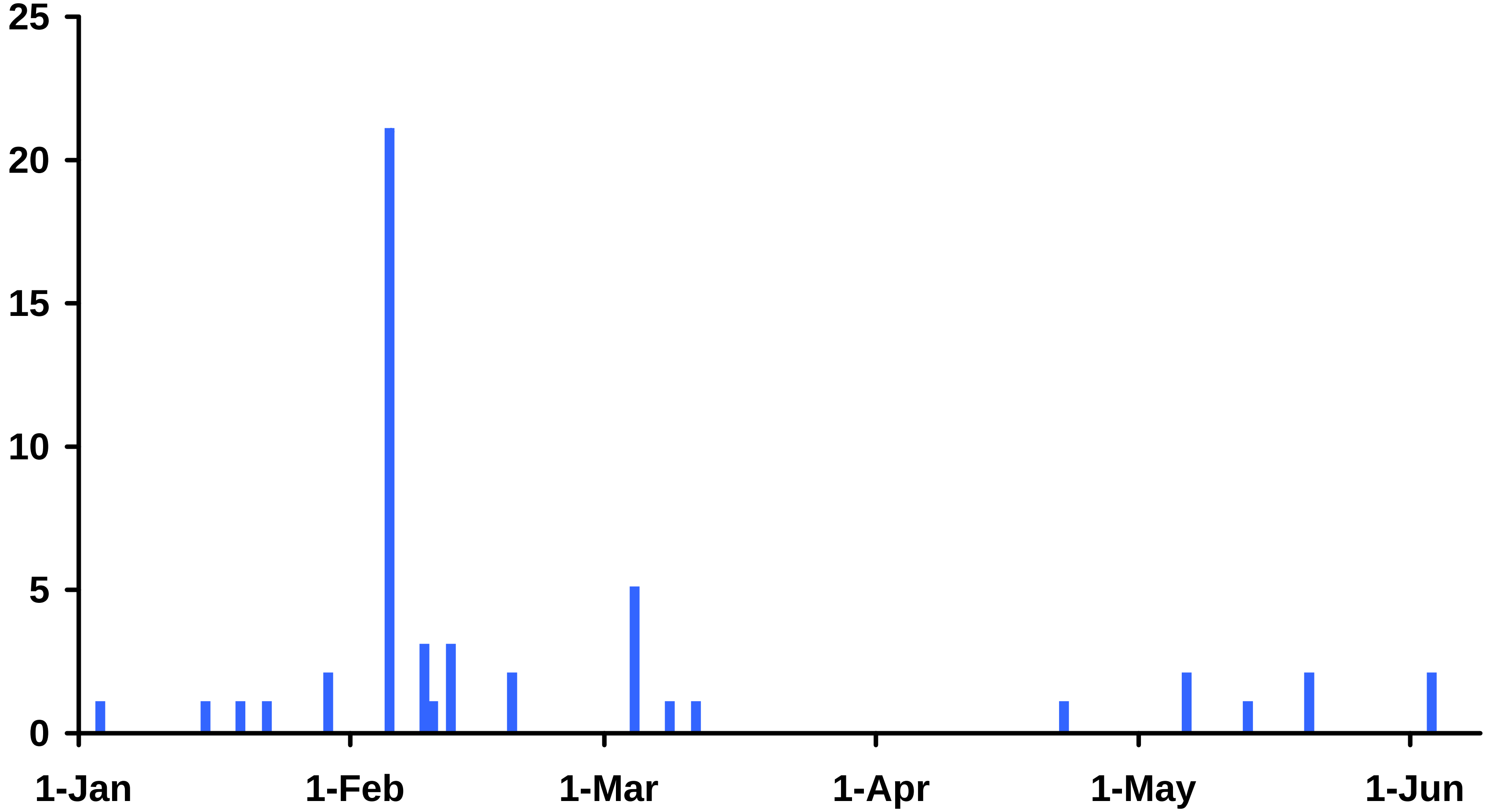


Figure 2: Number of Primaries by Date
2008 Democratic Primary Season



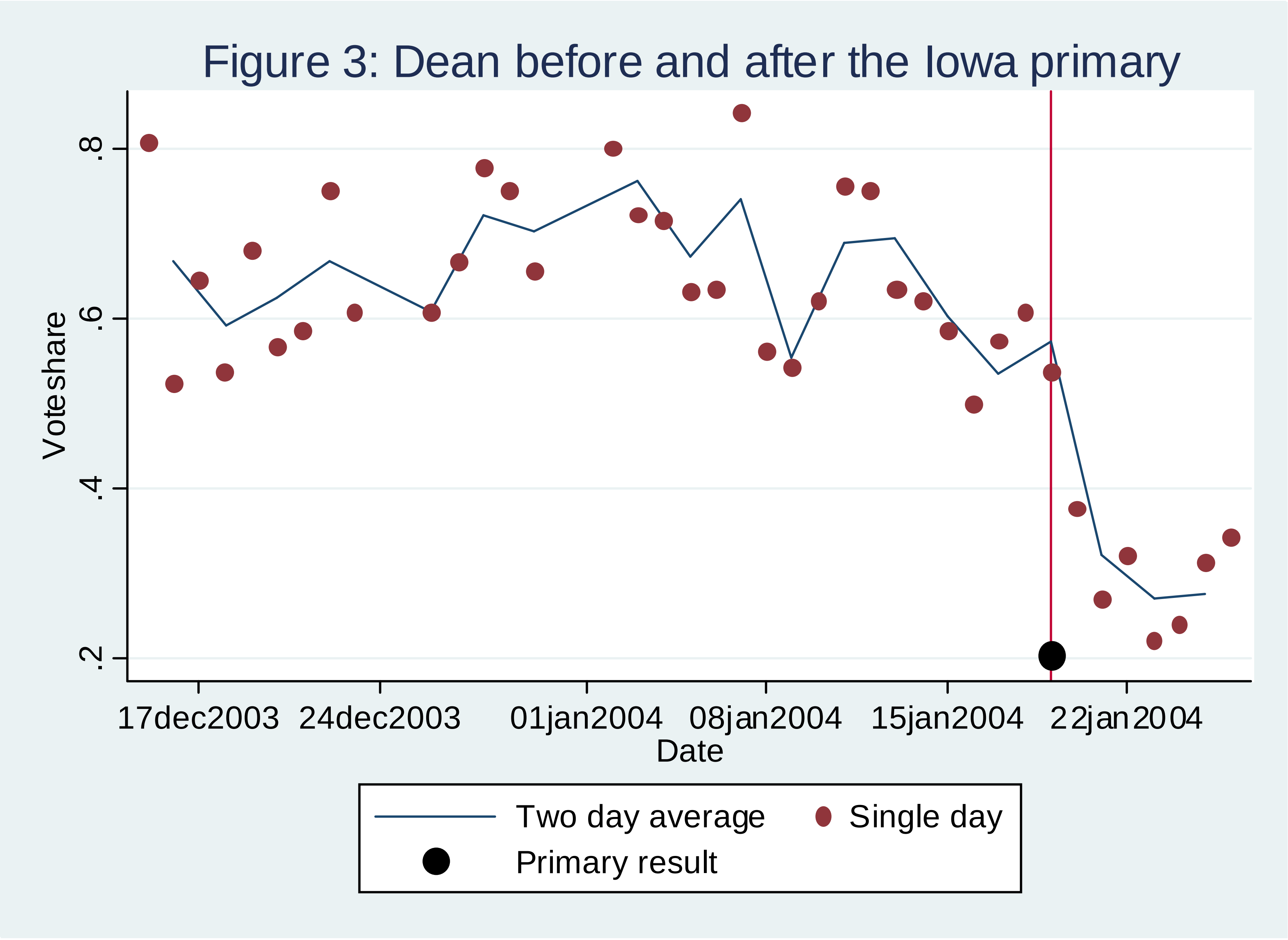


Figure 3: Dean before and after the Iowa primary
Voteshare
.2
.4
.6
.8
17dec2003
24dec2003
01jan2004
08jan2004
15jan2004
22jan2004
Date
Two day average
Single day
Primary result

