Learning and Coordination in the Presidential Primary System

George Deltas[*] Helios Herrera[†] Mattias K. Polborn[‡]

April 22, 2011

Abstract

To analyze the advantages and disadvantages of the U.S. presidential primary system, we develop a model in which candidates with different policy positions and qualities compete for the nomination, and voters are uncertain about the candidates' valence. This setup generates two effects, which we call vote-splitting (i.e., several candidates in the same policy position compete for the same voter pool) and voter learning (as the results in earlier elections help voters to update their beliefs on candidate quality). We analyze how different temporal organizations of primaries affect the trade-off between vote-splitting and voter learning: Sequential voting minimizes vote-splitting in late districts, but voters may coordinate on the wrong candidate. We structurally estimate the model using the 2008 Democratic presidential primaries. Using the parameter estimates, we conduct policy experiments such as replacing the current system with a simultaneous system or other proposed systems.

Keywords: Voting, Presidential primary elections, simultaneous versus sequential elections.

[*]Department of Economics, University of Illinois, 1407 W. Gregory, Urbana, IL 61801 USA, E-mail: deltas@uiuc.edu

[†]School of International and Public Affairs (SIPA), Columbia University. Email: hhh2108@columbia.edu

[‡]Corresponding author. Department of Economics and Department of Political Science, University of Illinois, 1407 W. Gregory, Urbana, IL 61801 USA, E-mail: polborn@uiuc.edu.

1 Introduction

One of the fundamental questions in the analysis of politics is how the institutions of the political system influence election results and policy outcomes. An understanding of such effects should ideally guide the institutional designers (such as a constitutional convention) in their choice of the political system. Clearly, this approach to institutional design generally suffers from an important problem: Institutional arrangements are often fixed in a constitution for a long term, so once we observe how a particular political system works in practice, it has already become hard to change. In this article, we analyze a particular feature of the U.S. political system that does not suffer from this conundrum: The selection of candidates for the U.S. presidential election by means of a sequence of elections within each political party, the "primaries",[1] is not enshrined in the U.S. Constitution, and the structure of the sequence has actually changed substantially in the past and is likely to continue to be modified in the future.

The nomination process is one of the most controversial institutions of America's contemporary political landscape: Its sequential structure is perceived as inherently "unfair" because it shifts too much power to voters in early primary states. For this reason, many states have shifted their primaries earlier and earlier over the last several election cycles, while the national parties have tried to steer against this movement. For the 2008 cycle, both the Democratic and the Republican National Committee chose rules that prohibited all but a few states to hold their primaries before February 5th. Florida and Michigan violated these rules and were punished by the DNC and RNC by taking away half of their delegates at the convention.[2] Thus, it appears that states have a strong interest in voting early, at least enough to risk such a punishment. Moreover, if the national parties' decisions reflect their interest in the efficiency of the whole nomination process, then the states' "race to the front" must be inefficient. There are at least three different primary organization systems that have attracted considerable support among both commentators and politicians. The main alternatives to the current status quo of a sequential system appear to be a nationwide primary to be held on the same day, and a proposal by the National Association of the Secretaries of State (NASS) for regional primaries. According to the NASS proposal (see Stimson (2008)), Iowa and New Hampshire would always vote first, followed by four regional primaries (for the East, Midwest, South and West regions) scheduled on the first Tuesday in March, April, May or June of presidential election years. The sequence of the four regions would rotate over a 16-year cycle. In our framework, we can analyze (i) under which circumstances the temporal organization makes a difference for who wins the nomination, and (ii) whether such a change is beneficial for voters from an ex-ante or utilitarian perspective.

[1]Different states have their presidential nomination elections organized as either primaries or caucuses. Since we are only interested in the temporal organization of the entire nomination process, we will, in a slight abuse of terminology, call all of these contests "primaries."

[2]Throughout the primary process, the Democratic National committee even threatened to take away all of Florida's and Michigan's delegates, but then reduced the size of the penalty to one-half.

We argue that the primary system has to provide a trade-off between two different and potentially conflicting objectives, learning about candidate quality and voter coordination. To better understand our basic argument, consider the following – only half-fictional – example of a nomination contest with three serious contenders at the time of the first elections, whom we call C, E and O. These candidates differ in some characteristics that are relevant for voters. First, candidate C has experience in Washington and would know on day 1 where the light switches are in the White House, while candidates E and O run as "Washington outsiders" or "change candidates". Suppose that, ceteris paribus, some voters prefer a candidate with Washington experience, while others (the "change voters") prefer an outsider. In addition, there is uncertainty about the valence of each candidate. If the primary elections were to take place simultaneously in all states, then it is quite plausible that C wins most states, as E and O split the change voters.

In contrast, in a sequential system, change voters in states that hold their primaries after the first ones can observe the early election results and vote accordingly; also, in expectation of such coordination, the trailing candidate may drop out early. For example, if O gets more votes than E in the early elections, then even voters with ranking $E \succ O \succ C$ may vote for O, because they have determined that E has no chance of winning, and among the remaining relevant candidates, they prefer O. In this case, O will win the nomination if a majority of the electorate prefers him to C.

Such voter migrations between candidates may be crucial for election outcomes. For example, Moulitsas (2008a) cites a Rasmussen poll for Missouri from January 31 (the last one conducted with Edwards in the mix) before the primary one week later. The preference numbers in the Rasmussen poll were Clinton 47, Obama 38, Edwards 11, while the actual election results were Obama 49.3, Clinton 47.1, Edwards 1.7. These numbers suggest that a majority of Edwards supporters migrated to Obama, after Edwards dropped out of the race. Similarly, in a 12/26-30, 2007 poll by Opinion Research Corp for CNN (cited by Moulitsas (2008b)), 36% of Iowa Democrats polled declare that Edwards was their second choice, 25% name Obama, but only 11% name Clinton as their second choice. Since all three candidates were very close in terms of *first* preferences, this suggests that most Obama and Edwards supporters had the respective other candidate as their second preference.

The benefit of a sequential system in our example is that, in most districts, the change voters do not split their votes, thus increasing the likelihood that a change candidate wins. There is, however, also a disadvantage when voters are uncertain about candidate valences: Conditioning coordination on only one or few initial elections raises the possibility that the weaker change candidate comes out on top, and if such an early electoral mistake occurs, it cannot be corrected in the remaining districts precisely because of coordination resulting in candidate withdrawal. The objective of our model is to provide a formal framework for the analysis of the trade-off between coordination and voter learning.

2 Coordination, learning, and the trade-off between them

Learning about candidate quality is a very relevant problem in presidential primaries: While most candidates are accomplished politicians such as governors or members of Congress, very few of them are already household names for a truly national audience. Moreover, in addition to past achievements, voters also care about how candidates acquit themselves under the pressure of an intense campaign under the spotlight of the national media. Thus, learning about candidate quality naturally proceeds throughout the entire primary process. While, in principle, all voters agree on the desirability of nominating "the best" candidate, imperfect information implies that they may have different ideas about who the best candidate is, in particular early in the primary process.

While some of the candidates' characteristics can be thought of as pure valence (in the sense that all voters agree that they want to nominate the *best* possible candidate in those categories), there are also differences between candidates that are better thought of in terms of horizontal differentiation. For example, when candidates differ in ideological positions such as moderates and conservatives in the GOP, then different voters have conflicting preferences even if all information about candidates is common knowledge. For our purposes, it is immaterial whether the voters' preferences over positions are "sincere" or follow some strategic calculations based on the recognition that the nominee has to compete in a general election against the nominee of the other party (for example, a very conservative, but risk-averse voter might actually have a preference for nominating a moderate Republican as a candidate if he believes that the moderate's higher likelihood of winning in the general election relative to a conservative compensates for the less preferred policy position). In our formal model, we take voters' preferences for one of the positions as fixed and exogenously given.

However, we do not think of our horizontal dimension as necessarily exclusively capturing actual "policy" differences in a traditional sense. For example, one can argue that the policy positions of the three main candidates in the 2008 Democratic race on actual political issues were very close to each other. What matters for our argument is entirely that voters *perceive* a difference that is important to them between different sets of candidates, and the opinion polls cited above clearly indicate that Democratic primary voters perceived Edwards and Obama to be relatively similar to each other, and relatively different from Clinton.[3]

Our theoretical model, set up in Section 4 and analyzed in Section 5, develops the simplest framework in which the issues of learning and coordination can arise and interact with each other, and provides some guidance as to which factors affect this trade-off. The net effect can go in either direction, so that the question of the optimal voting system is a quantitative one. In Section 6, we estimate the structural parameters of our theoretical model using data from the 2008 Democratic primary. The estimated parameter values show that both key features of the theory (slow voter learning about candidate valence, and unequal substitutability of candidates with different political positions) are quantitatively

[3]One can also think that voters may have different views on the desirability of "political dynasties" (see Dal Bo, Dal Bo, and Snyder (2009)).

important. In the first primary contest, the variability of the voters' estimate of candidate valence is only about a third of the *true* valence variability (the reason is that signal quality is weak, and updating is thus not very responsive to the received signal in the first district). Moreover, the horizontal differences between candidates appear to be very important for voters' choice.

