Momentum and Social Learning in Presidential Primaries
Author(s): Brian Knight and Nathan Schiff
Source: *The Journal of Political Economy*, Vol. 118, No. 6 (December 2010), pp. 1110-1150
Published by: The University of Chicago Press
Stable URL: http://www.jstor.org/stable/10.1086/658372
Accessed: 01/06/2011 20:43

Momentum and Social Learning in Presidential Primaries

Brian Knight

Brown University

Nathan Schiff

University of British Columbia

This paper investigates social learning in sequential voting systems. In the econometric model, candidates experience momentum effects when their performance in early states exceeds expectations. The empirical application uses daily polling data from the 2004 presidential primary. We find that Kerry benefited from surprising wins in early states and took votes away from Dean. Owing to these momentum effects, early voters had up to five times the influence of late voters in the selection of candidates, and this helps to explain the distribution of advertising expenditures. Finally, we use the estimated model to conduct two counterfactual experiments.

I. Introduction

While voting occurs simultaneously in many elections, voters choose sequentially in other cases, such as in roll-call voting in legislatures and

We received helpful comments from two anonymous referees, Nageeb Ali, Eric Brunner, Andy Foster, Alan Gerber, Rob McMillan, and seminar participants at Yale, Johns Hopkins, University of Pennsylvania (Wharton), Northwestern, Toronto, Brown, the National Bureau of Economic Research political economy meetings, and the 2009 American Economic Association meetings. Data on advertising were obtained from a project of the University of Wisconsin Advertising Project and include media tracking data from TNSMI/Campaign Media Analysis Group in Washington, DC. The University of Wisconsin Advertising Project was sponsored by a grant from the Pew Charitable Trusts. The opinions expressed in this article are those of the authors and do not necessarily reflect the views of the University of Wisconsin Advertising Project or the Pew Charitable Trusts.

[*Journal of Political Economy*, 2010, vol. 118, no. 6]
 0022-3808/2010/11806-0003$10.00

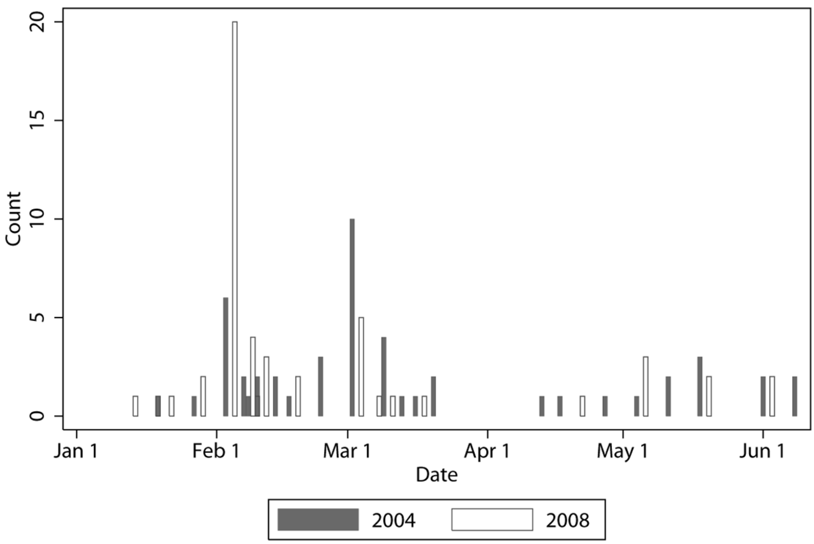

Fig. 1.—Number of primaries by date

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 then mini–Super Tuesday on February 3, 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 on June 8. In 2008 the schedule became increasingly front-loaded, with Nevada scheduled between Iowa and New Hampshire and, perhaps more important, many states moving their primaries to February 5.

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 on 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.[1] This opportunity for late voters to learn from

[1] Bikhchandani, Hirshleifer, and Welch (1992) mention presidential primaries as an example of sequential decision making and potential observational learning.

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 in 2004, 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. 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 those states. In the context of this model we show that candidates benefit from momentum effects when their performance in early states exceeds expectations.

In order to estimate the degree of social learning in sequential elections, we examine daily polling data in the 2004 Democratic primaries. Our empirical strategy involves comparing support for candidates among late voters before and after the release of voting returns from 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. Our estimates demonstrate substantial momentum effects. Using the estimated model, we next calculate the degree of any disproportionate influence for early voters and examine its implications for the distribution of campaign resources across states.

Finally, we use the model to simulate electoral outcomes under counterfactual electoral systems, an area in which the scope for real-world randomized experimentation is obviously quite limited. In particular, we show that the race would have been much tighter under a simultaneous system because of the lack of social learning. We also simulate the election under alternative calendars and show that electoral outcomes are sensitive to the order of voting.

The paper proceeds as follows. Section II provides an overview of the relevant theoretical and empirical literature. Section III lays out the basic theoretical and econometric model of momentum in primaries. Section IV describes our empirical application, Section V calculates voting weights, Section VI describes the counterfactual simulations, and Section VII describes possible extensions and summarizes our key findings.

II. Literature Review

Banerjee (1992), Bikhchandani et al. (1992), and Welch (1992) provide the first formal analysis of social learning. Agents choose actions sequentially and are uncertain about the correct action, which depends on the state of the world. 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.[2] Smith and Sorensen (2000) extend this model in two important ways, allowing for a continuum of signals and for heterogeneity in preferences across agents, and these two extensions will also be important ingredients of our model.

One of the key contributions of our paper is providing microfoundations for our empirical analyses in terms of building an econometric model from a theoretical model of social learning and sequential voting.

[2] This social learning framework has been applied in a variety of empirical settings. Welch (2000), e.g., 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 (2009) conduct a field experiment in which the top-selling dishes are 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 Bikhchandani, Hirshleifer, and Welch (1998) and Chamley (2004).

Of course, a key question in developing this application is whether these theoretical 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 on the actions of all agents.[3] Dekel and Piccione (2000) show that every equilibrium of the simultaneous game is also an equilibrium of the sequential game, regardless of the sequence. In their model, 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. Note, however, that this result does not demonstrate an equivalence between simultaneous and sequential elections; because of multiplicity, there are equilibria of the sequential game that are not equilibria of the simultaneous game. Indeed, Ali and Kartik (2010) construct an equilibrium in sincere voting in which voters do condition on the history of voting.[4]

Most of the empirical work on momentum in presidential primaries has come from the political science literature.[5] Bartels (1987, 1988) uses data from the National Election Study to predict the dynamics of the 1984 Democratic primary. He shows that simple ratings of candidates do not fit the dynamics as well as models that include measures of candidate viability. This model using viability suggests that candidate

[3] Several authors have suggested alternative theoretical reasons for momentum effects. Callander (2007) proposes a model in which every voter gains utility from both conforming, defined as supporting the eventual winner, and voting informatively, defined as supporting the best candidate on the basis of his or her 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. 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. This effect, which favors later winners, thus moves in the opposite direction of momentum, which favors early winners.

[4] 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.

[5] 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 candidates have a chance of winning. The authors test this hypothesis 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 (2007) test predictions of the sequential voting model of Battaglini (2005), which incorporates costly voting and endogenous turnout.

Gary Hart's surprising early victories convinced later voters of his viability.

The most closely related research is a series of papers that use state-level voting returns from multiple presidential primaries and examine the impact of returns from early states on overall primary outcomes. Adkins and Dowdle (2001) use data from 1980–96 and find that New Hampshire plays a key role in determining the ordinal ranking of candidate finishes. Using data from 1976–2000 and a richer set of control variables, Steger, Dowdle, and Adkins (2004) also find an important role for New Hampshire in determining the eventual nominee. Using data through 2004, Steger (2008) confirms the New Hampshire effect but also documents an important role for Iowa.

Relative to these studies using state-level voting returns, our study offers two contributions to the literature. First, our empirical strategy is based on the use of daily polling data, which allows us to compare the preferences of late voters before and after the release of returns from early states. The existing literature, which has effectively examined the correlation between early returns and late returns, may suffer from the problem of unobserved candidate characteristics discussed in the introduction. Second, we build an empirical model from microfoundations. This provides us a framework for measuring expectations and thus the informational content associated with returns from early states. This framework also helps in the development of weights that voters place on the returns from states voting earlier in the sequence. Finally, the microfoundations allow us to conduct several counterfactual policy experiments that are not possible using the methods in the existing literature.

III. Theoretical Framework

This section lays out our basic theoretical and econometric framework for measuring momentum effects in sequential elections, and the notation here roughly follows that in Chamley (2004). 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.

A. Setup

Consider a set of states (s) choosing between candidates ($c = 0, 1, \ldots, C$) 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, follows a type I extreme value, and is independently distributed across both candidates and voters. Note that quality is valued equally by all voters and can be interpreted as a fundamental positive characteristic, such as competence or integrity. We normalize utility from the baseline candidate to be zero for all voters ($u_{0is} = 0$). While underlying state 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 distributed normally ($\eta_{cs} \sim N(0, \sigma_\eta^2)$) 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 distributed normally ($\varepsilon_{cs} \sim N(0, \sigma_\varepsilon^2)$) and independently across states. These signals can be interpreted in a variety of ways, including town hall meetings with candidates, media coverage of candidate debates and appearances within the state, endorsements of candidates by either local media outlets or local politicians, political advertising on local television channels, and so forth. We assume that this signal is common across all voters within a state.[6] Finally, we assume that the signal is unobserved by voters in

[6] 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

other states. (Later in the paper we discuss the implications of an alternative model in which signals are observed by all voters on a national basis.)