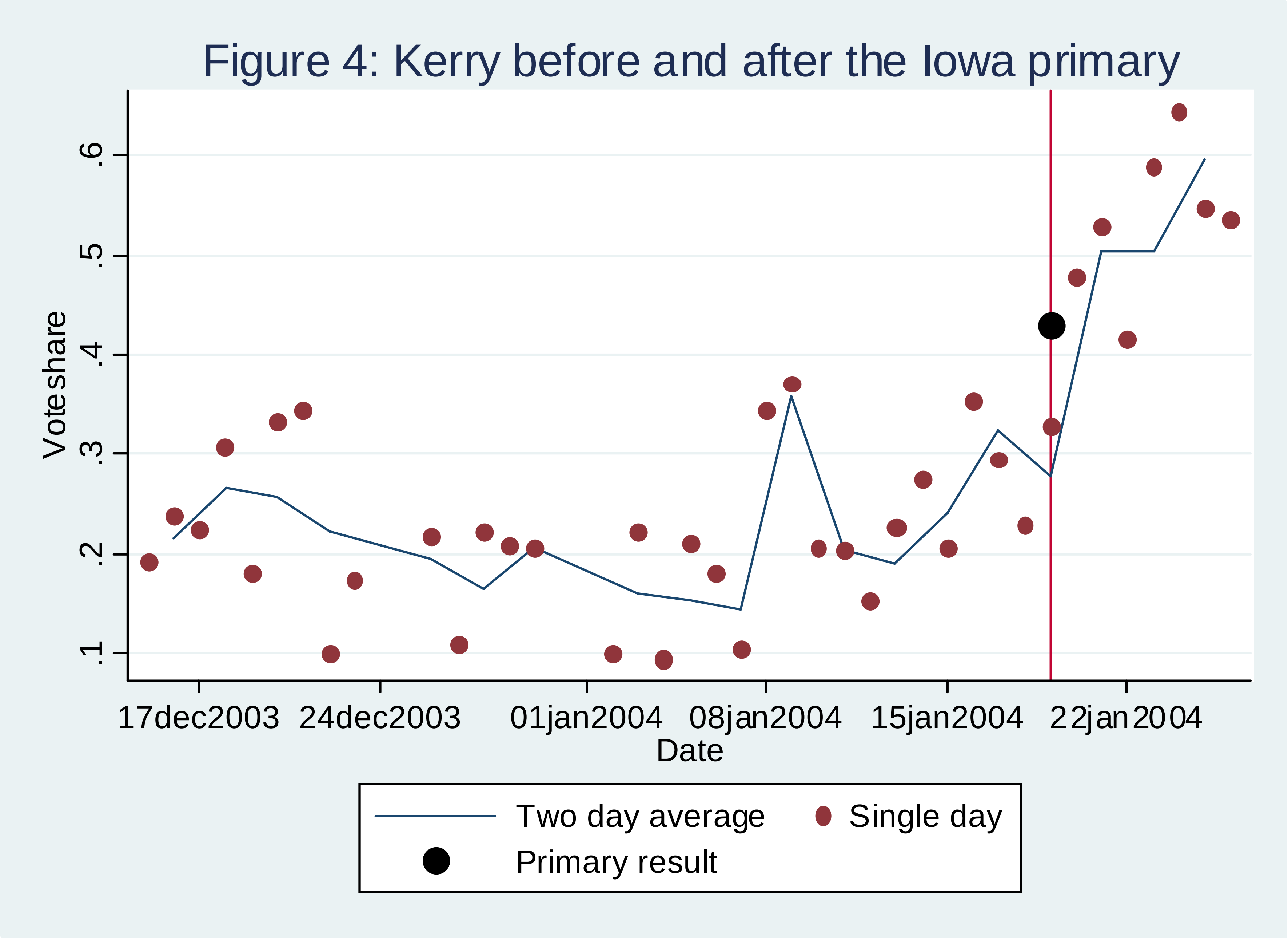


Figure 4: Kerry before and after the Iowa primary
Voteshare
.1
.2
.3
.4
.5
.6
17dec2003
24dec2003
01jan2004
08jan2004