However, the main point of the estimation is not to "test" the model in a classical sense. Rather, the purpose is to develop reasonable starting values for our institutional simulations in Section 7. All of our simulations consider races with three candidates competing for the nomination, two of whom share the same political position. We compute the distribution of election outcomes under several different sequencing scenarios of state voting. The first scenario assumes that all 50 states vote simultaneously; the second assumes that states vote sequentially and all three candidates remain in the race until the end; the third assumes that states vote sequentially but the candidate perceived as weaker (of the two candidates who share the same political position) drops out after the fifth state votes. Scenario 4 is modeled after the sequence in the 2008 Democratic race, and scenario 5 is the NASS proposal with dropout after the first regional contest.

Our results show that a sequential election with all candidates remaining in the race results in the highest expected valence and the highest probability that the Condorcet winner is elected, while a completely simultaneous election does worst. The other setups yield intermediate results, with the NASS proposal coming in as a very close second to completely sequential primaries. In fact, the impressive performance of the NASS proposal is particularly relevant because a completely sequential primary system with three candidates in the race for a very long time may not be practically feasible. After all, it is not just up to the candidates to decide when they want to give up, but also, voters may decide that only one of the two candidates in the shared position has a realistic probability of winning, and they may effectively eliminate a contender as a "serious candidate" even if he officially stays in the race. In contrast, it is quite plausible that candidates would remain in the race until after the first regional contest under the NASS proposal.

The intuitive reasons for the simulation results are as follows. A simultaneous election makes the nomination of the sole candidate very probable, independent of this candidate's valence, as vote-splitting between the two candidates in the same position is usually substantial and cross-over voting (i.e., voters with a preference for one position voting for a candidate in the other position) is only moderate. In the sequential election with all candidates staying in the race, there is some vote splitting in all districts, but the extent of it is sufficiently muted to be considerably less detrimental to the winning chances of the better of the two candidates in the same position. In the third scenario in which one of the two candidates who share the same position (namely, the one who is perceived as weaker by voters after the fifth district) drops out after the fifth state, the vote-splitting problem is reduced even further, but this comes at a substantial cost, as there is a distinct possibility that the wrong candidate is eliminated (i.e., the candidate whose true valence is higher than the one of his competitor). Consequently, expected valence decreases in this regime, relative to a completely sequential regime without dropout. We also find that the optimal dropout time from a social point of view is quite late (approximately after 30 districts), but that the overwhelming part of the expected utility increase can be achieved by moving to a dropout after about 15 states. This is the reason why the NASS proposal does very well from a welfare point of view in our simulations. Assuming that all candidates stay in the race until after the first large regional contest, there are sufficiently many early elections to be relatively confident that the strongest candidates survive, yet vote splitting is absent in three out of four large regional contests. Relative to a primary structure modeled after the 2008 temporal structure that the Condorcet winner wins increases from 59.9% to 73.4%.

Our baseline scenario takes the point estimates that we obtained in our estimation of the 2008 primaries, but we then check for robustness by increasing or decreasing each parameter value by one standard deviation while keeping the other parameter values constant. None of these changes changes the ranking of the different primary systems relative to the baseline case. This is important: While our estimation technique implies results about the ex-ante quality distribution from which candidates are drawn (as well as the distribution of signals), there is, of course, no guarantee that these distributions are constant throughout time. Thus, it is reassuring that our central result – the comparison between different primary systems – appears very robust with respect to reasonable variation in the parameter values.

3 Related Literature

Several studies analyze the relation between voters' expectations of which candidates will do well and their preference for these candidates. The study closest to our focus on the role of early primaries as a coordination device is Bartels (1987), who analyzes the 1984 Democratic presidential primary. Bartels (1987, pp.13) describes the coordination process of those Democratic voters unhappy with the establishment candidate, as follows.

> At the beginning of the 1984 primary season, the question facing prospective voters was whether or not to support the obvious front-runner, Walter Mondale. Those who were most predisposed to support Mondale (on the basis of issue preferences [...]) would do so without undue soul-searching. On the other hand, a fair number of Democrats who were lukewarm (or worse) about Mondale's candidacy may at least have entertained the possibility of supporting a different candidate. Their problem was to decide which alternative, if any, to turn to.
>
> Having framed the problem in this way, we may ask ourselves what a prospective voter with an eye out for an alternative to Mondale would have been likely to know about the other candidates in the race. At the beginning of the campaign, the best answer is probably "very little". But Hart's second-place finish in Iowa, followed by his dramatic upset victory in New Hampshire changed that. By the end of February, our prospective voter was quite

> likely to know at least one thing about at least one challenger: that Gary Hart was out there, an alternative to Mondale with significant popular support, [suggesting that] a vote for Hart would not be wasted.

In the empirical part of the paper, Bartels does not focus on this coordination aspect (i.e., Hart versus other non-Mondale candidates), but rather analyzes the dynamic aspects of how expectations about the candidates' winning chances influenced voters' preferences. Other studies analyzing similar relationships include Bartels (1985) for the 1980 Democratic primaries and Kenny and Rice (1994) for the 1988 Republican primary, but all of these focus implicitly on a two-candidate framework. An exception to this is Knight and Schiff (2007), who provide both a theoretical model and an empirical study of the 2004 Democratic primary. In contrast to our model, though, their model is not designed to analyze the optimality of different temporal structures of the primary process, and also does not have a trade-off between coordination and learning.

Most of the theoretical literature on primaries has focused on elections with two alternatives (in which, naturally, the issue of coordination does not arise). Dekel and Piccione (2000) analyze a model of sequential elections in which sophisticated voters try to aggregate their private information through voting. While, in principle, more information is available for voters in later elections, they show that every equilibrium of the simultaneous game is also an equilibrium of the sequential game, regardless of the sequence. The intuition for this result is that strategic voters know that their vote only matters if they are pivotal, and hence they behave as if they knew that all other voters are evenly divided between the two candidates. Thus, it does not matter for the election outcome which candidate is supported by the early voters.

While Dekel and Piccione (2000) show that the sequential structure does not allow voters to improve upon the information aggregation result that can be obtained with simultaneous elections, Ali and Kartik (2006) show that there are equilibria of the sequential game that do not correspond to equilibria of the simultaneous game. In particular, they construct an equilibrium in posterior-based voting in the context of a sequential election. In this equilibrium, if other voters play history dependent strategies, then it is individually optimal for each and every voter to do so as well. The information aggregation properties of such a herding equilibrium are worse than those of the equilibrium analyzed by Dekel and Piccione (2000). In summary, with two candidates, the design of a sequential primary system appears ill-advised from a social point of view, as the expected voting outcome is either the same or worse than in a simultaneous system.

Callander (2007) studies a sequential voting model which, on top of common value preferences, voters have an exogenous desire to vote for the winning candidate. Callander obtains equilibria which, at some point in the sequential election process, display bandwagon effects with certainty because the desire to conform eventually dominates information based voting. Battaglini (2005) shows that, when voting is costly, the set of equilibria under simultaneous and sequential models are generically disjoint. In a related experimental paper, Battaglini, Morton, and Palfrey (2007) explore empirically the implications of voting costs in sequential and simultaneous elections.

Schwabe (2010) compares a fully sequential primary system with a more simultaneous system in a model in which voters in each state can learn the candidates' valence only if the candidates have funds for their campaign. Both lobbies and voters want to select the best candidate in a common values setting, and lobbies must decide which candidates to fund, and in which state races. He finds that it is optimal for learning to have a primary system in which many states have simultaneous elections at the beginning of the contest (such as a Super-Tuesday).

Klumpp and Polborn (2006) analyze a contest model of sequential primaries in which two competing candidates choose how to allocate their campaign expenditures on the different states that hold their elections sequentially. In equilibrium, candidates allocate a large portion of their budget to the initial states. There is momentum in the sense that the currently leading candidate has an increased equilibrium probability of winning the next election. From a normative perspective, they show that a sequential organization of primaries has the advantage of leading to lower expected expenditures than a simultaneous system.

Fundamentally, our paper asks which primary structure would be socially optimal for voters in a world where both learning about candidate quality and coordination matter for voting outcomes. Conceptually, our paper is therefore related to a small literature that analyzes institutional design questions. For example, Diermeier and Myerson (1999) analyze the internal organizational choices of legislatures, taking as given the fundamental constitutional setup (i.e., the number and identity of the players involved in legislation). Similarly, we take as given that the candidates for the Presidential election are chosen through some sort of elections that involve the rank-and-file members of each party as voters, and analyze the consequences of different temporal voting structures in this general framework.

4 The model

Let $\mathcal{J} = \{1, \ldots, J\}$ denote the set of candidates who compete for their party's nomination, and let j denote a typical candidate. The set of states (i.e., electoral districts) is $\{1, \ldots, S\}$, with typical state s. We assume for simplicity that the number of states, S, is large and that all of them have the same size. States vote sequentially, though some states may vote at the same time. Voters can observe the outcome in all states that voted before their own state. The set of candidates in later elections may be a strict subset of the set of candidates in early elections, as some candidates may drop out.