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

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

Finally, regarding voter behavior, we assume sincere voting. In particular, given the information available, voter i in state s at time t supports the candidate who provides the voter with the highest level of expected utility.[7] This behavioral assumption is myopic in the sense that voters do not account for how their vote will influence later voters. We also abstract from several other commonly studied motives, such as bandwagon effects, under which voters have a preference for supporting the eventual winner. Bandwagon effects 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.

B. 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

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

Voters thus place more weight on their private signal (θ_{cs}) the higher the variance in the prior over quality (σ_t^2) and the lower the degree of noise in the signal (σ_ε^2).

Plugging equation (4) into equation (3), we have that

$$E(u_{cis}|\theta_{cs}, \eta_{cs}, \nu_{cis}) = \alpha_t\theta_{cs} + (1 - \alpha_t)\mu_{ct} + \eta_{cs} + \nu_{cis}. \tag{6}$$

Then, using the fact that ν_{cis} is distributed type I extreme value, we can write the probability that voter i supports candidate c conditional on the state-level variables (the signal θ_{cs} and unobservables η_{cs}) as follows:

---

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.

[7] These strategies are similar to the sincere voting strategies analyzed in Ali and Kartik (2010).

$$\Pr\left(E(u_{cis}|\theta_{cs}, \eta_{cs}, \nu_{cis}) > E(u_{dis}|\theta_{ds}, \eta_{ds}, \nu_{dis});\ \forall d \neq c\right) = \frac{\exp\left[\alpha_t\theta_{cs} + (1-\alpha_t)\mu_{ct} + \eta_{cs}\right]}{\sum_{d=0}^{C}\exp\left[\alpha_t\theta_{ds} + (1-\alpha_t)\mu_{dt} + \eta_{ds}\right]}. \tag{7}$$

Under the assumption of a continuum of voters, state-level vote shares are equal to these voting probabilities.[8] That is, the vote shares for candidate c relative to the baseline candidate 0 are given by

$$v_{cst}/v_{0st} = \frac{\exp\left[\alpha_t\theta_{cs} + (1-\alpha_t)\mu_{ct} + \eta_{cs}\right]}{\exp\left[\alpha_t\theta_{0s} + (1-\alpha_t)\mu_{0t} + \eta_{0s}\right]}. \tag{8}$$

Finally, when we take logs and use the normalization that the utility from the baseline candidate equals zero, vote shares in state s voting at time t can be described as follows:

$$\ln\left(v_{cst}/v_{0st}\right) = \eta_{cs} + \alpha_t\theta_{cs} + (1-\alpha_t)\mu_{ct}. \tag{9}$$

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 on 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.

C. *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 (9). However, rearranging equation (9) and using the fact that $\theta_{cs} = q_c + \varepsilon_{cs}$, we can say that transformed vote shares provide a noisy signal of quality:

$$\frac{\ln\left(v_{cst}/v_{0st}\right) - (1-\alpha_t)\mu_{ct}}{\alpha_t} = q_c + \frac{\eta_{cs}}{\alpha_t} + \varepsilon_{cs}, \tag{10}$$

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_{cs}/α_t); the combined variance of the noise in the voting signal thus

[8] In a theoretical model with a continuum of voters, sincere voting is an equilibrium since the behavior of individual voters has no effect on overall vote shares and hence the behavior of later voters. While this equilibrium may not be unique, it does help to motivate our assumption of sincere voting.

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:

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

$$\frac{1}{\sigma_{t+1}^2} = \frac{1}{\sigma_t^2} + \frac{N_t}{(\sigma_\eta^2/\alpha_t^2) + \sigma_\varepsilon^2}, \tag{12}$$

where the weight on the voting signals is given by

$$\beta_t = \frac{N_t \sigma_t^2}{N_t \sigma_t^2 + (\sigma_\eta^2/\alpha_t^2) + \sigma_\varepsilon^2}. \tag{13}$$

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 rewrite equation (11) as follows:

$$\mu_{ct+1} - \mu_{ct} = \frac{\beta_t/N_t}{\alpha_t} \sum_{s \in \Omega_t} [\ln (v_{cst}/v_{0st}) - \mu_{ct}]. \tag{14}$$

Social learning $(\mu_{ct+1} - \mu_{ct})$ thus depends on the surprises in voting returns, defined as the deviations in vote shares from expectations over candidate performance.[9] Interestingly, this learning rule implies that candidates who do not win the primary in state s can still benefit from momentum effects as 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.

[9] An important implicit aspect of the above formulation is that expectations over electoral outcomes, as captured by μ_{ct} in eq. (14), depend on national, but not state-specific, factors. This result follows from our assumption of unobserved state preferences. We make this assumption for three reasons. First, if state-level preferences are perfectly observed, then private signals can be inferred from voting returns, and preferences of early states would be assumed to have no disproportionate impact. Given this, we would be assuming away one of the most important controversies surrounding the primary system. Second, with perfectly observed preferences, we have that $\alpha_t = \beta_t$ for dates on which a single state votes $(N_t = 1)$. In this case, the updating rule takes the form $\mu_{ct+1} - \mu_{ct} = \ln (v_{cst}/v_{0st}) - \eta_{cs} - \mu_{ct}$, and momentum effects would be effectively assumed rather than estimated. Third, it may be difficult for voters to know the state preferences of most other states given that, with the exception of Iowa and New Hampshire, most polls of other states were reported infrequently, if at all. Even with Iowa and New Hampshire there could be error in the polls, and not all voters in other states may diligently follow this polling; it is thus far from clear that all voters could know the preferences of other states. Finally, we should note that later in the paper we relax the assumption of completely unknown state preferences by using an alternative specification in which state preferences are partially observed.

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

$$\frac{\partial \mu_{ct+1}}{\partial \ln (v_{cst}/v_{0st})} = \frac{\beta_t/N_t}{\alpha_t} = \frac{\sigma_t^2 + \sigma_\varepsilon^2}{N_t\sigma_t^2 + (\sigma_\eta^2/\alpha_t^2) + \sigma_\varepsilon^2}. \tag{15}$$

This parameter shows the ratio of the weight placed on a public signal, which is an indirect measure of the private signal, to the weight placed on a private signal that is received directly. Voters do not observe preferences of voters in other states, and this ratio is thus bounded between zero and one. When there is no heterogeneity in state preferences ($\sigma_\eta^2 = 0$) and for the special case of a single election ($N_t = 1$), public and private weights are equal ($\beta_t = \alpha_t$), and the ratio is one. As the degree of heterogeneity in state preferences increases, the private signal becomes more difficult to infer and the ratio converges to zero. As with updating from a private signal, the degree of social learning is decreasing in the degree of noise in the signal (σ_ε^2) and increasing in the variance of the prior (σ_t^2).

IV. 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, Senator John Kerry of 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.

A. 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. We focus on the campaigns of the three major candidates: Dean, Edwards, and Kerry, where Kerry is considered the baseline candidate.[10] Individual-level data are taken from the 2004 National Annenberg Election Survey (NAES), which conducted interviews on a daily basis in a rolling cross-section design beginning on October 7, 2003, and continuing through the general election in November 2004.[11] 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.[12] 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 this part of our analysis. Finally, as will be described below, we aim to estimate the state-specific preference parameters (η_{cs}), and thus insufficient data force us to also delete respondents from Washington, DC, and seven small states.[13] These individual-level data are then merged with state-aggregate vote shares from the 2004 primary season as reported on the Web site http://www.cnn.com.

While the econometric analysis to follow uses voting returns from all states, we next highlight our identification strategy in figures 2, 3, and 4 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 underperformed in Iowa relative to expectations, as captured by pre-Iowa polling levels, and voters in the Annenberg survey updated

[10] Another candidate, Wesley Clark, was considered viable in the months leading up to the primary season. He chose, however, not to 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 nonparticipation, 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.

[11] Sampling was done using random-digit dialing to create a national cross section in which each observation is a unique individual. To deal with differences in the population distribution of area codes, the NAES proportioned the first eight digits of the phone number to match the national incidence. We also include a separate oversample of New Hampshire voters covering the period surrounding the New Hampshire primary (January 9 to February 3, 2004).

[12] We assume that respondents report their voting intentions truthfully. Owing to any momentum effects, however, respondents may have incentives to misreport their preferences in polling data to dampen expectations for their preferred candidate. To address this issue, we test for the accuracy of these voting intentions using residents from New Hampshire, a state that was oversampled in the Annenberg survey, during the period after the Iowa caucuses but before the New Hampshire primary. According to these voting intentions, 33 percent planned to vote for Dean, 52 percent for Kerry, and 15 percent for Edwards. These intentions match the actual three-candidate vote shares surprisingly well. In particular, Dean received 34 percent, Kerry received 50 percent, and Edwards received 16 percent. While we cannot conduct similar tests for other states because of small sample sizes, the close match between intentions and voting returns in New Hampshire suggests that respondents do report truthfully.

[13] Specifically, we dropped Arkansas, Idaho, Kansas, North Dakota, South Dakota, Vermont, and Wyoming.