15jan2004
22jan2004
Date
Two day average
Single day
Primary result

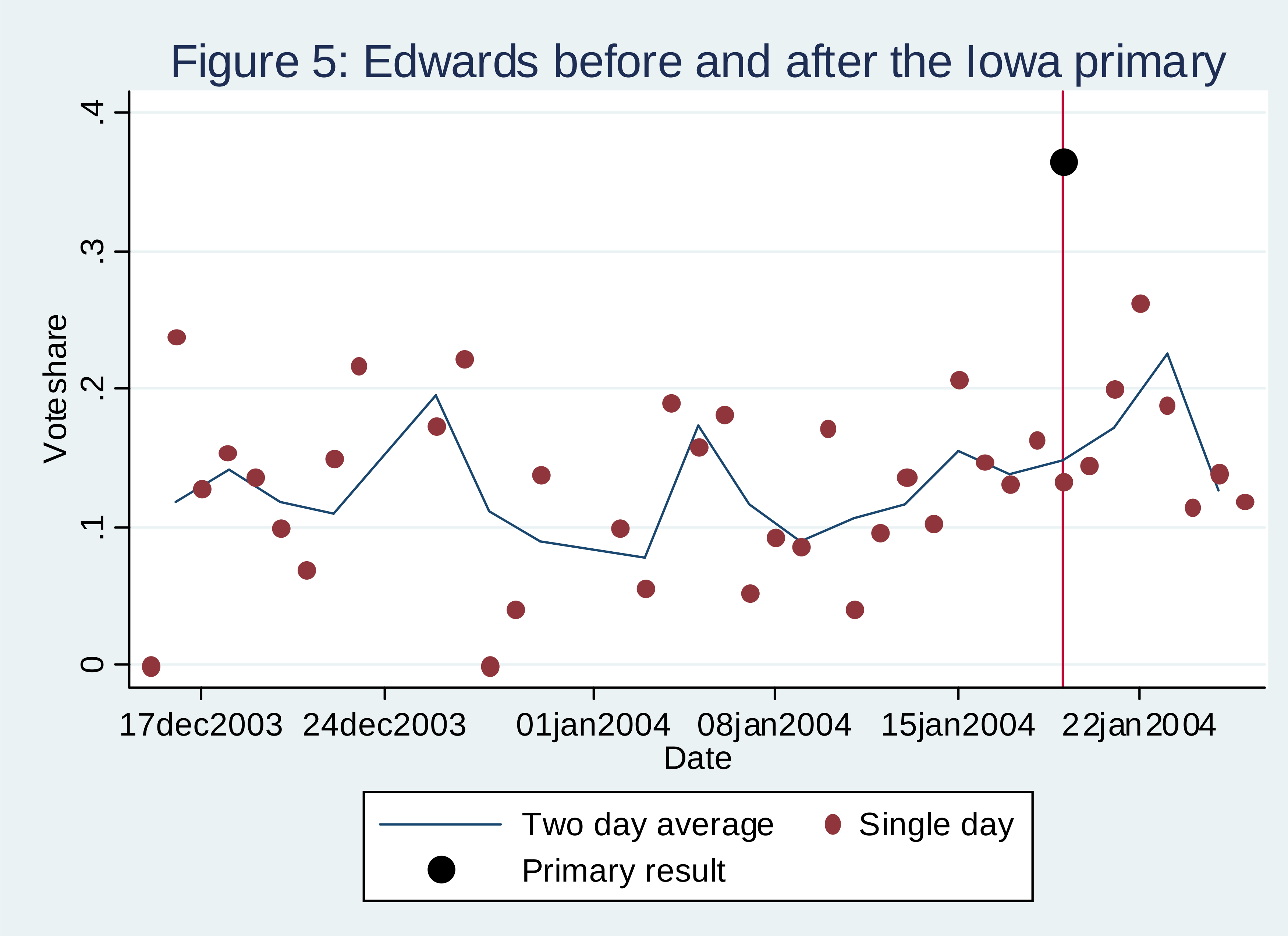
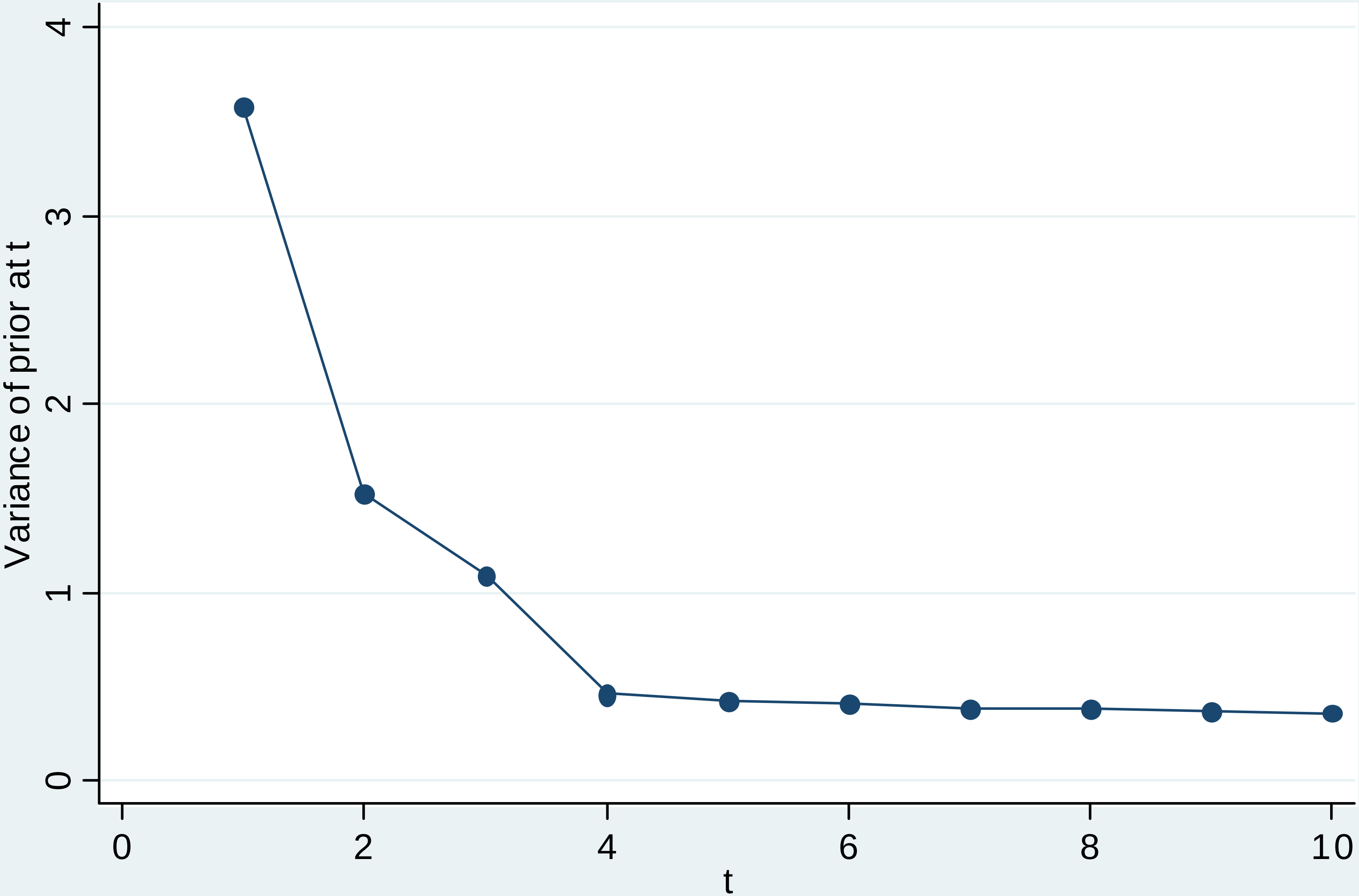