Candidates differ in two dimensions. First, parameter v_j measures Candidate j's valence (which is a characteristic like competence appreciated by all voters). Second, there is a policy issue on which candidates have either position 0 or 1. Without loss of generality, we assume that the first j_0 candidates are fixed at $a_j = 0$, while the other $j_1 = J - j_0$ candidates are fixed at $a_j = 1$.

The policy dimension is meant to capture the notion that some candidates are quite similar to each other and hence close substitutes for most voters, while there is a more substantial difference to some other candidates. Other issues are treated stochastically via a composite preference shock, as detailed below.[4] Voter i's utility from a victory of Candidate j is

$$U_j^i = v_j - \lambda|a_j - \theta^i| + \varepsilon_j^i. \tag{1}$$

Here, θ^i is voter i's preferred position on the policy issue, and λ measures the weight of the policy issue relative to valence. The proportion of the total population in district s with preference for $a = 1$ is $\mu^s \in (0, 1)$, which is common knowledge among all players.

The last term, ε_j^i, drawn from $N(0, {\sigma_\varepsilon}^2)$, is an individual preference shock of voter i for Candidate j, as in probabilistic voting models.[5] A possible interpretation of this term is that candidates also differ in a large number of other dimensions for which voters have different preferences. The policy dimension modeled explicitly ($a_j = 0$ or $a_j = 1$) should then be understood as the most important dimension.

Voters are uncertain about the candidates' valences. Specifically, each candidate's valence is an independent draw from a normal distribution $N(0, {\sigma_v}^2)$. Voters cannot observe v_j directly. Instead, voters in state s observe a signal $Z_j^s = v_j + \eta_j^s$, where the additional term for Candidate j, η_j^s, is an independent draw from a normal distribution $N(0, {\sigma_\eta}^2)$. Note that η_j^s is a state-specific (as opposed to voter-specific) observation error. The idea is that voters in the same state receive their news about the candidates from the same local news sources so that errors, if any, are not individual-specific.[6] If, instead, observation error terms were individual-specific, then the true valence of candidates would be perfectly known after the election results of the first state, which appears unrealistic.

Also, we assume that signals are state-specific rather than national, so that election results are informative for voters in later states. Even if information arrives from national news media, it appears likely that voters are particularly attentive before a state-wide election, while most voters who live in states that will only vote in a month or so may forget today's news stories before they decide about whom to vote for. Also, information may be interpreted differently in different states. If, instead, all information was broadcast nationally to all voters, then election results would not be incrementally informative about candidate valence.

Given their own signal, and possibly the election results in earlier states (from which the signals in those earlier states can be inferred), voters rationally update their belief about the valence of candidates. Let $\hat{v}_j^s$ denote the valence of Candidate j that is expected by voters in district s. Let J^t be the set of "relevant" candidates in period t elections. We assume that the set J^t is known to all voters, and that each voter votes sincerely given this set of relevant candidates.[7] That is, voter i in district s (which

[4] The assumption that policy differences can be expressed in binary form follows Krasa and Polborn (2010), and the assumption that there is only one major fixed characteristic in which candidates differ is very helpful for the empirical analysis.

[5] See, e.g., Lindbeck and Weibull (1987), Coughlin (1992) or Persson and Tabellini (2000) for a review of the various developments of this literature.

[6] Of course, in reality, there are plausibly both common and idiosyncratic observation errors. To simplify the model and gain some tractability, we focus on the state-specific observation error.

[7] In elections with more than two candidates, there are generally very many Nash equilibria in undominated strategies.

votes at time t) votes for Candidate j if and only if

$$j \in \arg\max_{j' \in J^t} \hat{v}^s_{j'} - \lambda|a_{j'} - \theta^i| + \varepsilon^i_{j'}.$$ [8]

Thus, the set of relevant candidates captures our notion of coordination among candidates and/or voters in later primaries. In practice, there are two ways how a candidate who participated in earlier rounds of elections may drop out of the set of relevant candidates, either by being generally considered to be a lost cause by all or most voters, or by officially withdrawing from the race. It is important to stress that a sequential structure of primaries *facilitates* coordination (and the particular form of coordination that we focus on is, in our opinion, fairly natural), but, of course, sequential primaries do not *enforce* any particular form of coordination. We discuss this issue further below.

5 Analysis

5.1 Roadmap

Ideally, we would want to solve for the equilibrium in all possible primary structures and then find the optimal primary structure for each set of parameters. Clearly, this model is much too complex to allow for such a strategy. However, for a special case of the model, we obtain an analytical solution that provides some intuition for the trade-off between learning and coordination. In this scenario, presented in detail in the supplemental appendix, we analyze two candidates in position 1 competing with a single candidate in position 0, and assume that λ is so large that all voters vote for a candidate who is in the same position as they themselves, and that μ is constant across districts.

When μ, the share of voters who prefer position 1, is between $1/2$ and $2/3$, vote splitting in simultaneous primaries may have the effect that the minority candidate (i.e., the Condorcet loser in this setting) wins. In a sequential primary system, one of the two candidates in the majority-preferred position wins, but the probability that it is the higher valence candidate is bounded away from 1, and is decreasing in σ_η and increasing in σ_v. Thus, sequential primaries solve the coordination problem, but sometimes at the expense of worse learning (i.e., a lower chance of selecting the best candidate). As we show, there are parameter values such that that sequential system is better for voters than a simultaneous one, and other parameters such that the reverse relation holds.

While these theoretical results are instructive, they suffer from the fact that we need to impose relatively restrictive assumptions in order to keep that scenario analytically tractable with a closed-form solution. Even then, the ranking between the various electoral systems would be an empirical question, since it depends on parameters. The results that we present in the main text are therefore

However, sincere voting is a standard assumption in the literature for multicandidate elections, and also appears to capture voter behavior in many elections (see Degan and Merlo (2006)).

[8] Since the distribution of ε is continuous, the measure of voters who are indifferent between 2 or more candidates is equal to zero, so it is irrelevant for the election outcome how those voters behave.

based on (i) developing a theoretical foundation for an empirical strategy to estimate the parameters of the model in a particular primary, (ii) to use the estimated parameters to conduct policy experiments such as changing the temporal structure of the primaries and (iii) a sensitivity analysis that inquires how robust the results are to changes of the parameters.

5.2 Updating and vote-shares

We now focus on deriving theoretical foundations of voter updating about candidate valence and vote-share determination for the empirical analysis in Section 6. In particular, we show how vote shares in the entire sequence of elections are determined given the fundamentals (candidate valences, the set of competing candidates, and voter initial beliefs) and the signals that voters observe over the course of the campaign.

We start with an analysis of the vote shares of candidates in district s, given that the beliefs of voters in district s are given by the vector $\hat{v}^s = (\hat{v}_1^s, \hat{v}_2^s, \dots, \hat{v}_J^s)$. We then turn to the determination of $\hat{v}^s$. Let J_0^s denote the set of candidates with position 0 who are running in district s, and J_1^s the set of candidates with position 1 who are running in district s. Beliefs about candidate valence, together with an individual's idiosyncratic preferences, determine the candidate that he will vote for. In particular, a voter of type θ votes for Candidate $j \in J_0^s$ if and only if, for all $j' \neq j$,

$$\hat{v}_j^s + \varepsilon_j - \lambda d(j,\theta) > \hat{v}_{j'}^s + \varepsilon_{j'} - \lambda d(j',\theta), \tag{2}$$

where $d(j,\theta)$ measures the distance between Candidate j and voter type θ (i.e., $d = 0$ if voter type and candidate agree, and $d = 1$ when they disagree). For a given ε_j, (2) is satisfied if and only if

$$\varepsilon_{j'} < \hat{v}_j^s - \hat{v}_{j'}^s + \varepsilon_j - \lambda[d(j,\theta) - d(j',\theta)] \text{ for all } j' \neq j\,. \tag{3}$$

First consider a voter of type $\theta = 0$. Since the ε's are distributed independently, the probability that such a voter votes for Candidate j is

$$\prod_{J_0^s \setminus \{j\}} \Phi\left(\frac{\hat{v}_j^s - \hat{v}_{j'}^s + \varepsilon_j}{\sigma_\varepsilon}\right) \cdot \prod_{J_1^s} \Phi\left(\frac{\lambda + \varepsilon_j + \hat{v}_j^s - \hat{v}_{j'}^s}{\sigma_\varepsilon}\right). \tag{4}$$

Integrating over the possible realizations of ε_j shows that the proportion of type 0 voters who vote for Candidate $j \in J_0^s$ is

$$\int_{-\infty}^{\infty} \prod_{J_0^s \setminus \{j\}} \Phi\left(\frac{\hat{v}_j^s - \hat{v}_{j'}^s + \varepsilon_j}{\sigma_\varepsilon}\right) \cdot \prod_{J_1} \Phi\left(\frac{\lambda + \varepsilon_j + \hat{v}_j^s - \hat{v}_{j'}^s}{\sigma_\varepsilon}\right) \cdot \phi_\varepsilon(\varepsilon_j) d\varepsilon_j. \tag{5}$$