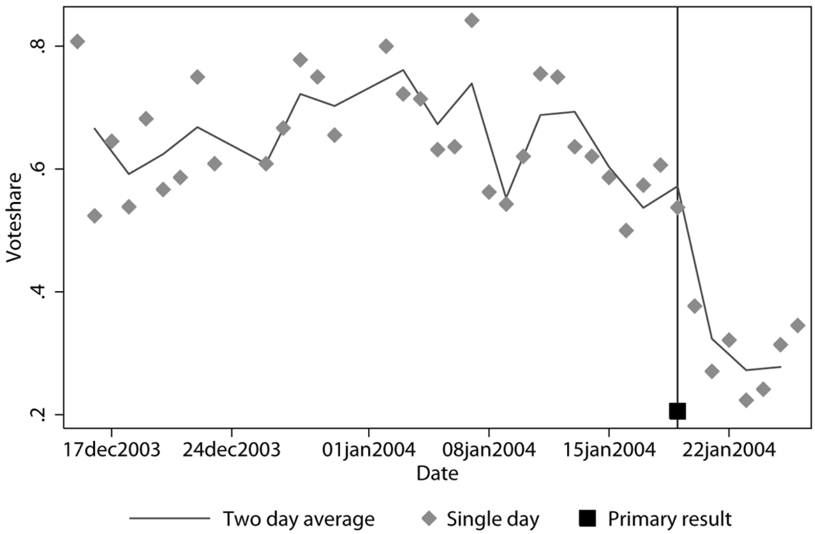

FIG. 2.—Dean before and after the Iowa primary

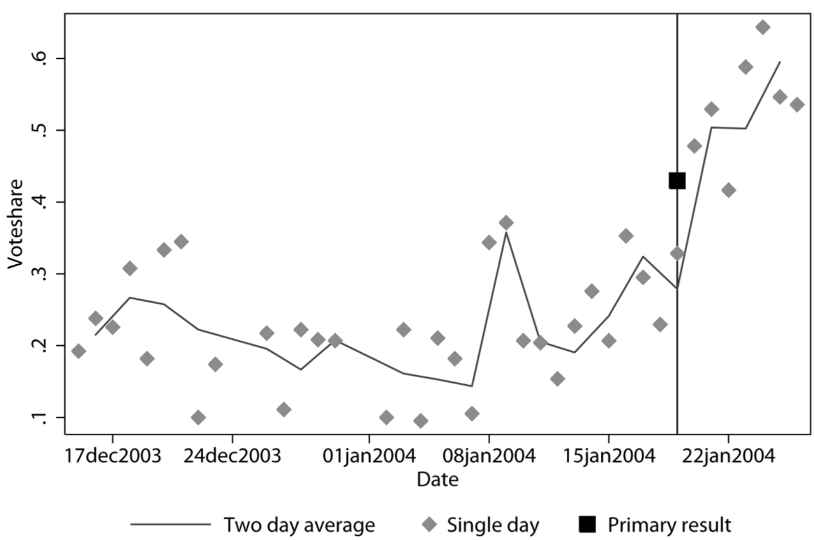

FIG. 3.—Kerry before and after the Iowa primary

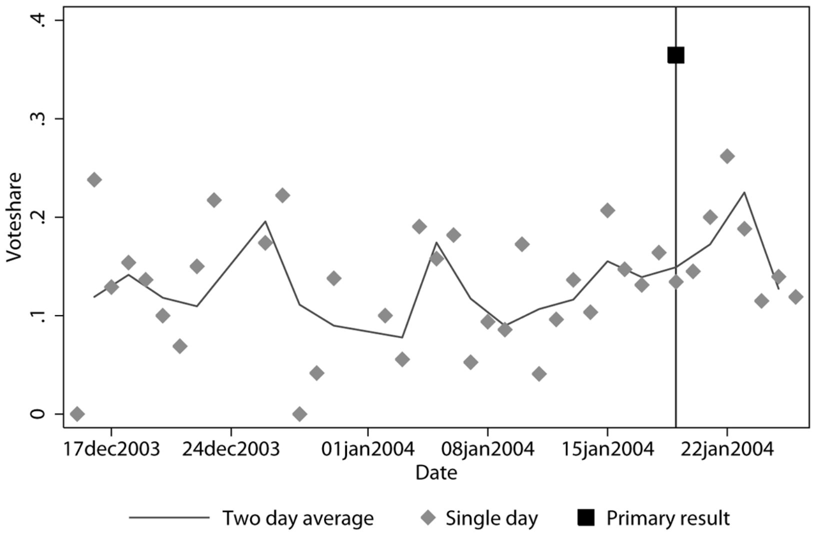

FIG. 4.—Edwards before and after the Iowa primary

appropriately.[14] Kerry, by contrast, outperformed expectations in Iowa, and survey respondents updated accordingly. Edwards also outperformed his pre-Iowa polling numbers, and his polling numbers increased following Iowa. After a few days, however, his support fell back to pre-Iowa levels.[15]

B. *Empirical Model*

As noted above, our empirical strategy involves comparing support for candidates among late voters before and after the release of voting returns from early states. In our econometric specification, we assume that voters receive their signals just prior to the date of their primary. That is, these respondents from late states have not yet observed their

[14] The "Dean scream" will be discussed in detail below.

[15] These patterns are similar to those in prices from the Iowa Electronic Market, in which market participants purchased contracts that would pay $1 in the event that Kerry, e.g., 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, who 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 to follow.

private signals.[16] In this case, their voting intentions can thus be summarized as follows:

$$\Pr\left(E(u_{cis}|\eta_{cs}, \nu_{cis}) > E(u_{dis}|\eta_{ds}, \nu_{dis});\, \forall d \neq c\right) = \frac{\exp(\eta_{cs} + \mu_{ct})}{\sum_{d=0}^{C} \exp(\eta_{ds} + \mu_{dt})}. \tag{16}$$

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:

$$\alpha_t = \frac{\sigma_t^2}{\sigma_t^2 + \sigma_\varepsilon^2}, \tag{17}$$

$$\beta_t = \frac{N_t \sigma_t^2}{N_t \sigma_t^2 + (\sigma_\eta^2/\alpha_t^2) + \sigma_\varepsilon^2}, \tag{18}$$

$$\mu_{ct+1} - \mu_{ct} = \frac{\beta_t/N_t}{\alpha_t} \sum_{s \in \Omega_t} [\ln(v_{cst}/v_{0st}) - \mu_{ct}], \tag{19}$$

$$\frac{1}{\sigma_{t+1}^2} = \frac{1}{\sigma_t^2} + \frac{N_t}{(\sigma_\eta^2/\alpha_t^2) + \sigma_\varepsilon^2}. \tag{20}$$

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 on 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 $(\mu_{c3}, \sigma_3^2, \alpha_3, \beta_3)$, and so on. Each day in which there is one or more elections is the start of a new period. There are thus 10 periods in our sample; see table 4 below for periods, states, and election dates.

Thus, it should be clear that the key parameters to be estimated are

[16] We also considered an alternative formulation under which all voters receive their signals prior to the start of the primary season. For two reasons, we chose not to pursue this alternative formulation. First, identification of the key parameters is more complex since expected quality is now a weighted average of signals and priors. Second, as will be seen below, our analysis demonstrates that respondents who have already voted are better informed when responding to auxiliary questions regarding candidate characteristics than respondents who have not yet voted.

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).[17] 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

$$\Pr\left(E(u_{cis}|\eta_{cs}, \nu_{cis}) > E(u_{dis}|\eta_{ds}, \nu_{dis});\, \forall d \neq c\right) = \frac{\exp(\eta_{cs} + \mu_{c1})}{\sum_{d=0}^{C} \exp(\eta_{ds} + \mu_{d1})}. \tag{21}$$

We estimate the state-level preference parameters (η_{cs}), which are normalized to sum to zero and 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 all states voting prior to March 2, the effective end of the primary season, in order to estimate the key parameters (σ_ε^2, σ_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.[18]

The key social learning parameters are identified by voter responses (in the NAES data) 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.

C. *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 pri-

[17] Note that we are assuming that the variance in the noise in the signal (σ_ε^2) is stable throughout our sample period. If more information is released in the early part or the later part of the campaign, then the variance in the noise in the signal should depend on the time period, i.e., $\sigma_\varepsilon^2(t)$. In this case, our estimate can be considered as the average variance in the noise in the signal across the primary season.

[18] 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. Owing to this variability in the number of states in a replication, we report the median coefficients from our bootstrap samples in tables 1 and 2.

TABLE 1
FIRST-STAGE MULTINOMIAL LOGIT

	Base Specification		Includes Distance		Includes Time Trend	
	Dean (1)	Edwards (2)	Dean (3)	Edwards (4)	Dean (5)	Edwards (6)
Constant	.938**	−.701**	.938**	−.701**	1.404**	−.32
	[.773, 1.14]	[−.913, −.433]	[.773, 1.14]	[−.913, −.433]	[1.108, 1.691]	[−.761, .125]
AL	.64	1.114	.676	.75	.642	1.119
	[−.427, 1.849]	[−.784, 2.697]	[−.38, 1.886]	[−1.272, 2.301]	[−.382, 1.878]	[−.826, 2.662]
AZ	.169	−.316	.123	−.496	.188	−.313
	[−.716, 1.86]	[−1.665, .973]	[−.742, 1.825]	[−1.849, .703]	[−.716, 1.866]	[−1.695, 1.105]
CA	.071	−.235	.015	−.377	.134	−.195
	[−.24, .425]	[−.813, .316]	[−.292, .376]	[−.946, .194]	[−.193, .495]	[−.775, .37]
CO	−.53	−.737	−.592	−.881	−.469	−.744
	[−1.219, .611]	[−2.128, .559]	[−1.246, .53]	[−2.271, .432]	[−1.113, .612]	[−2.163, .591]
CT	−.103	.028	−.03	.654	.052	.152
	[−1.641, 1.345]	[−1.088, 2.113]	[−1.52, 1.414]	[−.214, 2.47]	[−1.48, 1.534]	[−1.047, 2.162]
DE	−1.352	.736	−1.308*	1.075	−1.252	.789
	[−2.832, .239]	[−.863, 2.352]	[−2.777, .305]	[−.667, 2.703]	[−2.804, .685]	[−.742, 2.268]
FL	.116	−.141	.195	−.559	.091	−.177
	[−.32, .871]	[−1.293, .718]	[−.205, .939]	[−1.67, .36]	[−.342, .861]	[−1.32, .696]
GA	.332	.835	.405	.444	.315	.797
	[−.578, 1.475]	[−.515, 1.971]	[−.528, 1.538]	[−.864, 1.525]	[−.524, 1.47]	[−.536, 1.954]
IA	.014	.348	−.037	.302	−.054	.313
	[−.674, 1.146]	[−.861, 1.606]	[−.713, 1.095]	[−.911, 1.581]	[−.78, 1.16]	[−.917, 1.572]
IL	.23	−.701	.213	−.837	.258	−.671
	[−.286, .975]	[−2.29, .27]	[−.297, .944]	[−2.395, .141]	[−.254, .961]	[−2.295, .33]
IN	−.325	−1.202*	−.332	−1.31*	−.266	−1.145*
	[−1.079, .446]	[−2.349, .169]	[−1.094, .433]	[−2.408, .021]	[−1.085, .542]	[−2.282, .213]
KY	−.122	.426	−.111	.165	−.119	.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	−.095	.449	−.062	.046	.046	.573