Figure 5: Edwards before and after the Iowa primary
Voteshare
0
.1
.2
.3
.4
17dec2003
24dec2003
01jan2004
08jan2004
15jan2004
22jan2004
Date
Two day average
Single day
Primary result

Figure 6: Uncertainty over quality falls



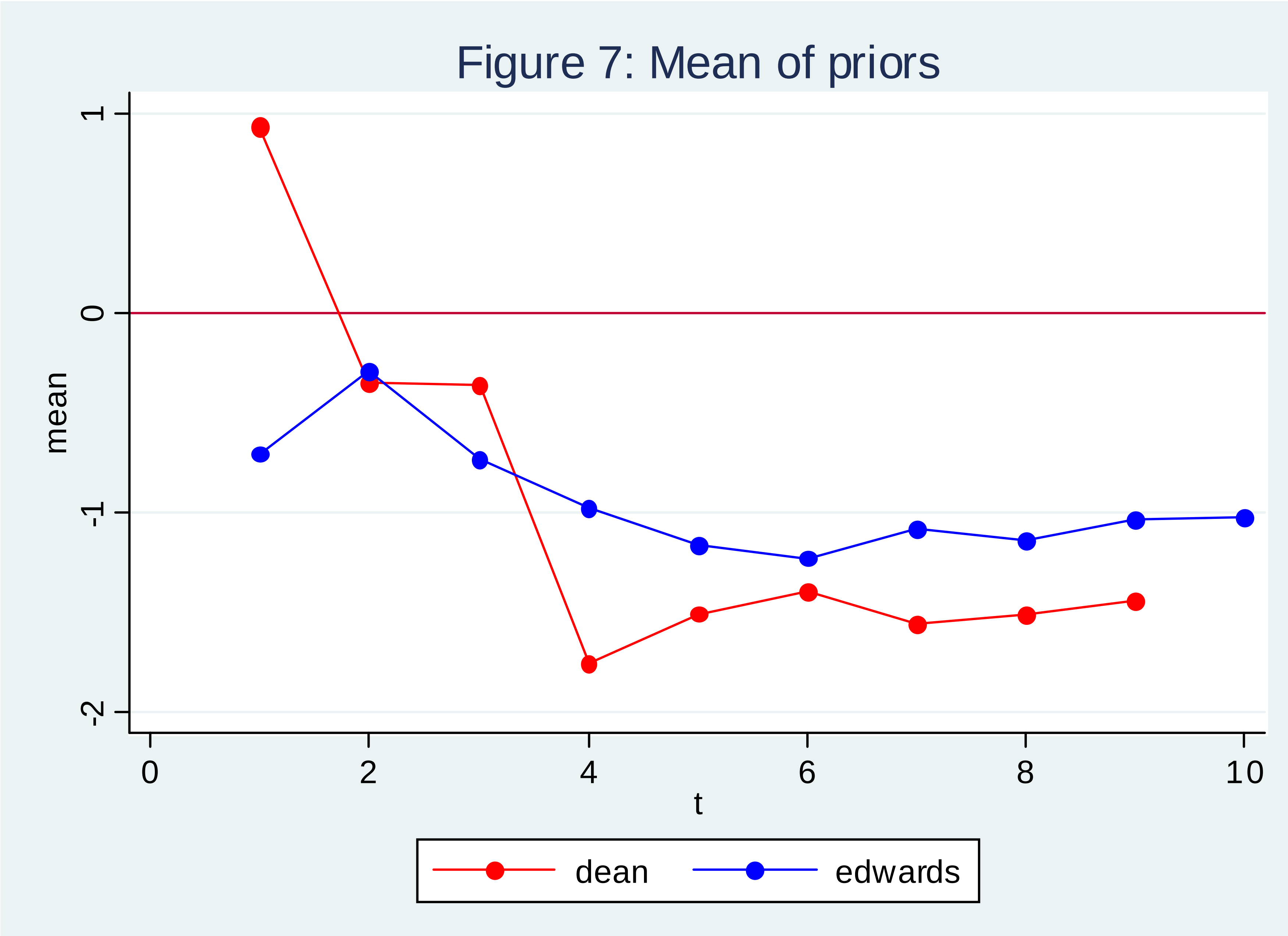


Figure 7: Mean of priors
mean
1
0
-1
-2
0
2
4
6
8
10
t
dean
edwards

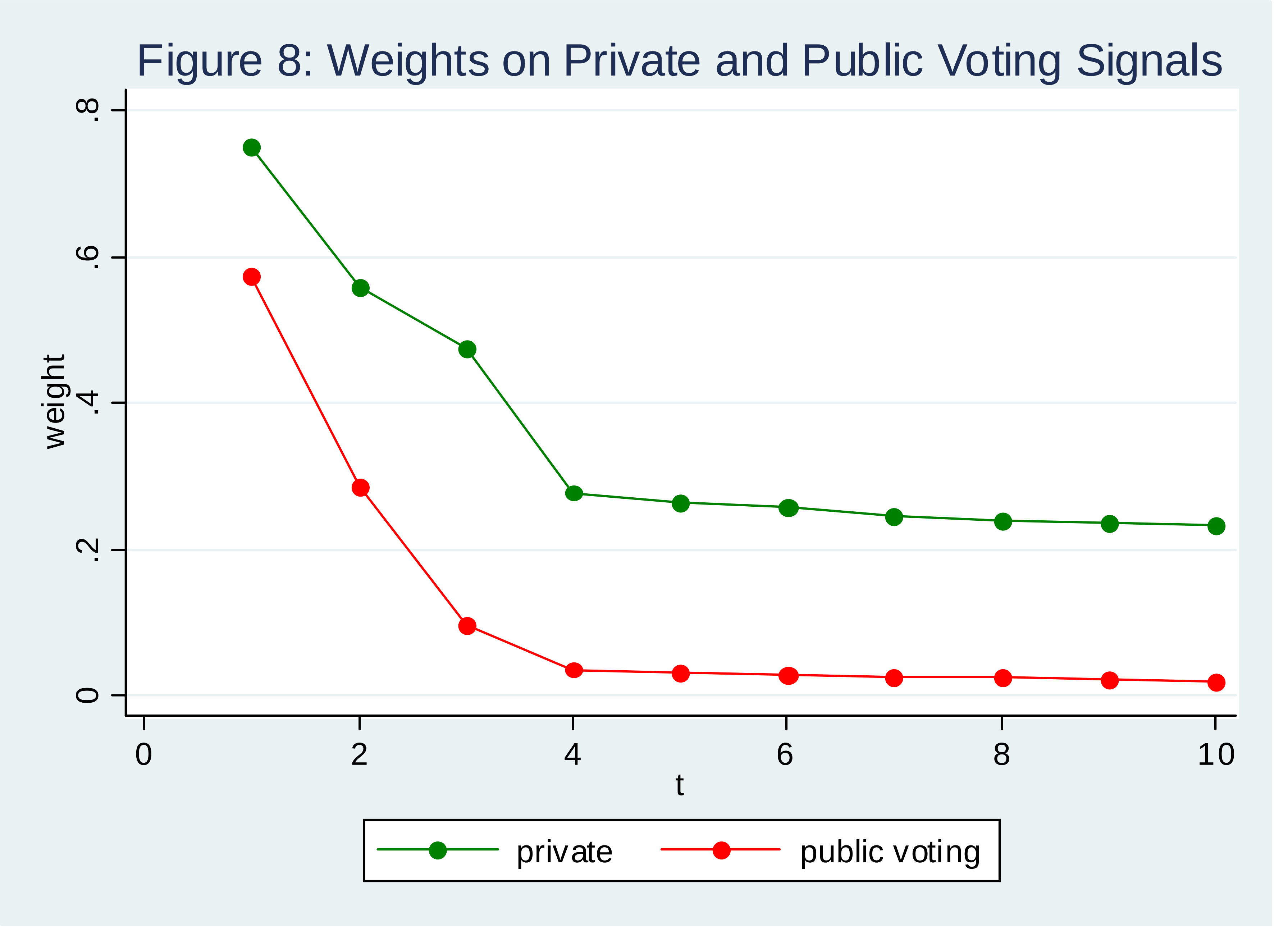


Figure 8: Weights on Private and Public Voting Signals
weight
0
.2
.4
.6
.8
0
2
4
6
8
10
t
private
public voting