Similarly, the share of type 1 voters who vote for Candidate j is

$$\int_{-\infty}^{\infty} \prod_{J_0^s \setminus \{j\}} \Phi\left(\frac{\hat{v}_j^s - \hat{v}_{j'}^s + \varepsilon_j}{\sigma_\varepsilon}\right) \cdot \prod_{J_1^s} \Phi\left(\frac{-\lambda + \varepsilon_j + \hat{v}_j^s - \hat{v}_{j'}^s}{\sigma_\varepsilon}\right) \cdot \phi_\varepsilon(\varepsilon_j) d\varepsilon_j. \tag{6}$$

The total vote share of Candidate $j \in J_0^s$ is then given by the weighted average of (5) and (6), where the weights are $(1-\mu^s)$ and μ^s. In an analogous way, the total vote share of Candidate $j \in J_1^s$ can be derived. Thus, the vote shares of candidates in state s satisfy the following equation system

$$
\begin{aligned}
W_j^s &= (1-\mu^s)\int_{-\infty}^{\infty}\prod_{J_0^s\setminus\{j\}}\Phi\left(\frac{\hat{v}_j^s-\hat{v}_{j'}^s+\varepsilon_j}{\sigma_\varepsilon}\right)\cdot\prod_{J_1^s}\Phi\left(\frac{\lambda+\varepsilon_j+\hat{v}_j^s-\hat{v}_{j'}^s}{\sigma_\varepsilon}\right)\cdot\phi_\varepsilon(\varepsilon_j)d\varepsilon_j + \\
&\quad \mu^s\int_{-\infty}^{\infty}\prod_{J_0^s\setminus\{j\}}\Phi\left(\frac{\hat{v}_j^s-\hat{v}_{j'}^s+\varepsilon_j}{\sigma_\varepsilon}\right)\cdot\prod_{J_1^s}\Phi\left(\frac{-\lambda+\varepsilon_j+\hat{v}_j^s-\hat{v}_{j'}^s}{\sigma_\varepsilon}\right)\cdot\phi_\varepsilon(\varepsilon_j)d\varepsilon_j\ ,\forall j\in J_0^s \\
W_j^s &= (1-\mu^s)\int_{-\infty}^{\infty}\prod_{J_0^s}\Phi\left(\frac{-\lambda+\hat{v}_j^s-\hat{v}_{j'}^s+\varepsilon_j}{\sigma_\varepsilon}\right)\cdot\prod_{J_1^s\setminus\{j\}}\Phi\left(\frac{\varepsilon_j+\hat{v}_j^s-\hat{v}_{j'}^s}{\sigma_\varepsilon}\right)\cdot\phi_\varepsilon(\varepsilon_j)d\varepsilon_j + \\
&\quad \mu^s\int_{-\infty}^{\infty}\prod_{J_0^s}\Phi\left(\frac{\lambda+\hat{v}_j^s-\hat{v}_{j'}^s+\varepsilon_j}{\sigma_\varepsilon}\right)\cdot\prod_{J_1^s\setminus\{j\}}\Phi\left(\frac{\varepsilon_j+\hat{v}_j^s-\hat{v}_{j'}^s}{\sigma_\varepsilon}\right)\cdot\phi_\varepsilon(\varepsilon_j)d\varepsilon_j\ ,\forall j\in J_1^s \qquad (7)
\end{aligned}
$$

To compute the vote shares given the sequence of signals and the set of candidates competing in every state, we now need to determine the ex-ante beliefs about candidate valences for the voters in each state. Consider the situation in the state(s) that vote first. Voters know that candidate valences are drawn from $N(0,\sigma_v^2)$. In addition, voters in state s receive a state-specific signal Z_j^s that is normally distributed with expected value v_j (i.e., the true valence of Candidate j) and variance σ_η^2. Voters can use Bayes' rule to derive the ex-post density of the candidate's valence, which is again the density of a normal distribution, but now with expected value

$$\hat{v}_j^s = \frac{\sigma_v^2}{\sigma_v^2+\sigma_\eta^2}Z_j^s \tag{8}$$

and variance

$$(\sigma_{v_j}^s)^2 = \frac{\sigma_v^2\sigma_\eta^2}{\sigma_v^2+\sigma_\eta^2}. \tag{9}$$

For any subsequent state, if a voter has an ex-ante belief (i.e., before seeing his own state-specific signal) about candidate j's valence that is distributed according to $N(\hat{v}_{j0},\sigma_{j0}^2)$ and receives a state-specific signal Z_j^s, the ex-post density of the candidate's valence is again the density of a normal distribution, but now with expected value

$$\hat{v}_j^s = \frac{\sigma_\eta^2}{\sigma_{j0}^2+\sigma_\eta^2}\hat{v}_{j0} + \frac{\sigma_{j0}^2}{\sigma_{j0}^2+\sigma_\eta^2}Z_j^s \tag{10}$$

and variance

$$(\sigma_{v_j}^s)^2 = \frac{\sigma_{j0}^2\sigma_\eta^2}{\sigma_{j0}^2+\sigma_\eta^2}. \tag{11}$$

Applying (11) recursively shows that the coefficient of the candidate valence signal in state j in (10) takes the same value for all candidates. Thus, an increase in the values of all valence signals by a constant increases ex-post valence estimates of all candidates by the same amount. Since vote shares are determined by differences in ex-post valences, they are unaffected. Therefore, signal realizations can be normalized by subtracting a constant so that the signal of the first candidate is equal to zero.

We now turn to the calculation of $\hat{v}_{j0}$. If voters can infer the signals observed in all prior states, then they can obtain v_{j0} (and σ^2_{j0}) by applying (8) and (9) to the states that vote in the first round, and (10) and (11) sequentially to all states that vote subsequently.[9] Proposition 1 shows that this approach is indeed feasible: Observing the outcome in state s allows voters in later states to essentially recover the estimated vector of candidate valences in state s, and thus, as Corollary 1 shows, the valence signals Z^s_j. This method can be applied recursively to recover the valence signals in all states that vote earlier.

Proposition 1 *Consider* (7) *as an equation system in* $\{\hat{v}^s_1, \hat{v}^s_2, \ldots\}$*. There exists a unique vector of valence values* $(0, x_2, x_3, \ldots x_k)$ *such that all solutions of* (7) *are of the form* $(0, x_2, x_3, \ldots, x_k)+(c, c, \ldots, c)$*,* $c \in \mathbb{R}$.

Proof. Existence follows by construction: Since the vector W^r is generated using the realized vector of estimated valences $(\hat{v}^r_j)_{j=1,\ldots,k}$, a solution to (7) exists. Furthermore, it is clear that any vector of the form $(0, x_2, x_3, \ldots, x_k) + (c, c, \ldots, c)$ also satisfies (7). It remains to be shown that there cannot be a solution of the form $(0, y_2, y_3, \ldots, y_k)$ with $(0, y_2, y_3, \ldots, y_k) \neq (0, x_2, x_3, \ldots, x_k)$. Assume to the contrary, and let $\bar{k}$ be the candidate for whom $y_j - x_j$ is maximal. If $y_{\bar{k}} - x_{\bar{k}} > 0$, then substituting in the corresponding equation of (7) shows that candidate $\bar{k}$ receives a strictly higher vote share than $W^r_{\bar{k}}$, a contradiction. Similarly, let $\underline{k}$ be the candidate for whom $y_j - x_j$ is minimal. If $y_{\underline{k}} - x_{\underline{k}} < 0$, then substituting in the corresponding equation of (7) shows that candidate $\underline{k}$ receives a strictly smaller vote share than $W^r_{\underline{k}}$, a contradiction. But then, it must be true that $y_j = x_j$ for all $j = 2, \ldots, k$. ■

Note that vote shares are determined only by the *difference* between the candidates' estimated valences, so we can only determine those differences. However, it is also immaterial which of these possible beliefs a voter in a later state uses to infer the signals observed by the voters of that state.

Corollary 1 *Given a set of ex-post valence beliefs* $(0, x_2, x_3, \ldots, x_k) + (c, c, \ldots, c)$*,* $c \in \mathbb{R}$*, there is a unique vector of signals* $(0, y_2, y_3, \ldots, y_k)$ *such that all solutions to the system of equations given in* (10)*, for* $j \in \{1, \ldots, k\}$*, are of the form* $(0, y_2, y_3, \ldots, y_k) + (\gamma, \gamma, \ldots, \gamma)$.

Proof. This follows from the fact that equations (10) form a linear system in ex-post valances and observed signals for all candidates competing in state s. ■

By observing vote shares in the election of a prior state, a voter can infer signals up to a constant. As already pointed out, voters determine their preferred candidate on the basis of differences in ex-post

[9]The application of (10) and (11) is by round, i.e., all states voting in a particular round use values of v_{j0} and σ^2_{j0} as obtained from the signals up to the end of the previous round.

perceived valence, and these differences are determined by differences in the valence signals observed by voters of the state. In other words, a uniform shift of the ex-ante beliefs about all candidates by c translates into a uniform shift of the ex-post beliefs (i.e., after the state-specific signal), leaving the difference between the valence estimates for the different candidates, and hence the voter's voting decision, unaffected. The value of γ is immaterial in determining voting shares and can be normalized to zero.