	[−1.255, .982]	[−.873, 1.621]	[−1.228, 1.006]	[−1.268, 1.284]	[−1.136, 1.104]	[−.678, 1.74]
MA	−1.346**	−2.1**	−1.195**	−1.184**	−1.27**	−2.066**
	[−1.885, −.819]	[−3.299, −1.194]	[−1.706, −.724]	[−2.181, −.22]	[−1.799, −.759]	[−3.267, −1.091]
MD	−.195	.066	−.139	.234	−.158	.11
	[−.938, .789]	[−1.518, .95]	[−.836, .838]	[−1.382, 1.086]	[−.844, .965]	[−1.498, .994]
ME	.25	.004	.197	.604	.233	−.017
	[−.803, 1.524]	[−1.18, 1.865]	[−.861, 1.476]	[−.367, 2.504]	[−.869, 1.545]	[−1.167, 1.891]
MI	.278	−.244	.204	−.029	.3	−.285
	[−.339, 1.363]	[−1.581, .975]	[−.404, 1.288]	[−1.464, 1.142]	[−.38, 1.335]	[−1.664, .985]
MN	−.038	−.188	−.138	−.06	−.039	−.19
	[−.772, .803]	[−1.95, .795]	[−.837, .72]	[−1.806, .945]	[−.756, .807]	[−2.015, .859]
MO	.421	−.007	.399	−.197	.419	.027
	[−.809, 2.074]	[−1.555, 1.986]	[−.817, 2.055]	[−1.703, 1.811]	[−.784, 2.107]	[−1.552, 2.058]
MS	.588	.697	.619	.245	.498	.687
	[−.743, 1.594]	[−.367, 1.763]	[−.712, 1.622]	[−.701, 1.409]	[−.86, 1.669]	[−.454, 1.736]
MT	−.922	.724	−.995	.742	−1.064	.608
	[−2.347, .47]	[−.706, 2.035]	[−2.43, .374]	[−.627, 2.067]	[−2.526, .319]	[−.787, 1.973]
NC	.639	3.13**	.714	2.717**	.639	3.143**
	[−.51, 2.133]	[2.005, 4.578]	[−.415, 2.184]	[1.55, 4.179]	[−.523, 2.154]	[2.017, 4.588]
NE	−.764	−.445	−.82	−.522	−.839	−.474
	[−2.817, .734]	[−1.667, 1.094]	[−2.868, .681]	[−1.762, 1.035]	[−2.73, .62]	[−1.578, 1.043]
NH	−.192	−.362	−.283**	.503	−.576**	−.685**
	[−.504, .035]	[−.743, .11]	[−.587, −.048]	[−.111, 1.099]	[−.931, −.27]	[−1.135, −.196]
NJ	−.252	−.491	−.188	.004	−.208	−.464
	[−.838, .73]	[−1.688, .601]	[−.747, .819]	[−1.352, 1.003]	[−.832, .828]	[−1.689, .614]
NM	−.018	−.458	−.047	−.667	−.042	−.442
	[−.921, 1.491]	[−1.246, 1.888]	[−.95, 1.481]	[−1.465, 1.644]	[−1.096, 1.54]	[−1.291, 1.866]
NV	−.368	−.05	−.422	−.168	−.377	−.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]
NY	.35	−.911**	.319	−.391	.397*	−.868*
	[−.078, .793]	[−2.593, −.01]	[−.116, .763]	[−1.84, .476]	[−.025, .845]	[−2.553, .022]
OH	.124	−.094	.122	−.115	.154	−.065
	[−.465, .955]	[−1.175, .885]	[−.469, .951]	[−1.18, .876]	[−.419, .989]	[−1.163, .89]
OK	−.533	.744	−.541	.431	−.601	.748
	[−2.321, 1.008]	[−.656, 2.116]	[−2.328, 1.004]	[−.888, 1.993]	[−2.44, .983]	[−.63, 2.127]

TABLE 1 (*Continued*)

	Base Specification		Includes Distance		Includes Time Trend	
	Dean (1)	Edwards (2)	Dean (3)	Edwards (4)	Dean (5)	Edwards (6)
OR	−.127	−.669	−.194	−.67	−.2	−.741
	[−.963, .61]	[−2.099, .644]	[−1.055, .537]	[−2.099, .653]	[−1.021, .546]	[−2.102, .619]
PA	−.231	−1.116**	−.213	−.776*	−.204	−1.085**
	[−.649, .419]	[−2.392, −.117]	[−.638, .426]	[−2.125, .164]	[−.671, .458]	[−2.392, −.096]
RI	−.527	−.438	−.393	.409	−.577	−.458
	[−1.642, 1.2]	[−1.536, 1.127]	[−1.494, 1.314]	[−.715, 1.713]	[−1.751, 1.317]	[−1.699, 1.187]
SC	.908	2.031**	.971	1.527**	.898	2.05**
	[−.493, 1.986]	[.705, 3.389]	[−.434, 2.084]	[.244, 2.873]	[−.444, 2.015]	[.805, 3.357]
TN	−.115	−.074	−.073	−.368	−.069	−.042
	[−.833, 1.216]	[−1.434, 1.773]	[−.81, 1.257]	[−1.942, 1.273]	[−.806, 1.297]	[−1.416, 1.83]
TX	.034	.493	.035	.152	.019	.462
	[−.408, .663]	[−.319, 1.372]	[−.399, .665]	[−.645, .987]	[−.43, .673]	[−.333, 1.309]
UT	.433	.249	.398	.118	.325	.172
	[−.815, 1.35]	[−.866, 1.388]	[−.876, 1.309]	[−.981, 1.322]	[−.979, 1.331]	[−.871, 1.394]
VA	.387	.221	.456	.144	.476	.245
	[−.576, 1.175]	[−1.397, 1.439]	[−.537, 1.214]	[−1.493, 1.343]	[−.495, 1.208]	[−1.313, 1.486]
WA	.177	−.287	.082	−.226	.187	−.284
	[−.408, .618]	[−1.343, .681]	[−.51, .533]	[−1.29, .734]	[−.383, .657]	[−1.302, .726]
WI	.207	−.028	.146	.109	.262	.031
	[−.769, 1.083]	[−1.35, 1.122]	[−.838, .975]	[−1.21, 1.23]	[−.719, 1.091]	[−1.295, 1.102]
WV	.078	.008	.102	−.048	.038	−.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			−.062**	−.062**		
			[−.103, −.025]	[−.103, −.025]		
Trend					.01**	.008**
					[.005, .015]	[0, .018]

NOTE.—Bootstrap 95 percent confidence intervals are in brackets.
*Significant at 10 percent
**Significant at 5 percent.

TABLE 2
Estimated Parameters

	Base Specification (1)	Includes Distance (2)	Includes Time Trend (3)
σ_η^2	.815**	.707**	.829**
	[.551, 1.194]	[.402, 1.05]	[.546, 1.192]
σ_1^2	3.577**	2.793**	4.525**
	[1.497, 7.129]	[1.221, 6.656]	[2.608, 9.652]
σ_ε^2	1.197**	1.115*	1.833**
	[.062, 4.097]	[−.015, 6.642]	[.309, 5.99]

Note.—Bootstrap 95 percent confidence intervals are in brackets.
*Significant at 10 percent
**Significant at 5 percent.

mary 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 (19). The substantial 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 in an alternative specification, which relaxes the assumption that such regional 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 shown in column 1, the key learning parameters, the initial prior (σ_1^2) and the degree of noise in the signal (σ_ε^2), are both positive and statistically significant.

To provide additional interpretation of these results, we next present in figures 5 and 6 the key dynamics of the model as implied by these estimated parameters and the aggregate returns. As shown in figure 5*A*, voters learned about the quality of the candidates relative to one another from the returns in other states. Prior to the first primary, voters viewed Dean as the highest-quality candidate 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 that candidates can benefit, even relative to first-place finishers, from surprisingly strong second-place finishers. However, 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.

As voters learned from election results, the degree of variance in the beliefs over candidate quality (σ_t^2) fell substantially over the primary season, as depicted in figure 5*B*. 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.

To provide further interpretation of these results, figure 6 plots the implied weights on the private signals observed by voters (α_t), the weights placed on aggregate vote shares after scaling by the number of primaries (β_t/N_t), and the ratio of these two, which is the key social learning parameter ($\beta_t/\alpha_t N_t$). 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 on 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.