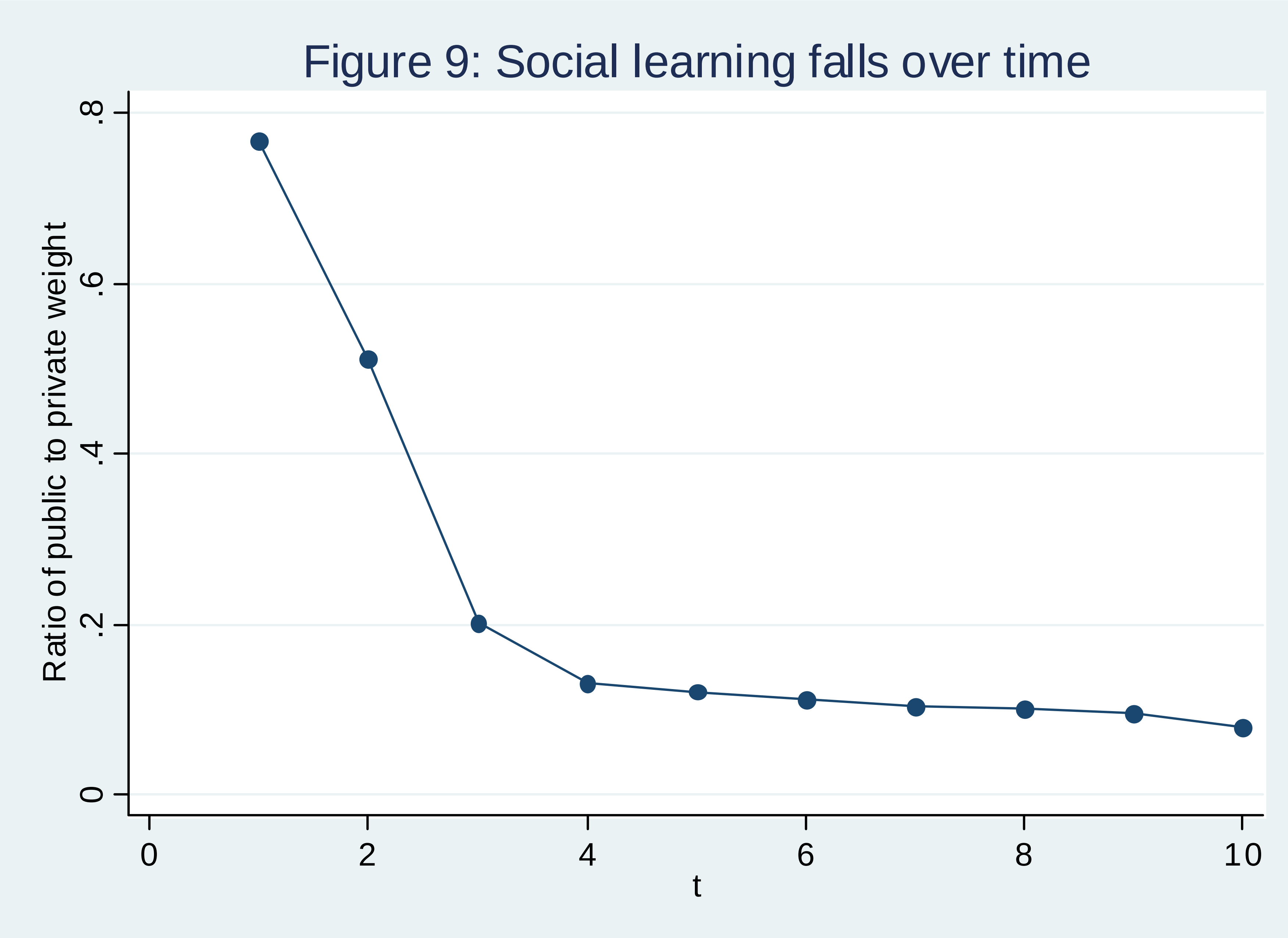


Figure 9: Social learning falls over time
Ratio of public to private weight
0
.2
.4
.6
.8
0
2
4
6
8
10
t