To recapitulate, this section shows that the vote shares of candidates in a sequence of state contests can be obtained on the basis of equations (7) – (8), and given (a) the number of candidates in each position in each state contest, (b) the valence of these candidates, (c) the signals for every candidate observed by the voters in each state, (d) the fraction of voters μ_j in each state, j, who are of political position 1, and (e) the values of four parameters: σ_v, σ_η, λ, and σ_ε.

Finally, note that the right-hand sides of (7) are homogeneous of degree 0 in $(\varepsilon, \hat{v}^s, \sigma_\varepsilon)$. It is therefore useful to normalize $\sigma_\varepsilon \equiv 1$. Thus, all other parameters in the model are effectively expressed as multiples of the standard deviation of the idiosyncratic preference shock ε.

6 Empirical analysis of the 2008 Democratic primaries

We now turn to the empirical analysis, using data from the 2008 United States Presidential primary of the Democratic Party. However, our ultimate objective is not primarily to test our theoretical model for this particular primary race, but rather to obtain roughly plausible values for parameters on which we can base simulations of the effects of different primary structures. Using the point estimates as a starting point, we then analyze the robustness of the qualitative results to changes in parameters.

6.1 Data

Our dataset consists of the vote shares from the 2008 Democratic Presidential primary.[10] The three candidates that are included in our analysis are Barack Obama, Hillary Clinton, and John Edwards, while we exclude Dennis Kucinich and other minor candidates. We consider primaries in all states except Michigan,[11] plus the District of Columbia, yielding a total of 50 contests.

The prices on the Iowa Election Market for the 2008 Democratic nomination support this selection of candidates.[12] For example, on December 31, 2007 (i.e., just before the first primaries), the Arrow security that paid $1 if Hillary Clinton won the nomination had an average price of 63.8 cents, and the

[10]The reason for why we do not use the vote shares from the Republican primary is that the Republican primary displayed an unusual pattern of candidate withdrawal. Arguably, among the top two "conservative" candidates challenging the "moderate" McCain, Mike Huckabee was a weaker competitor than Mitt Romney, who, however, dropped out before. This sequence of exits is inconsistent with the spirit of the theoretical model and would render our estimation strategy problematic.

[11]The Michigan primary was held earlier than allowed for by Democratic party rules, and the names of Obama and Edwards were not on the ballot in Michigan.

[12]Market data are available at http://iemweb.biz.uiowa.edu/WebEx/marketinfo_english.cfm?Market_ID=214.

prices for Edwards and Obama were 11.5 cents and 24 cents, respectively. Thus, the three candidates that we focus on each had perceived winning chances greater than 10 percent. In contrast, the average price for the "rest of field" contract (i.e., any other person winning) on 12/31/2007 was 1.7 cent. Thus, even though Kucinich did receive a non-trivial vote share in some states, the market prices indicate that he was never perceived as a plausible nominee by market participants. Since such "protest candidates" do not fit our theoretical framework, we exclude Kucinich and other minor candidates.

A key component of the model is that candidates are distinguished by their horizontal position. In the introduction, we have presented evidence that voters viewed Edwards and Obama as relatively close substitutes for each other, while Clinton is farther away. There are certainly different potential explanations for why this was the case, and which one applies is immaterial for our estimation. Our preferred interpretation is that Obama and Edwards were perceived as outsiders, while Hillary Clinton was seen as part of the Democratic establishment and representing a continuation of the political philosophy of her husband's administration.[13] Voters may have different views on the desirability of such political dynasties (Dal Bo, Dal Bo, and Snyder (2009) document the importance of family connections for political careers in the U.S. Congress).

For the three major candidates, we obtain the vote percentage in the primary or caucus of each state from the Federal Election Commission. We rescale the data such that the vote shares of the candidates we consider add up to 100% (as assumed by the model). This data, along with the information about the round in which each state voted, is presented in Table 4 in the Appendix.

6.2 Identification

Our data consists of the number of candidates who compete in each state contest, along with their political position, vote shares, and the round of each state contest in the primary run. We do not observe voter signals, the distribution of voters to political positions (μ^s), or the candidate valence. We also do not observe the value of the parameters σ_υ, σ_η, and λ. Thus, we do not have all the information needed to calculate vote shares predicted by the model in a specific primary campaign for various configurations of state vote sequencing, as described at the end of Section 5. With our data being obtained from a single primary run, it is not feasible to obtain credible estimates of μ^s; we instead posit that μ^s is a random draw from the uniform distribution with mean equal to one half and support S_μ.[14]

Given that we do not estimate state specific values of μ^s, inverting the vote shares to obtain the state

[13] Deltas and Polborn (2009) argue that the single most salient partition of the Democratic candidates between in the three last presidential primaries was whether a candidate is perceived to be an insider of the Washington establishment, or rather draws his strength from the grass roots, and runs as an "outsider." In contrast, the liberal versus moderate distinction appears to be of lesser importance. Yet, even if the driving factor for the closer substitutability between Edwards and Obama was rather a male-female divide among voters, the implications for our estimation do not change.

[14] Deltas and Polborn (2009) find that the political positions of the candidates (i.e., "insider" or "outsider") do not significantly affect the candidates vote shares in the 2000-2008 Democratic primaries. This finding can be used as a (rough) justification for our assumption here that $E(\mu^s) = 1/2$.

signals is not a feasible strategy. Rather, we only aim to estimate (i) the standard deviation of candidate valence, denoted by σ_v; (ii) the standard deviation of state-specific information shocks, denoted by σ_η; (iii) the salience of the two major political positions, denoted by λ; and (iv) the support of electoral preferences for the two main political positions, denoted by S_μ.

In the estimation, we consider the withdrawal of Edwards after the fifth state contest as exogenous. That is, we do not use the exit of Edwards to draw any inference about state-specific signals beyond the first five states in which we observe Edwards's vote shares. The four parameters listed above then pin down the stochastic process that generates the vote shares. These parameters can also be used to obtain the stochastic process of vote shares under different state voting sequences, and under different assumptions about how long the third candidate (i.e., the equivalent of Edwards in a future race) stays in the race. We cannot infer what the outcome of the 2008 primary, holding state signals fixed, would have been with each different rule because we cannot estimate individual state signals with our approach. However, we can predict how the distribution of outcomes differs across different rules, if we were to draw candidate valences, voter preferences for positions and signals again and again from the estimated distributions. Thus, if the parameters remain stable over time, we can predict the outcome distribution under different primary systems in hypothetical future races. We describe these prospective simulations in detail in Section 7 below.

We now turn to a somewhat informal discussion of identification, where we consider the four parameters separately, taking the values of other parameters as given. This provides a useful intuition about the main sources of identification, even though all four parameters are estimated jointly and more than one source of variation in the data helps to pin down any given parameter.

The parameters S_μ and σ_η are identified jointly from the time variation of vote share volatility. Holding the candidates fixed, the model predicts that vote share volatility declines over time as voter beliefs about candidates' valence become more precisely concentrated around the true value. In the limit, once candidates' valences become known, share variability will be driven solely by variability in μ^s. Thus, holding other parameters constant, S_μ is identified from the limit share variability, and σ_η is obtained from the rate of decline in share variability towards that limit.

The parameters λ and σ_v are identified jointly from the mean vote shares and how these change after Edwards withdraws. High values of λ, holding other parameters constant, imply that a higher percentage of voters whose first choice is Edwards will vote for Obama in the absence of Edwards, as high values of λ make Clinton a worse substitute for Edwards. The value of σ_v is identified from the share of candidates in the two political position as a function of the number of candidates in each political position, both initially and in later election rounds. The higher the value of σ_v, the higher the expected difference in valence between the best of Obama and Edwards, and Clinton. Thus, higher values of σ_v are associated with lower vote shares for Clinton.

As noted above, identification of any particular parameter comes from multiple sources of data variation, and the informal discussion above focuses on the main sources of identification. To see the interdependence of parameter estimates, consider the following example: A higher value of λ would increase the value of σ_η implied by any given observed vote share volatility of Clinton. Since Clinton would be a poorer substitute for the other candidates, higher vote share variability could be rationalized by higher signal volatility. Similarly, changes in the two parameters that drive vote share volatility also have an impact on average shares (given that the vote share functions are non-linear). Our estimation procedure jointly pins down the parameter values from all these variations in the data.

Finally, note that the share of Clinton is sufficient for all of the above identification arguments to go through. We therefore only utilize her vote share for each of the 50 states. Following the withdrawal of Edwards, the vote shares of Obama provide no additional information, as vote shares add to 100 percent. For the first five contests, Edwards's vote shares add some information, but this information is not needed for identification. Omitting it yields substantial computational advantages, with a very small loss of efficiency.