In terms of the weights on public signals, late voters initially place roughly 60 percent weight on these signals and 40 percent on their priors. As more primary returns come in, voters place less weight on voting returns and more weight on their prior, and by the end of the sample voters are largely unmoved by primaries held in other states. As this weight falls to zero, the ratio of these two weights ($\beta_t/\alpha_t N_t$) also falls from its starting point of 75 percent to around 10 percent. The reason is that as states place less weight on their private signal (α_t), the variance of the noise in the public signal ($\sigma_\eta^2/\alpha_t^2 + \sigma_\varepsilon^2$) increases and voting returns are thus less informative about quality.

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 matter, however, and Edwards benefited relative to Kerry from a surprisingly strong second-place finish in Iowa. While Dean came in third place in Iowa, he benefited from strong voter

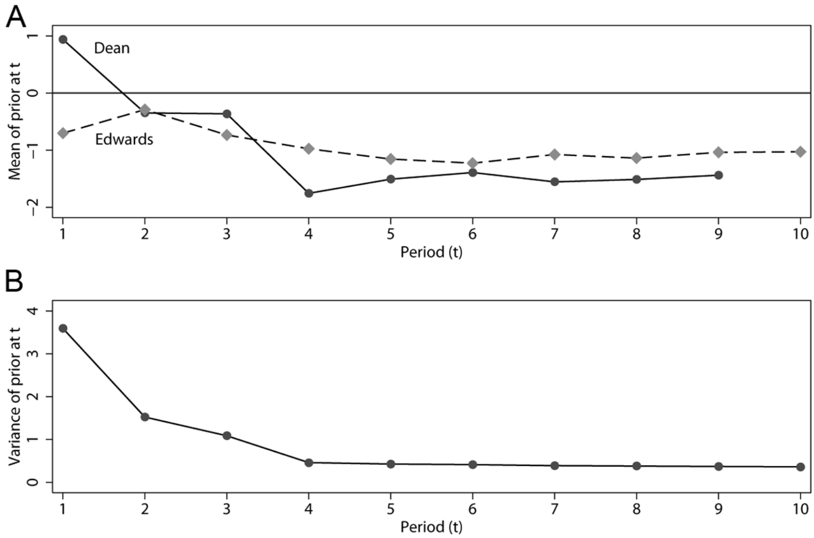

FIG. 5.—*A*, Mean of prior (Kerry = 0); *B*, variance of prior

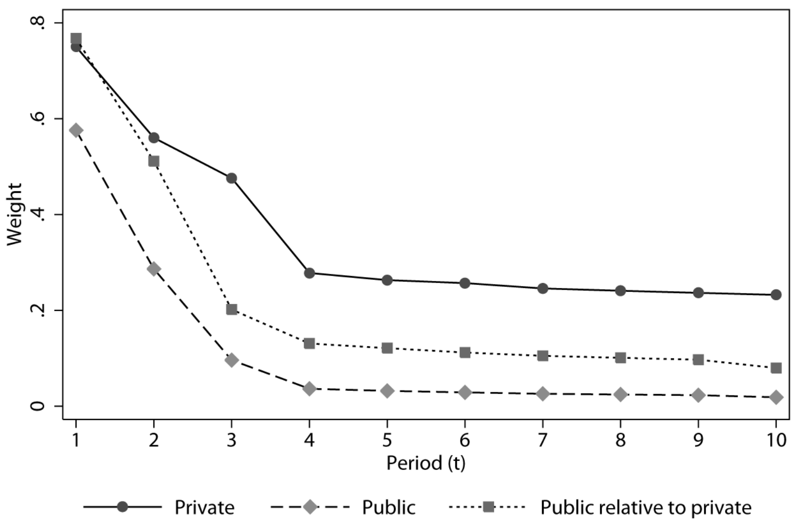

FIG. 6.—Weights on private and public voting signals

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 both on their private signals and on 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.

D. *Additional Specifications*

As noted above, the baseline model assumes that voters observe their own state-level preferences but not those in other states. 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:

$$\ln (v_{cst}/v_{0st}) = \eta_{cs} + \gamma X_{cs} + \alpha_t \theta_{cs} + (1 - \alpha_t)\mu_{ct}, \tag{22}$$

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:

$$\mu_{ct+1} - \mu_{ct} = \frac{\beta_t/N_t}{\alpha_t} \sum_{s \in \Omega_t} [\ln (v_{cst}/v_{0st}) - \gamma X_{cs} - \mu_{ct}]. \tag{23}$$

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. In this way we allow for voters to incorporate perceived regional advantages into their expectations of candidate performance. For example, voters may expect Edwards to outperform national polling trends in his home state of North Carolina. 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 intentions. After this observed dimension of preferences is accounted for, 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 is 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:

$$\ln (v_{cst}/v_{0st}) = \eta_{cs} + \gamma_c\tau + \alpha_t\theta_{cs} + (1 - \alpha_t)\mu_{ct}, \tag{24}$$

where τ indexes the interview date and is normalized to equal zero on the date of the Iowa primary. The social learning rule then becomes

$$\mu_{ct+1} - \mu_{ct} = \frac{\beta_t/N_t}{\alpha_t}\sum_{s\in\Omega_t}[\ln (v_{cst}/v_{0st}) - \gamma_c\tau - \mu_{ct}]. \tag{25}$$

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, whereas Kerry was disadvantaged. Thus, at the time of the Iowa primary, the mean evaluations of Dean and Edwards are higher than 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 are similar to those in the baseline specification.

E. *Alternative Explanations*

In this subsection, we address three alternative explanations for our baseline findings: departures from sincere voting, national information, and campaign finance and persuasion.

1. Departures from Sincere Voting

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, nonlearning 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 electability-driven switching provides an alternative explanation for our results.[19] The second form of strategic voting involves 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 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. Therefore, 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 respondents 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.

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

$$\text{quality}_{itc} = \delta_c + \kappa \times \mu_{ct} + \xi_{itc}, \tag{26}$$

where μ_{ct} is the mean candidate quality at time t, as implied by our estimated model and reflected in figure 5*A*; and quality_{itc} represents responses to the auxiliary questions regarding candidates, relative to Kerry, described above. 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 learn about the preferences of other voters only from early returns.

As shown in panel A of table 3, there is a positive relationship between estimated mean evaluations of candidate quality (μ_{ct}) and respondent

[19] 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, e.g., 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.

TABLE 3

CANDIDATE QUALITY AND VOTER INFORMATION

A. ADDITIONAL MEASURES OF CANDIDATE QUALITY

	Favorability	Cares about People Like Me	Inspiring	Strong Leader	Trustworthy	Shares My Values	Knowledgeable	Reckless
Mean candidate quality	1.135**	.468**	.750**	.672**	.365**	.601**	.115	−.130
	[.052]	[.136]	[.164]	[.153]	[.132]	[.136]	[.171]	[.196]
Dean	−.587**	−.583**	−.598**	−.429**	−.494**	−.706**	−.194	1.015**
	[.070]	[.189]	[.199]	[.187]	[.179]	[.197]	[.246]	[.294]
Constant	.152**	.381**	.490**	−.194	.026	.303*	−.802**	−.242
	[.057]	[.160]	[.188]	[.175]	[.155]	[.162]	[.209]	[.236]
Observations	6,374	965	991	972	962	954	488	479
R^2	.085	.018	.029	.027	.013	.027	.002	.028

B. VOTER INFORMATION ANALYSIS

	Correct Answer That Edwards Is Son of Mill Worker	Correct Answer That Dean Was Governor of VT	Correct Answer That Edwards Was a Trial Lawyer	Total Correct Answers	Correct Answer That Edwards Is Son of Mill Worker	Correct Answer That Dean Was Governor of VT	Correct Answer That Edwards Was a Trial Lawyer	Total Correct Answers
Already voted	.199**	.306**	.153**	.739**	.070**	.052	.050	.169**
	[.024]	[.027]	[.025]	[.066]	[.031]	[.033]	[.031]	[.077]
Observations	4,007	3,154	4,007	3,154	4,007	3,154	4,007	3,154
R^2	.041	.107	.026	.068	.086	.218	.083	.165
Date fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	No	No	No	No	Yes	Yes	Yes	Yes

NOTE.—Robust standard errors are in brackets.

*Significant at 10 percent.

**Significant at 5 percent.

evaluations of candidate quality, 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.[20] Thus, this analysis demonstrates that, in addition to voting intentions, perceptions of candidate quality also respond to the release of voting returns from early states and thus provide support for our social learning interpretation of our baseline findings.

2. National Information

A key assumption in our model is that signals are observed only within the state. If all signals are observed nationally, by contrast, then our model does not predict social learning since early voters do not hold an informational advantage. A key question, however, is whether our estimator would mistakenly detect social learning due to purely national signals. To examine this issue, consider a model with national information, under which all voters, both early and late, update their beliefs upon receiving the signal available at time t (θ_{ct}) as follows:

$$\mu_{ct+1} = \alpha_t\theta_{ct} + (1 - \alpha_t)\mu_{ct}, \tag{27}$$

and voting returns in state s at time t can thus be summarized by

$$\ln (v_{cst}/v_{0st}) = \eta_{cs} + \alpha_t\theta_{ct} + (1 - \alpha_t)\mu_{ct}. \tag{28}$$

Averaging (28) across all states voting at time t, solving for $\alpha_t\theta_{ct}$, and substituting into (27), we have that

$$\mu_{ct+1} - \mu_{ct} = \frac{1}{N_t}\sum_{s\in\Omega_t} [\ln (v_{cst}/v_{0st}) - \eta_{cs} - \mu_{ct}]. \tag{29}$$

Note that equation (29) is similar to equation (14) and thus also predicts that updating by early voters will be correlated with voting returns. The key issue is that any information released between periods t and $t + 1$ will be reflected in voting returns at the end of t and in the updating at the beginning of $t + 1$. So, it is clear that our estimator may detect momentum effects even in the absence of social learning due to the presence of national information.