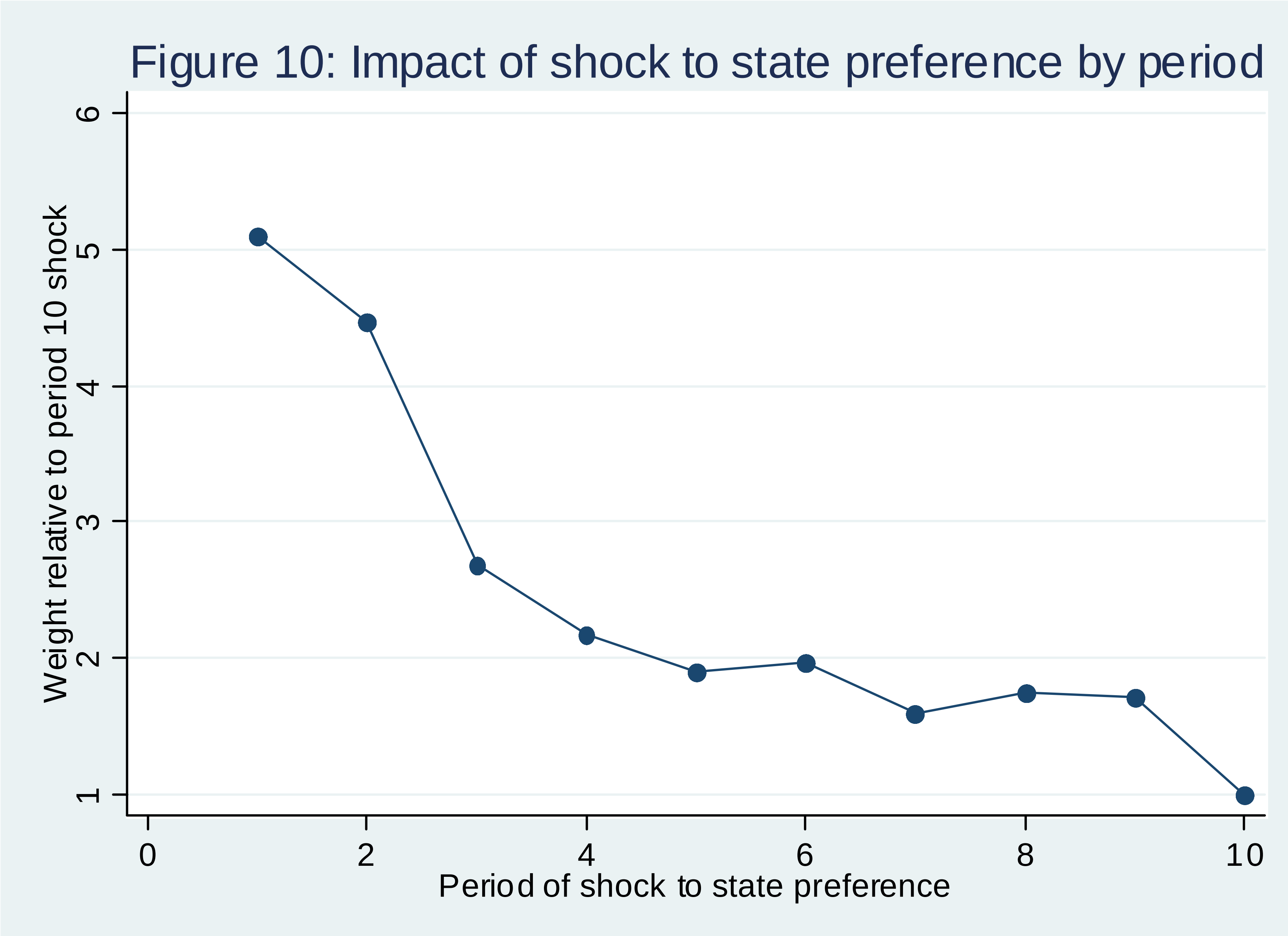


Figure 10: Impact of shock to state preference by period
Weight relative to period 10 shock
1
2
3
4
5
6
0
2
4
6
8
10
Period of shock to state preference