6.3 Estimation

We estimate the unknown parameters S_μ, σ_v, σ_η, and λ from the 2008 Democratic primary using the method of moments. Given that our emphasis is on obtaining plausible parameters values for the purpose of simulation rather than for model testing, we utilize four moments of the data based on the identification arguments outlined in the preceding section. This leads to exact identification.[15]

We now describe our estimation approach. Let W_C^s denote the observed vote share of Clinton in state s. We partition states into three groups. The first group consists of the 5 states in which there was a three way race between Clinton, Edwards, and Obama; denote this group by $3WAY$, used (in the absence of any ambiguity) alternatively as a set or superscript. The second group consists of the 22 states that voted on Super Tuesday, denoted by ST. The last group consists of the 23 states that voted after Super Tuesday, denoted by pST. The union of the last two groups is denoted by the set or subscript $2WAY$. The indicator variable $1_{s\in A}$ takes the value of 1 if state s belongs in the group A and zero otherwise. Denote the sample average share of Clinton in the group of states A by $\bar{W}_C^A$.

Consider an election with two candidates located in position 1 and one candidate in position 0. The value of μ^s for each state is a random draw from the uniform distribution with mean 0.5 and support S_μ. Valences and signals are distributed normally with means 0 and v_j, and variances σ_v^2 and σ_η^2, respectively. There are five sequential contests in states $s = 1, \ldots, 5$, at the end of which the weaker of the two candidates in position 1 ($j = 1b$) withdraws.[16] The stronger one of the two candidates in position 1 ($j = 1a$) competes with the candidate in position 0 ($j = 0$) in two more rounds, one consisting of 22 states ($s = 6, \ldots, 27$), and the other one consisting of 23 states ($s = 28, \ldots, 50$).

[15] Incorporating additional moments would increase efficiency, but at substantial computational cost, primarily due to the iterative procedure needed to obtain the optimal weight matrix.

[16] Our definition of "weaker" is the candidate with the lowest valence draw. This is clearly the case in the 2008 primary, as Obama is ex-post widely understood to be of higher valence than Edwards.

The first moment in our analysis is based on the expectation of candidate 0's vote share in the first five states, and is given by

$$m1(\sigma_v, \sigma_\eta, \lambda, S_\mu) = E_{\mathbf{v},s}\left\{1_{s\in 3WAY} E_{\mu^s,\mathbf{Z}}[W_0^s | v_0, v_{1a}, v_{1b}, \sigma_\eta, \lambda, S_\mu]\right\}, \tag{12}$$

where the inner expectation is taken with respect to the distribution of μ_s and the signal histories $\mathbf{Z}$ and the outer expectation is taken with respect to the joint distribution of valence draws and over all states. Note that W_0^s does not depend on the values of μ^t for $t \neq s$. Thus, the inner expectation can be obtained by integrating W_0^s over the distribution of signal histories (conditional on the vector of valences), and then integrating the result with respect to the distribution of μ^s, resulting in a random variable whose value depends on the random valence draws and the state s.

The second moment in our analysis is based on the expectation of candidate 0's vote share in the last 45 states, and is given by

$$m2(\sigma_v, \sigma_\eta, \lambda, S_\mu) = E_{\mathbf{v},s}\left\{1_{s\in 2WAY} E_{\mu^s,\mathbf{Z}}[W_0^s | v_0, v_{1a}, \sigma_\eta, \lambda, S_\mu]\right\}, \tag{13}$$

where the expectations are taken as in (12). The inner expectation is a random variable at the start of primaries (its value depends on the valence draws) whose value does not depend on the time at which candidate $1b$ drops out, but which depends on the vote order of states.

The next two moments are based on vote share variability. The third moment refers to the elections on Super-Tuesday and is given by

$$m3(\sigma_v, \sigma_\eta, \lambda, S_\mu) = E_{\mathbf{v},\mathbf{Z},\mu^s,s}\left\{1_{s\in ST} |W_0^s - E_{\mu^s,\mathbf{Z}}[W_0^s | v_0, v_{1a}, \mathbf{Z_{3WAY}}, \sigma_\eta, \lambda, S_\mu, s \in ST]|\right\}. \tag{14}$$

The outer expectation is taken over all states with respect to valence draws, signal histories, and the distribution of voter preferences. The inner expectation is the expected value of vote shares of candidate 0 in the group of states belong in Super Tuesday, conditional on candidate valence and signal draws prior to the voting in those states. The expectation integrates out the variability in the state voter preferences, μ^s, of the Super Tuesday states and the signals received by their voters.

The last moment used in our analysis refers to the elections after Super-Tuesday and is given by

$$m4(\sigma_v, \lambda, S_\mu) = E_{\mathbf{v},\mu^s,s}\left\{1_{s\in pST} |W_0^s - E_{\mu^s}[W_0^s | v_0, v_{1a}, \sigma_\eta = 0, \lambda, S_\mu, s \in pST]|\right\}. \tag{15}$$

where, unlike in (14), both expectations ignore signal histories. The inner expectation gives the expected vote share of candidate 0 in post Super Tuesday states, conditional on candidate valence and assuming that this valence is known to the voters. That is, to simplify computations, we assume that valence is perfectly revealed in the last 23 states.[17] The outer expectation integrates over candidate valences and voter preferences across post Super-Tuesday states.

[17] This simplifies the estimation algorithm considerably, as we do not need to update candidate valence after Super Tuesday and can treat all subsequent states as voting simultaneously. For the estimated parameter values, this assumption appears largely justified: Using (20) to calculate the ratio of the standard deviation of perceived valence relative to actual valence shows that after Super-Tuesday, in expectation over 94 percent of the uncertainty about candidate valence is resolved.

Our estimates are based on the four by four equation system obtained by setting the moments equal to their sample analogs, where the vote shares of Clinton are considered to be the realizations of the vote shares of the candidate 0. The system that generates the estimates can be written as[18]

$$m1(\sigma_v, \sigma_\eta, \lambda, S_\mu) - \frac{1}{50} \sum_s \left\{ 1_{s \in 3WAY} W_C^s \right\} = 0, \tag{16}$$

$$m2(\sigma_v, \sigma_\eta, \lambda, S_\mu) - \frac{1}{50} \sum_s \left\{ 1_{s \in 2WAY} W_C^s \right\} = 0, \tag{17}$$

$$m3(\sigma_v, \sigma_\eta, \lambda, S_\mu) - \frac{1}{50} \sum_s \left\{ 1_{s \in ST} |W_C^s - \bar{W}_C^{ST}| \right\} = 0, \tag{18}$$

$$m4(\sigma_v, \lambda, S_\mu) - \frac{1}{50} \sum_s \left\{ 1_{s \in pST} \left| W_C^s - \bar{W}_C^{pST} \right| \right\} = 0. \tag{19}$$

Given exact identification, one can find parameter values so that these four equations will be satisfied with equality.[19] The expectations with respect to the distribution of valences and signals are obtained via Monte Carlo integration. Thus, the estimates we obtain contain some simulation error. The number of valence draws and sequences of signals was equal to 18,000, resulting in a simulation error that is less than 5 percent of the standard error (see Appendix for details).

6.4 Estimation Results

The estimation results and associated standard errors are $\hat{\sigma}_v = 0.92 \pm 0.29$, $\hat{\sigma}_\eta = 2.8 \pm 1.9$, $\hat{\lambda} = 1.5 \pm 0.17$, and $S_\mu = 0.67 \pm 0.04$. The standard errors are valid asymptotically as the number of candidates goes to infinity. While this is clearly not satisfied in our sample, the standard errors are nevertheless somewhat indicative of the relative confidence in our point estimates, with the dispersion in voter preferences being most precisely estimated (largely because it is pinned down by all 50 observations) and confidence in the variance of signals being least precisely estimated (because it is pinned down mainly from the results of the first 5 states).

Our primary interest lies in simulating the effects of different temporal organizations of the primaries (and not in a "test" of the model for the particular 2008 primary considered here). For the simulations, the point estimates of parameters are used as inputs for the base scenario, and we then analyze the qualitative robustness of results by changing the parameters one at a time. However, before we proceed to these simulations, it is useful to briefly discuss the relative importance of candidate

[18] Moment conditions are often written in terms of the contributions of each observation to the each moment. This can also be done in the system (16) – (19): Substituting for $m1(\cdot)$ from (12) and replacing the expectation with respect to the indicator variables by the sum over the observations, (16) can be written in terms of the contributions of each observation in the moment equations as $\frac{1}{50} \sum_s \left\{ 1_{s \in 3WAY} E_{\mathbf{v}} \left\{ E_{\mu^s, \mathbf{Z}} [W_0^s | v_0, v_{1a}, v_{1b}, \sigma_\eta, \lambda, S_\mu \right\} - 1_{s \in 3WAY} W_C^s \right\} = 0$, with analogous expressions for the other three moments through the corresponding manipulations of (17), (18) and (19).

[19] In general, it is not guaranteed that such a solution exists, but it does for this system of equations.

valence, voter preferences, differences in these preferences across states, and voter uncertainty about candidates implied by our estimation results.