We address this alternative explanation in two ways. Our first test examines whether any shift in voting intentions in the daily polling data seems to correspond with key election dates. While our baseline estimator uses daily polling data in conjunction with information on the dates of primaries and their associated electoral outcomes, this alter-

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

native analysis uses only the daily polling data in order to detect the most likely date for a break in support for candidates. If our results are driven solely by nationally available information, there is no reason that a shift in support for candidates should correspond with key election dates and may even occur before such events. According to our social learning model, by contrast, a shift in support for candidates should occur on or just following key election dates.

More concretely, using methods analogous to those from the econometrics literature on unknown structural breaks, we first locate the single break in support for which the model best fits the data.[21] That is, we estimate the parameters from the following model for all possible break dates j and choose the date that maximizes the likelihood:

$$\Pr(\text{prefer } c) = \frac{\exp[\eta_{cs} + \mu_c + 1(\tau > j)\mu_{cj}]}{\sum_{d=0}^{C} \exp[\eta_{ds} + \mu_d + 1(\tau > j)\mu_{dj}]}, \tag{30}$$

where η_{cs} is the state-candidate fixed effect, τ indexes the date of the interview, μ_c is a time-invariant candidate-specific constant, $1(\tau > j)$ indicates that the interview occurred after date j, and μ_{cj} represents the shift in support for candidate c occurring after date j. Relative to equation (16), which formed the basis for our baseline estimator, this alternative model also includes state fixed effects but allows for a more flexible pattern in mean support for the candidates, which is no longer required to follow the updating rule in equation (19) and thus does not incorporate information on the timing and outcomes of elections.

To estimate the most likely break date, we use interviews between January 2 and February 17, the day before Dean dropped out of the race.[22] As shown in figure 7, the likelihood is clearly maximized by choosing January 19, the date of the Iowa caucus, as the break date.[23] That is, the model best fits the data by allowing for a break in support between January 19 and January 20.[24] This evidence, which finds a break in support just after, and not before, a key election date, thus supports our social learning interpretation of the baseline results over this interpretation based on nationally available information.

This reaction in polling data to the Iowa outcome is potentially confounded by Dean's reaction, which was dubbed the "Dean scream" and was televised extensively. For a variety of reasons, we feel that this re-

[21] For an overview of this literature, see Hansen (2001).

[22] For the purposes of this analysis, it is helpful not to have time gaps in the data. No interviews occurred on January 1, and some dates prior to January 1 had a very small number of interviews.

[23] Following standard techniques, we trim 5 percent of the observations on each side and thus allow for all possible break dates between January 4 and February 15, for a total of 45 days.

[24] The associated coefficients capturing the size of the break (μ_{cj}) are jointly significant at conventional levels.

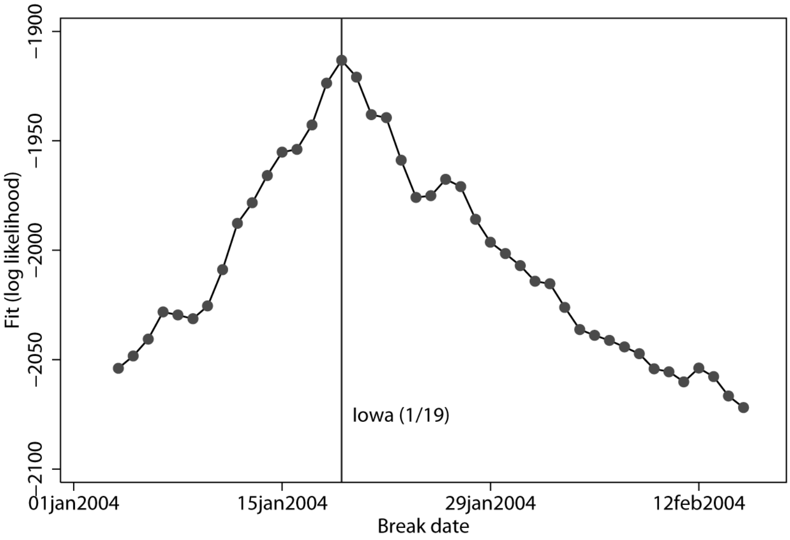

Fig. 7.—Testing for structural break

action does not fully explain these documented shifts in support. First, this media coverage would not arguably 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. Second, the votes that Dean lost went to Kerry and Edwards, the winners in Iowa, rather than to the losers, including Gephardt and Joe Lieberman. Thus, even if Dean's vote loss was due to his reaction, the reallocation of those votes is consistent with our story of momentum associated with social learning from aggregate returns. Third, the television coverage was concentrated in the days following the election, whereas our finding is that the break in support occurred on the election day itself.[25] Fourth, when we exclude Dean from the analysis and focus on the choice between Kerry and Edwards, we find a break on January 22, which is just 3 days after Iowa. Thus, our results are not entirely driven by Dean himself. Finally, when we take the Iowa break as given and allow for a second break in support, we find another shift in support on January 27, which is the date of the New Hampshire primary.[26]

Our second test examines the degree to which respondents who have already voted are better informed than respondents who have not yet

[25] In the first 4 days after the Iowa caucuses, the Dean scream was televised 633 times (*Boston Globe*, January 29, 2004).

[26] These two sets of results are available on request from the authors.

voted.[27] If information is available nationally and is processed in the same manner in every state, as outlined above, then those who have not yet voted and those who have already voted should be equally well informed. If information is available on a state-level basis, by contrast, then, among voters interviewed on the same date, those who have already voted have received their signals and should thus be better informed. To measure the degree to which voters are informed about candidates, we next examine whether respondents answered correctly or incorrectly three factual questions involving candidate characteristics. For example, respondents were asked which candidate was the son of a mill worker, with the correct answer being Edwards.

As shown in panel B of table 3, which reports coefficients from a linear probability model with interview date fixed effects, those who have already voted were 20 percentage points more likely to know that Edwards was the son of a mill worker. Similarly, those who have already voted were more likely to know that Dean was the governor of Vermont. Third, those who have already voted are also more likely to know that Edwards was a trial lawyer. As shown in column 4, those who have already voted on average answered 0.74 more questions (out of three) correctly. One potential limitation of this analysis is that early voters, such as those from Iowa, may be better informed in general than late voters. To address this concern, we next estimate a model with state fixed effects, and these results are presented in columns 5–8. While the key coefficient in two of the four specifications is no longer statistically significant, in part because of a loss in power associated with estimating a large number of additional parameters, we still find evidence that those who have already voted answered more questions correctly in total. Thus, this test documents that those who have already voted are better informed and supports our social learning interpretation of the baseline results over an interpretation based on nationally available information.

3. Campaign Finance and Persuasion

A final possible alternative explanation for our results involves campaign contributions and expenditures. Suppose that surprising wins in early states lead to an increase in campaign contributions from influence-motivated contributors—those who want to alter the platform of the likely winner of the election.[28] Suppose further that these contributions

[27] Note that we did not use the first set of voters in our baseline analysis since they were not asked about their primary voting intentions.

[28] Grossman and Helpman (2001) distinguish between the influence motive for contributing and the electoral motive. According to the latter motive, candidates have fixed platforms and contributors give to enhance the electoral success of their preferred candidate.

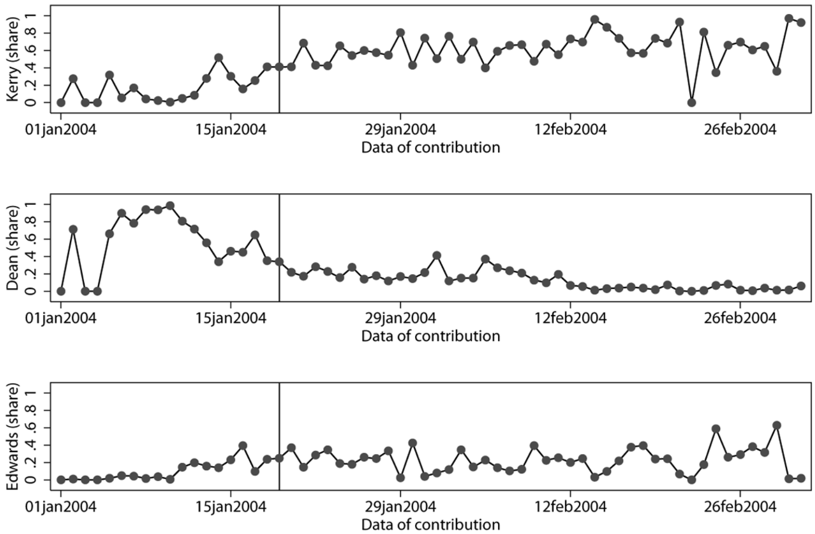

Fig. 8.—Candidate share of campaign contributions

are used to fund political advertisements that persuade late voters to support the candidate sponsoring the advertisement. In this case, our results, which document an increase in support among late voters for candidates who outperform expectations in early primaries, could be driven by this alternative explanation.