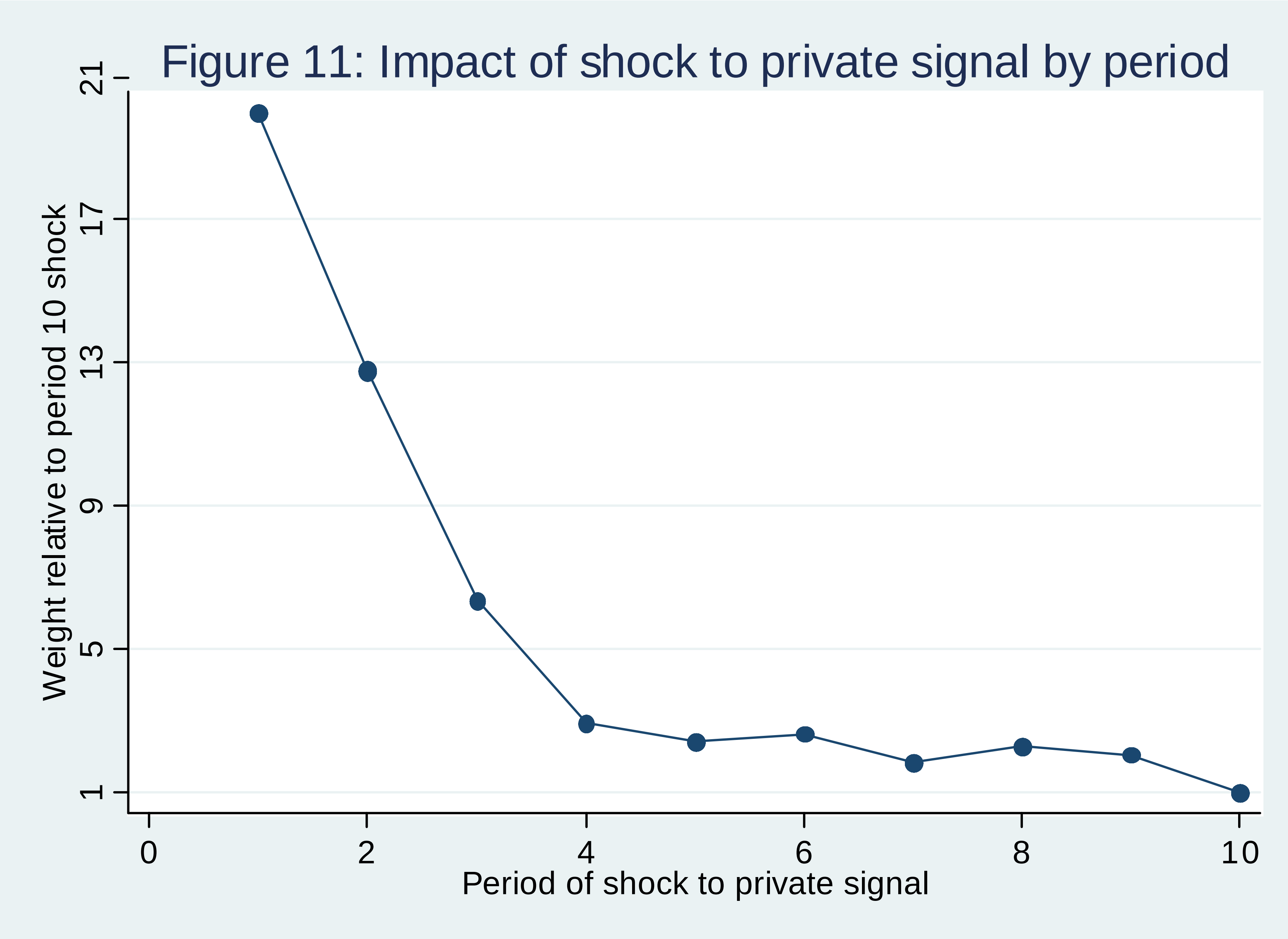


Figure 11: Impact of shock to private signal by period
Weight relative to period 10 shock
21
17
13
9
5
1
0
2
4
6
8
10
Period of shock to private signal

TABLE 1: First Stage Multinomial Logit

	Base Specification		*Includes Distance*		*Includes Time Trend*	
	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	3.577**	5.421**	5.967**
	[1.497, 7.129]	[2.55, 23.572]	[3.492, 14.178]
σ_ε^2	1.197**	1.545**	3.987**
	[0.062, 4.097]	[0.26, 8.501]	[1.619, 9.646]

[bootstrap 95% confidence interval]

Table 3: Additional Measures of Candidate Quality

	Favorability	*Cares About People Like Me*	*Inspiring*	*Strong Leader*	*Trustworthy*	*Shares My Values*	*Knowledge-able*	*Reckless*
Mean Candidate Quality	1.135**	0.468**	0.750**	0.672**	0.365**	0.601**	0.115	-0.130
	[0.052]	[0.136]	[0.164]	[0.153]	[0.132]	[0.136]	[0.171]	[0.196]
Dean	-0.587**	-0.583**	-0.598**	-0.429**	-0.494**	-0.706**	-0.194	1.015**
	[0.070]	[0.189]	[0.199]	[0.187]	[0.179]	[0.197]	[0.246]	[0.294]
Constant	0.152**	0.381**	0.490**	-0.194	0.026	0.303*	-0.802**	-0.242
	[0.057]	[0.160]	[0.188]	[0.175]	[0.155]	[0.162]	[0.209]	[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%

Table 4: Counterfactual Primary

State	Period	Date	*Sequential Primary*			*Simultaneous (3 way)*			*Simulaneous (2 way)*	
			Dean	*Edwards*	*Kerry*	*Dean*	*Edwards*	*Kerry*	*Edwards*	*Kerry*
IA	1	1/19/2004	21%	36%	43%	21%	36%	43%	46%	54%
NH	2	1/27/2004	34%	16%	50%	34%	15%	51%	23%	77%
AZ	3	2/3/2004	22%	11%	67%	24%	6%	70%	8%	92%
DE	3	2/3/2004	14%	15%	70%	27%	5%	68%	7%	93%
MO	3	2/3/2004	10%	29%	60%	7%	27%	66%	29%	71%
NM	3	2/3/2004	23%	16%	60%	29%	12%	59%	17%	83%
OK	3	2/3/2004	7%	49%	44%	8%	45%	47%	49%	51%
SC	3	2/3/2004	6%	56%	38%	4%	41%	55%	42%	58%
MI	4	2/7/2004	20%	16%	63%	60%	6%	34%	14%	86%
WA	4	2/7/2004	35%	8%	57%	92%	0%	8%	3%	97%
ME	5	2/8/2004	33%	10%	57%	86%	1%	14%	4%	96%
TN	6	2/10/2004	6%	37%	57%	2%	72%	26%	73%	27%
VA	6	2/10/2004	8%	31%	61%	3%	43%	54%	44%	56%
NV	7	2/14/2004	19%	12%	70%	73%	1%	26%	3%	97%
WI	8	2/17/2004	20%	37%	43%	38%	53%	9%	85%	15%
UT	9	2/24/2004		35%	65%				37%	63%
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%				44%	56%
MA	10	3/2/2004		20%	80%				90%	10%
MD	10	3/2/2004		30%	70%				30%	70%
MN	10	3/2/2004		35%	65%				59%	41%
NY	10	3/2/2004		25%	75%				62%	38%
OH	10	3/2/2004		40%	60%				71%	29%
RI	10	3/2/2004		21%	79%				21%	79%
Number of States Won			0	2	23	4	2	9	6	19
Percentage of Delegates Won			7%	28%	65%	34%	26%	39%	41%	59%

Table 5: Sequential Elections with Randomized Order		
Two Way Sequential	**Won Plurality of States**	**Won Plurality of Delegates**
Kerry	92.3%	89%
Edwards	7.7%	11%
Three Way Sequential		
Kerry	99.1%	94.9%
Edwards	0.5%	3.8%
Dean	0.4%	1.4%