The point estimate of σ_v indicates that the better of two candidates in the same political position who differ in one standard deviation of valence will obtain $\Phi(0.92) \approx 82\%$ of the voters who share the same political position when voters know the true valences. (Remember that the standard deviation of idiosyncratic preference shocks, σ_ε, is normalized to 1, so that Φ is the cdf of ε.)

The point estimate of λ indicates that a candidate in position 0 who is one standard deviation better (in terms of valence) than a candidate in position 1 will obtain $\Phi(2.42) \approx 99\%$ of the voters in position 0 and $\Phi(-0.58) \approx 28\%$ of the voters in position 1. Two candidates of equal valence but different positions get $\Phi(1.5) \approx 93\%$ of the voters with the same position and $\Phi(-1.5) \approx 7\%$ of the voters with the opposite position. Thus, the data imply that political positions are very important.

The point estimate of σ_η indicates that uncertainty about candidate valence is substantial in the states that vote early. For example, suppose that the valence difference between the two candidates in the same position is one standard deviation of valence. In this case, the chance that voters in the first district will actually perceive the better candidate as indeed better is only $\Phi(0.92/2.8) \approx 0.629$. Moreover, even if the better candidate receives the better signal and is thus also perceived as better, voters are aware that their signal has a relatively low quality and therefore put a low weight on it. Thus, the perceived valence difference between the two candidates is initially (in expectation) substantially smaller than the true valence difference, so that there is substantial vote-splitting between two candidates in the same position. In contrast, as argued above, if valence is known (which is almost the case in the final elections of a sequential primary system), then about 82% of the voters prefer the candidate with the higher valence over his competitor in the same position, and vote splitting will be minor.

More generally, consider Candidate j's perceived valence after N signals have been observed, $\hat{v}_j^N$. From an ex-ante point of view (i.e., before valence and signal realizations have been drawn), this is a random variable with expected value 0 (by the fact that the expected value of valence is zero, and expectations after signals follow a martingale). Given realized signals $(Z_j^s)_{s=1..N}$, expected valence is[20]

$$\frac{\sigma_v^2}{\sigma_v^2 + \frac{\sigma_\eta^2}{N}} \cdot \frac{\sum_{s=1}^N Z_j^s}{N}.$$

Thus, the variance of perceived valence after N signals have been observed is

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\sigma_v^2}{\sigma_v^2 + \frac{\sigma_\eta^2}{N}} x \right)^2 \phi \left(\frac{x - v}{\sqrt{\sigma_\eta^2/N}} \right) dx \quad \phi(v/\sigma_v) dv = \frac{\sigma_v^4}{\sigma_v^2 + \frac{\sigma_\eta^2}{N}}. \tag{20}$$

Note that this variance is always smaller than σ_v^2, because signal uncertainty implies that non-mean realizations of v are only learned over time, and the fact that voters know that signals are imperfect

[20]This is a weighted average of the ex-ante expected valence, 0, and the average signal realization (the second fraction), where the weight depends on the precisions of the ex-ante distribution of v and the precision of the signal distribution for N signals.

means that their updating to their signals is damped. Moreover, the variance of perceived valence is increasing in N and goes to σ_v^2 in the limit of $N \to \infty$; this is intuitive because, when valence is eventually revealed, the variance of perceived valence is the same as the ex-ante variance of valence. For our point estimates, (20) implies that the standard deviation of perceived valence is less than 0.3 in the first district, about 0.5 by the fifth district, and about 0.75 for district 20.

Finally, the point estimate of the support of μ indicates that the percentage of voters in each political position can be as low as 17 percent of the electorate and as high as 83 percent of the electorate. In the typical state, in terms of deviation from the 50/50 voter partition, a third of the voters support one position and two-thirds the opposite. Suppose that two candidates of equal perceived valence compete in that state. Then, the candidate with the less popular position in the state will obtain $\frac{1}{3}93 + \frac{2}{3}7 \approx 36$ percent of the votes and the candidate with the more popular position will obtain 64 percent of the votes. Vote shares are less variable than μ^s since a candidate obtains positive vote shares from voters in both positions. Suppose instead that the candidate with the less popular position was one standard deviation better (in terms of valence) than the candidate with the more popular position. Then, the vote share of the better candidate (who has, however, the less popular position) would be $\frac{1}{3}99 + \frac{2}{3}28 \approx 52$ percent of the electorate. Thus, the better candidate can overcome the typical electoral swing against him/her, but not by that much (however, the average difference between two randomly chosen candidates is in fact somewhat more than one standard deviation of valence).

7 Simulated effects of different institutions

We now use the point estimates of parameters to generate a baseline scenario that quantifies the implications of different primary systems. In Section 7.2, we then analyze the robustness of these results to changes in the parameters. Finally, we discuss our assumptions and limitations of the results in Section 7.3.

7.1 The baseline scenario

Our basic approach is as follows: We always consider races with three candidates, two of whom share a position while the third one is in the other position. In each simulation run, we first draw candidate valences from the estimated normal distribution $N(0, 0.92^2)$. Among the candidates who share a position, this generates two candidates with different valences, whom we denote B (for "better") and W (for "worse"). The other, "solitary", candidate is denoted S. We then draw state-specific signals according to $N(0, 2.8^2)$. Depending on the temporal structure of elections (and hence, on which signals are effectively observable in a state), this generates, according to Bayesian updating, voters' beliefs in a state.[21] We also draw aggregate position preferences in state s, μ^s, from a uniform distribution on

[21] As explained in Section 5, voters in later-voting states can essentially recover the realized state-specific signals of all states that voted before them.

[0.165, 0.835]. Together with the distribution of individual preference shocks (normalized to be drawn from $N(0, 1)$), this generates the vote distribution for candidates in a state. Aggregating over all states, we find the average vote share of each candidate, and the candidate with the most votes wins the nomination for a given run. (For the purpose of calculating aggregate vote shares, we assume that all states have the same size so that a candidate's aggregate vote share is simply the unweighted average of the candidate's vote shares in all states). We repeat this process 25,000 times to generate a probability distribution over outcomes, e.g., the proportion of times that B, W and S win the nomination.

We start by comparing the following three primary systems. The first system is a completely simultaneous primary in which all states vote at the same time. The second system is a completely sequential primary in which only one state votes at any given time. The third system is also a completely sequential primary, but, in contrast to the second system where we assume that all three candidates compete in all states, we now assume that all three candidates compete only for the first five states. Then, the candidate from the two that share a common position who is, after the fifth round of voting, perceived to be the weaker candidate (i.e., whose valence estimate at the beginning of the sixth district is lower) drops out. The remaining two candidates compete in the remaining 45 districts. Table 1 summarizes the results.

	I: Simultaneous elections	II: Purely Sequential, no dropout	III: Sequential with dropout after 5 rounds
S vote share	40.7%	38.7%	44.6%
B vote share	31.3%	41.2%	39.6%
W vote share	28.0%	20.0%	15.9%
S wins	98.4%	45.0%	39.0%
B wins	1.6%	48.9%	47.1%
W wins	0%	6.1%	13.9%
Exp. valence if S wins	0.016	0.519	0.578
...B wins	1.494	0.880	0.827
...W wins	n.a.	0.105	-0.012
S wins if CW	100%	88.5%	82.9%
B wins if CW	2.4%	68.5%	63.8%
Prob. that CW wins	35.2%	75.5%	70.2%
Winner's exp. valence	0.039	0.670	0.613

Table 1: Simulation results

The first and second three rows provide the mean vote shares and winning percentages of candidates S, B and W in the different primary systems, respectively. The nest three rows report the average valence of the nominee in the different primary systems, respectively. The next two rows give the winning probabilities of candidates S and B, conditional on being the Condorcet winner under full information. (Remember that Candidate W is never the Condorcet winner, because his position is the same as that of Candidate B, and his valence is lower).[22] Finally, the last two rows report the overall probability that the Condorcet winner wins, and the winner's expected valence.

The results indicate that, from a welfare perspective, a completely sequential voting system without dropout (regime II) performs best, independent of whether this performance is measured by the probability that the Condorcet winner wins, or the winner's expected valence. Simultaneous voting in all 50 states (regime I) does worst, with regime III intermediate, but closer to the pure sequential system.

For an intuition, consider first the simultaneous system. Candidate S wins almost all races, even though his average vote share is only 40.7%, because the two other candidates often split their votes almost evenly. As argued above, the variance in the voters' perception of valence is small in the first district, and, in a simultaneous system, all states are effectively a "first" state (i.e., they only observe their own state-specific signal). Vote-splitting is thus a prevalent problem, and almost always prevents the two candidates with a shared position from winning. Expected valence of the election winner is thus close to zero, the ex-ante expected valence of Candidate S. Also, Candidate B has a chance of winning only when he is significantly better than both Candidate S and Candidate W. Therefore, B's valence in those few instances where he wins is actually very high (more than 1.5 standard deviations above the expected valence).