We address this alternative explanation by examining the timing of both campaign contributions and expenditures in the 2004 primaries. We first compute the share of contributions, as measured in dollars, flowing to each of the three candidates during the key primary season, using data from the Federal Election Commission. For consistency with the structural break analysis described above, we use data beginning on January 1, 2004. We also compute the share of advertising expenditures, as measured in the Wisconsin Advertising Data, by each of the three candidates during this same time period.[29]

As shown in figure 8, which focuses on campaign contributions, it is indeed the case that contributions to Kerry, relative to Dean, spiked following his surprising victory in Iowa on January 19. As shown in figure 9, there is also a spike in advertising expenditures by Kerry, relative to Dean, following his victory in Iowa. Importantly, however, this spike did

[29] In this data set researchers recorded all the political advertisements in 100 media markets from February 2003 to November 2004, coding characteristics for each ad including candidate, date, time of day, length, cost, tone, and other features. The cost of each ad was estimated on the basis of the time slot and geographic market.

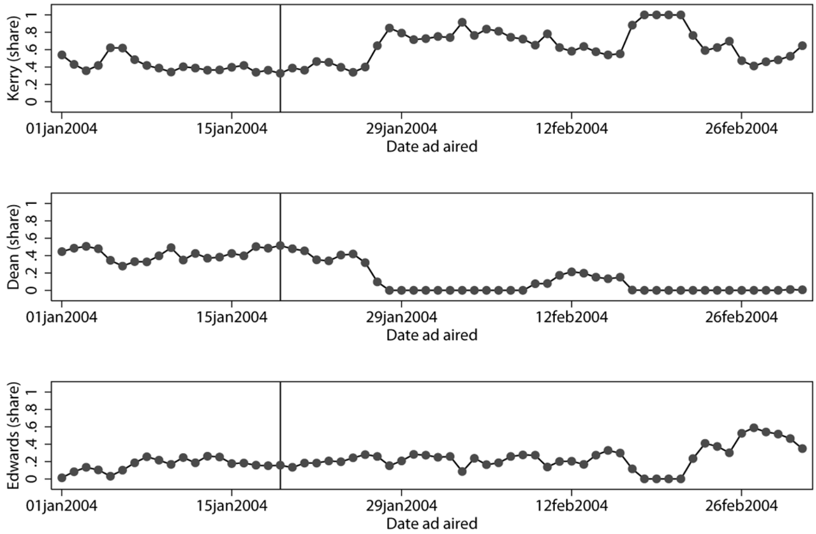

FIG. 9.—Candidate share of campaign expenditures

not occur until around January 26, a full 7 days after the Iowa caucus.[30] As shown previously in figure 7, however, the most likely date for a shift in support occurred just after the Iowa election. Given this discrepancy in the timing of the shift in campaign expenditures and support for the candidates, the increase in campaign expenditures cannot fully explain our baseline results.

To summarize, we have addressed and ruled out three alternative interpretations of our baseline results. We should emphasize that this analysis demonstrates that our results cannot be explained solely by these alternative interpretations. It is of course possible that multiple factors shape voter perceptions of candidates. For example, it could be that both national and state-level information plays an important role in sequential elections. Even if other motives are present, however, this analysis does provide strong support for the presence of social learning in this election.

V. Implied Voting Weights and the Allocation of Campaign Resources

Owing to these documented momentum effects, early voters may have a disproportionate influence over the selection of candidates. In this

[30] This lag presumably reflects the delay in processing contributions and in placing new advertisements in the field.

section, we attempt to quantify any such influence by measuring the weights, or number of votes, afforded to voters from different states during the primary season. Any overweighting of early voters associated with sequential voting represents a deviation 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 its magnitude. After measuring these weights, we then examine the consequences of any disproportionate influence for the distribution of campaign resources across states.

A. *Measuring Voting Weights*

We first use the estimated model to explicitly calculate the voting weights associated with sequential voting in the 2004 primary. Our measure of voting weights is based on the effect of changes in state-level preferences (η_{cs}) on national vote shares:

$$\omega_{cst} = \frac{1}{N_s}\sum_{r}\frac{\partial v_{crt}}{\partial \eta_{cs}} = \frac{1}{N_s}\left(\frac{\partial v_{cst}}{\partial \eta_{cs}} + \sum_{r\neq s}\frac{\partial v_{crt}}{\partial v_{cst}}\frac{\partial v_{cst}}{\partial \eta_{cs}}\right), \qquad (31)$$

where N_s is the number of states.[31] As shown, these weights depend on two effects. The direct effect ($\partial v_{cst}/\partial \eta_{cs}$) captures the change in candidate vote shares in the home state associated with a shift in preferences, and the second term captures the indirect multiplier effect, which equals zero in the absence of social learning. Even with social learning, however, the second term equals zero for voters from states voting in the final period ($t = 10$). For earlier states, by contrast, the multiplier effects will be positive in the presence of social learning.

We compute this derivative for each state voting on each of the primary dates ($t = 1, 2, \ldots, 10$) and present in figure 10 the average effect by time period.[32] We normalize the weight for the final time period to be equal to one. The plotted measures can thus be interpreted as the number of votes afforded to a representative voter from a state voting at time t. As shown, preferences of voters from the state of Iowa, the first state to vote, have roughly five times the influence relative to those

[31] In order to focus on differences in political power due to the timing of elections, the calculation of these weights abstracts from population differences across states. Given that delegates are allocated roughly proportional to population, however, incorporating state population would not significantly change the analysis.

[32] Specifically, we increase state-level preferences by one unit and recompute vote shares for that state, as expressed in eq. (9). In order to predict vote shares for subsequent states, we recompute the posterior mean quality, as expressed in eq. (11), and ultimately vote shares, as expressed in eq. (9). Note that simulating these vote shares requires explicit measures of the voting signal (θ_{cs}), which can be backed out of eq. (9) with information over state-level preferences (η_{cs}).

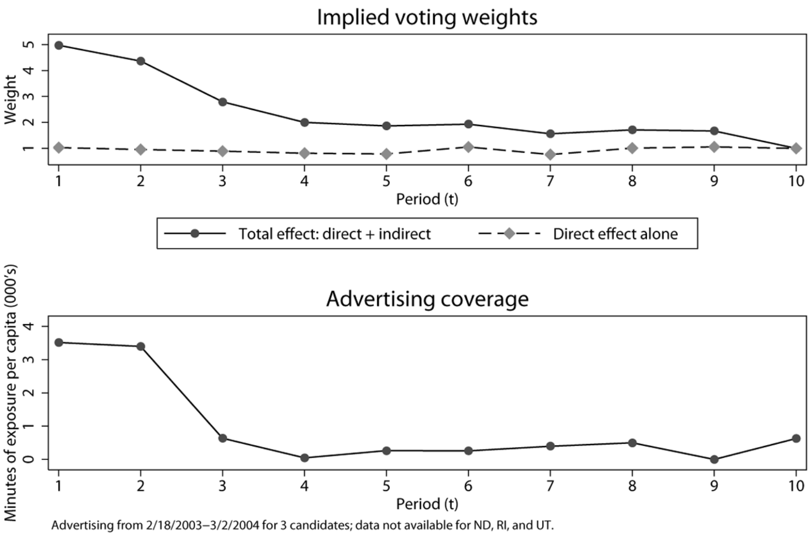

FIG. 10.—Advertising coverage versus implied voting weights

individuals voting on Super Tuesday ($t = 10$), and voters from New Hampshire have over four times the influence. For comparison purposes, we also plot voting weights in the absence of social learning, those associated with only the direct effect ($\partial v_{cst}/\partial \eta_{cs}$). These may vary across states because preferences enter vote shares in a nonlinear manner. As shown, however, there are not large differences across states. 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 due to momentum effects and thus represent a significant departure from "one person, one vote."[33]

B. Distribution of Campaign Resources

We next examine how this disproportionate influence of early voters affects candidates' incentives in terms of the allocation of campaign resources, as measured by advertising expenditures. To motivate our empirical exercise, we extend our theoretical model by examining a

[33] We also investigated a measure based on the effect of changes in state-level information (θ_{cs}) on candidate vote shares. We find that the information held by Iowa voters has roughly 20 times the influence of information held by Super Tuesday voters. These information-based weights are substantially larger than the preference-based weights, as described above, given that voters in late states place less weight on their own signal. This underweighting 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.

candidate charged with allocating a fixed budget among each of the states, where A_{cs} denotes the advertising expenditure in state s. In developing this simple model of the allocation of campaign expenditures, we make four key assumptions. First, the budget is fixed and does not depend on election outcomes.[34] Second, candidates commit to an allocation of campaign expenditures in advance of the primary season.[35] Third, candidates choose an allocation across states in order to maximize their overall vote share. Fourth, campaign spending in state s is productive via its shifting of the preferences of voters.[36] In particular, we assume that preferences depend on spending as follows:

$$\eta_{cs} = \eta_{cs}^{0} + f(A_{cs}), \tag{32}$$

where η_{cs}^{0} is the baseline support for candidate c in state s and $f(A_{cs})$ is an increasing and concave function normalized so that $f(0) = 0$.

Given these assumptions and recalling the definition of ω_{cst}, the optimal allocation of campaign spending can be characterized as follows:

$$\omega_{cst} f'(A_{cs}) = \lambda, \tag{33}$$

where λ represents the Lagrange multiplier on the candidate budget constraint. Thus, as weights increase, $f'(A_{cs})$ must decrease, which requires that A_{cs} increase. In short, the model predicts that advertising is increasing in preference weights.

To test this prediction, we use the Wisconsin Advertising Data to examine the geographic distribution of campaign spending, as measured by the exposure (minutes) of residents to candidate advertisements. An important consideration is that these data are measured at the designated market area (DMA) level. DMAs do not cover the entire United States and also often cross state boundaries. To convert these market-based measures to state-based measures, we take a weighted average across markets, where the weights are based on the number of residents of state s living in the market relative to the number of residents in state s covered by the Wisconsin Advertising Data.

As shown in the bottom panel of figure 10, advertising exposure in state s, aggregated across the three candidates (Dean, Kerry, and Edwards), is increasing in our estimated preference weights. Thus, the prediction of this simple model of the allocation of campaign resources is supported, and this pattern demonstrates that candidates appear to respond to the momentum effects documented in this paper.

[34] This assumption rules out the possibility that campaign contributions and hence budgets may change following election outcomes in early states.

[35] This assumption rules out the possibility that candidates may adjust their allocations in response to these election outcomes.

[36] This assumption rules out informative advertising, under which voters learn about candidate attributes, such as quality. See Prat (2006) for a review of these models.

VI. Counterfactual Simulations

One advantage of our empirical model is our ability to conduct counterfactual experiments. We examine two counterfactual scenarios: simultaneous voting and alternative ordering of states under a sequential system.

A. *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:

$$\ln (v_{cs1}/v_{0s1}) = \eta_{cs} + \alpha_1\theta_{cs} + (1 - \alpha_1)\mu_{c1}. \tag{34}$$

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 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 on the distribution of the realized signals.

Table 4 provides the key results from the actual sequential primary and the counterfactual simultaneous primary based on 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.[37] 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 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 Oklahoma, Tennessee, Virginia, and Wisconsin. While Kerry would have won a plurality with 40 percent of the delegates, he would not have won a majority. Similarly, the two-candidate simultaneous primary would have been much closer, with Edwards winning nine states. We do not wish to overemphasize the predictive nature of the results from this simulation for specific states, such as the surprising finding

[37] 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.

TABLE 4
Counterfactual Primary

			Sequential Primary (%)			Simultaneous (Three Way) (%)			Simultaneous (Two Way) (%)	
State	Period	Date (2004)	Dean	Edwards	Kerry	Dean	Edwards	Kerry	Edwards	Kerry
IA	1	1/19	21	36	43	21	36	43	46	54
NH	2	1/27	34	16	50	45	11	45	19	81
AZ	3	2/3	22	11	67	19	8	73	10	90
DE	3	2/3	14	15	70	22	7	71	8	92
MO	3	2/3	10	29	60	5	31	63	33	67
NM	3	2/3	23	16	60	24	15	61	20	80
OK	3	2/3	7	49	44	5	51	44	54	46
SC	3	2/3	6	56	38	3	46	51	47	53
MI	4	2/7	20	16	63	48	10	43	18	82
WA	4	2/7	35	8	57	88	0	11	4	96
ME	5	2/8	33	10	57	79	1	20	6	94
TN	6	2/10	6	37	57	1	79	20	79	21
VA	6	2/10	8	31	61	2	52	46	53	47
NV	7	2/14	19	12	70	62	2	36	4	96
WI	8	2/17	20	37	43	22	70	8	89	11
UT	9	2/24		35	65				46	54
CA	10	3/2		23	77				28	72
CT	10	3/2		29	71				36	64
GA	10	3/2		47	53				53	47
MA	10	3/2		20	80				93	7
MD	10	3/2		30	70				38	62
MN	10	3/2		35	65				68	32
NY	10	3/2		25	75				70	30
OH	10	3/2		40	60				78	22
RI	10	3/2		21	79				28	72
States won (*N*)			0	2	23	4	4	7	9	16
Delegates won (%)			7	28	65	29	31	40	46	54

that Edwards would have defeated Kerry in Massachusetts.[38] Rather, we hope that the heightened competition under a counterfactual simultaneous election helps to further reinforce our finding that Kerry benefited from the sequential primary system.

[38] 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 on signals at the beginning of the primary season relative to the end of the primary season.

TABLE 5
SEQUENTIAL ELECTIONS WITH RANDOMIZED ORDER

	Won Plurality of States (%)	Won Plurality of Delegates (%)
Two-way sequential:		
Kerry	92.3	89
Edwards	7.7	11
Three-way sequential:		
Kerry	99.1	94.9
Edwards	.5	3.8
Dean	.4	1.4

B. Alternative Sequential Schedules

Our second counterfactual election involves changes in the voting order under a sequential schedule. In the context of our model, there are two reasons why the sequence of elections can affect voting outcomes: (1) early signals have more weight and (2) preferences of early states have more weight. For these reasons, 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 and assigning each state the same signal it received in the actual sequence. 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. When delegate weights are used, 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 sizable 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.

VII. 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 exogenous here, one could introduce a media outlet that reports election results in early states to voters in late states. Second, one could model candidate entry and exit, which we have taken as given in this paper. Candidate exit would presumably depend on the degree of social learning, which may limit the ability of trailing candidates to make up lost ground in late states. A third extension would be to examine whether candidates alter their platforms in favor of issues that are most important to voters in early states. Whether or not such strategies are effective presumably depends on 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 overweighted, and state preferences may be misinterpreted as signals of quality; outliers in these early signals or early state preferences could lead to selection of a lower-quality candidate.

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. Our empirical application focuses on the 2004 Democratic primary. We find that Kerry benefited 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. Early states have up to five times the influence of late states, and candidates respond to these differences by funneling campaign expenditures into early 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 and also hope that the methods developed in this paper will be applied in other settings as well.

References

Adkins, Randall, and Andrew Dowdle. 2001. "How Important Are Iowa and New Hampshire to Winning Post-reform Presidential Nominations?" *Polit. Res. Q.* 54 (June): 431–44.

Ali, Nageeb, and Navin Kartik. 2010. "Observational Learning with Collective Preferences." Manuscript, Columbia Univ.

Banerjee, Abhijit. 1992. "A Simple Model of Herd Behavior." *Q.J.E.* 107 (August): 797–817.

Bartels, Larry. 1987. "Candidate Choice and the Dynamics of the Presidential Nominating Process." *American J. Polit. Sci.* 31 (February): 1–30.

———. 1988. *Presidential Primaries and the Dynamics of Public Choice.* Princeton, NJ: Princeton Univ. Press.

Battaglini, Marco. 2005. "Sequential Voting with Abstention." *Games and Econ. Behavior* 51 (May): 445–63.

Battaglini, Marco, Rebecca Morton, and Thomas Palfrey. 2007. "Efficiency, Equity, and Timing of Voting Mechanisms." *American Polit. Sci. Rev.* 101 (August): 409–24.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades." *J.P.E.* 100 (October): 992–1026.

———. 1998. "Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades." *J. Econ. Perspectives* 12 (Summer): 151–70.

Cai, Hongbin, Yuyu Chen, and Hanming Fang. 2009. "Observational Learning: Evidence from a Randomized Natural Field Experiment." *A.E.R.* 99 (June): 864–82.

Callander, Steven. 2007. "Bandwagons and Momentum in Sequential Voting." *Rev. Econ. Studies* 74 (July): 653–84.

Chamley, Christophe. 2004. *Rational Herds: Economic Models of Social Learning.* Cambridge: Cambridge Univ. Press.

Dekel, Eddie, and Michele Piccione. 2000. "Sequential Voting Procedures in Symmetric Binary Elections." *J.P.E.* 108 (February): 34–55.

Devenow, Andrea, and Ivo Welch. 1996. "Rational Herding in Financial Economics." *European Econ. Rev.* 40 (April): 603–15.

Foster, Andrew, and Mark Rosenzweig. 1995. "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture." *J.P.E.* 103 (December): 1176–1209.

Glaeser, Edward, and Bruce Sacerdote. 2007. "Aggregation Reversals and the Social Formation of Beliefs." Working Paper no. 13031, NBER, Cambridge, MA.

Grossman, Gene, and Elhanan Helpman. 2001. *Special Interest Politics.* Cambridge, MA: MIT Press.

Hansen, Bruce. 2001. "The New Econometrics of Structural Change: Dating Breaks in US Labor Productivity." *J. Econ. Perspectives* 15 (Fall): 117–28.

Klumpp, Tilman, and Mattias Polborn. 2006. "Primaries and the New Hampshire Effect." *J. Public Econ.* 90 (August): 1073–1114.

McKelvey, Richard, and Peter Ordeshook. 1985. "Elections with Limited Information: A Fulfilled Expectations Model Using Contemporaneous Poll and Endorsement Data as Information Sources." *J. Econ. Theory* 36 (June): 55–85.

Morton, Rebecca, and Kenneth Williams. 1999. "Information Asymmetries and Simultaneous versus Sequential Voting." *American Polit. Sci. Rev.* 93 (March): 51–67.

Munshi, Kaivan. 2004. "Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution." *J. Development Econ.* 73 (February): 185–215.

Prat, Andrea. 2006. "Rational Voters and Political Advertising." In *The Oxford Handbook of Political Economy*, edited by Barry R. Weingast and Donald A. Wittman. Oxford: Oxford Univ. Press.

Smith, Lones, and Peter Sorensen. 2000. “Pathological Outcomes of Observational Learning.” *Econometrica* 68 (March): 371–98.

Steger, Wayne. 2008. “Forecasting the Presidential Primary Vote: Viability, Ideology and Momentum.” *Internat. J. Forecasting* 24 (April–June): 193–208.

Steger, Wayne, Andrew Dowdle, and Randall Adkins. 2004. “The New Hampshire Effect in Presidential Nominations.” *Polit. Res. Q.* 57 (September): 375–90.

Strumpf, Koleman. 2002. “Strategic Competition in Sequential Election Contests.” *Public Choice* 111:377–97.

Welch, Ivo. 1992. “Sequential Sales, Learning, and Cascades.” *J. Finance* 47 (June): 695–732.

———. 2000. “Herding among Security Analysts.” *J. Financial Econ.* 58:369–96.

Wolfers, Justin, and Eric Zitzewitz. 2004. “Prediction Markets.” *J. Econ. Perspectives* 18 (Spring): 107–26.