Now consider regime II, the purely sequential system in which all candidates stay in the race. The learning facilitated by the sequential structure has the effect that vote share shifts from W to B (while S's vote share is just a bit lower than in regime I). As a consequence, B now wins much more often (48.9% of races). Note, however, that Candidate S still has an advantage in this system, as S still wins in many cases when he is not the Condorcet winner. This is reflected in the candidates' winning probability conditional on being Condorcet winner: While S wins over 88.5% of the races when he is the Condorcet winner, B wins only with probability 68.5% when he is the Condorcet winner.[23]

In regime III, we assume that during the first five elections, all candidates compete. Then, the candidate from the two that share a common position who is, after the fifth round of voting, perceived to be the weaker candidate (i.e., whose ex-ante valence estimate at the beginning of the sixth district is lower) drops out. The remaining two candidates compete in the remaining 45 districts. From a positive point of view, this modification has the expected effect of reducing the winning probability of Candidate S (from 45% to 39%), as there is now less vote-splitting for most of the election sequence. Surprisingly though, Candidate B's winning probability also decreases (from about 49% to 47%), while Candidate W's winning probability increases from 6% to almost 14%. The reason is that the probability

[22]Hence, all voters with $\varepsilon_W \leq \varepsilon_B$ (i.e., half of the population) strictly prefer B over W. By continuity, the set of voters who prefer B to W is always larger than the set of voters who prefer W to B.

[23]The reason that B wins absolutely more often than S is that B's expected valence is higher than S's, since he is the better of two candidates in his position – since valence draws are iid, the probability that B's valence is higher than S's is 2/3.

for a "mistake", i.e., the better Candidate B being forced to drop out after 5 rounds, is quite substantial (approximately 30%). As a consequence, this system performs worse from a welfare point of view than the purely sequential system without dropout.

In terms of overall welfare of the election outcome, the difference between simultaneous and sequential elections is substantial. If we take expected valence as our welfare measure, the valence increase of $0.670 - 0.039 = 0.631 \approx 0.686\sigma_v$. Also, the probability that the Condorcet winner is selected as nominee is substantially higher under sequential voting than under simultaneous voting.

Regime III in Table 1 provides just one sequential voting regime with dropout. Together with Regime II (which can be interpreted as "dropout" after 50 rounds), it raises the question when the socially optimal dropout time is that would optimally trade-off between coordination and learning. To investigate this question, we perform simulations of a purely sequential contest (no two states vote at the same time) in which the candidate who is perceived to be the lowest valence among B and W withdraws after state K. We vary K from 1 to 50, and plot the results in Figure 1.[24]

The results show that the electoral prospects of Candidate S are best for low and high values of K. When K is low, Candidate S faces a single opponent for most states; thus, vote splitting is kept at a minimum. However, the opponent is often the low-valence Candidate W, as Candidate B can easily be eliminated by a few bad draws in the first couple of states. For high values of K, S faces two opponents for most races and vote splitting is substantial; thus, S also often emerges as the winner. Intermediate levels of K (around 7 to 20) allow B to very likely dominate W, who then withdraws, and do so sufficiently early so that vote splitting is not excessive. This reduces the probability of winning for S. The electoral prospects of B more or less mirror those of S: They are low for low and high values of K and highest for intermediate values of K. They peak at somewhat higher values of K because a marginal increase in K reduces the probability of win for candidate W almost throughout the range. Finally, the electoral prospects of W decline monotonically until nearly the very end.[25]

The socially optimal value of K (using either reasonable measure of optimality) is even higher than the value of K that maximizes B's probability of winning. This is because higher values of K provide better information for comparing B and S, conditional on these two candidates remaining in the race. While expected valence and the probability that the Condorcet winner emerges as nominee both decline for K higher than about 30, the decline is very small. This suggests that for election contests of this type, the biggest concern is that the third candidate withdraws too soon rather than too late.

In practice, it may not be feasible to keep three candidates in the race for a very long time in a sequential primary system. After all, it is not just up to the candidates to decide when they want to give up, but also, voters may decide that only one of the two candidates in the shared position has a realistic

[24] Values for $K = 1, 2, \ldots 10$ and $K = 15, 20, 25, 30, 35, 40, 45, 50$ are as obtained from simulations, based on 25,000 replications. Values for remaining values of K are linear interpolations.

[25] A small uptick at the end is driven by the fact that incremental increases in K do not substantially affect the probability that it is Candidate W who withdraws (which is close to 1 anyway when K is high), but the increase in K increases W's cumulative vote share since he competes in more states.

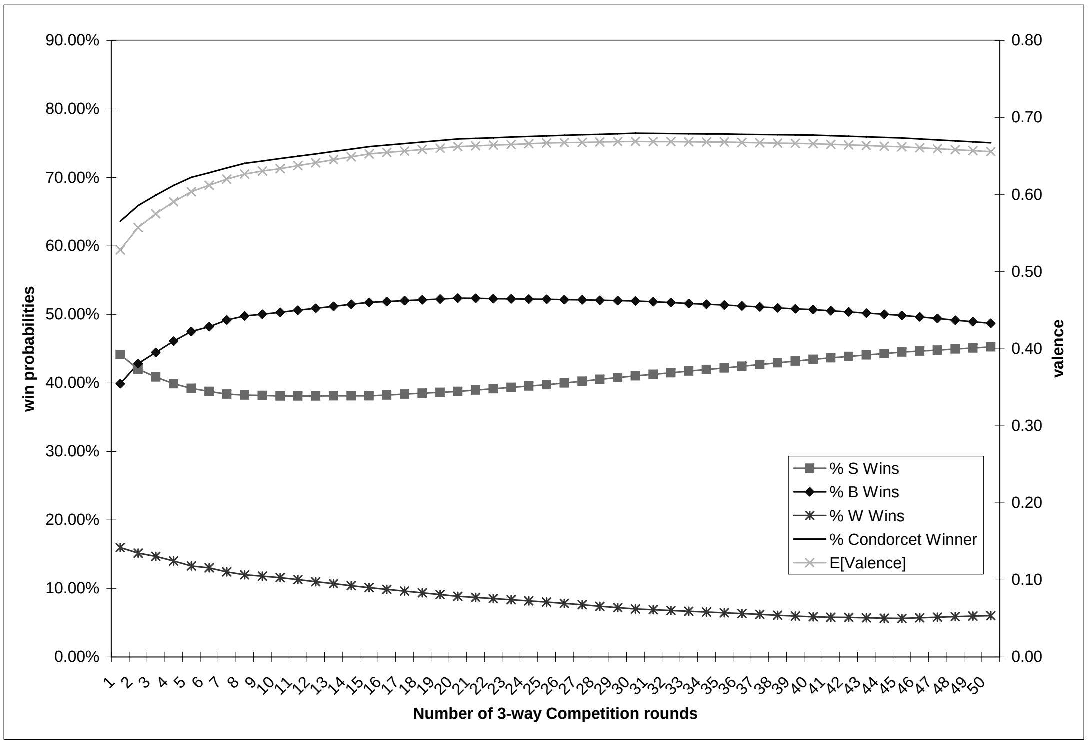

Figure 1: Winning probabilities and expected valence for different dropout rounds K

probability of winning, and they may effectively eliminate a contender as a "serious candidate" even if he officially stays in the race.

Figure 1 suggests that it would be very desirable to organize the primary sequence in a way that all three candidates remain in the race for at least ten districts or so, as the increase in expected valence is steepest in that range and then flattens out. The reform proposal by the National Association of Secretaries of State (NASS) has a very good chance to achieve this objective: There are only two initial elections in Iowa and New Hampshire, followed by four regional contests of approximately twelve states voting simultaneously, respectively. It appears plausible that all candidates remain in the race (at least) until after the first large regional contest.

Table 2 therefore compares the NASS proposal (Regime V) with Regime IV whose structure is modeled after the existing primary system. Specifically, in Regime IV, there are 5 initial sequential elections, followed by "Supertuesday", and another round in which all remaining states vote.[26] Like in Regime III, the candidate perceived as weaker after the fifth election drops out.

	IV: 2008 primary sequence w/ dropout after 5 states	V: NASS proposal w/ dropout after first regional primaries
S vote share	47.0%	42.9%
B vote share	36.7%	42.0%
W vote share	16.3%	15.2%
S wins	37.5%	38.7%
B wins	45.7%	50.8%
W wins	16.8%	10.5%
Exp. valence if S wins	0.458	0.640
. . . B wins	0.726	0.808
. . . W wins	-0.173	0.014
S wins if CW	65.9%	81.9%
B wins if CW	56.9%	69.0%
Prob. that CW wins	59.9%	73.4%
Winner's exp. valence	0.474	0.640

Table 2: Simulation results: Status quo vs. NASS proposal

From Table 2, it is apparent that the NASS structure does a considerably better job at eliminating the low valence candidate W, whose winning probability decreases from 16.8% to 10.5%. Interestingly, while most of those cases where W would win in Regime IV lead to a victory of Candidate B under the NASS structure, Candidate S's winning probability also increases, as S, while facing a stronger opponent more often, also benefits from vote splitting in 14 rather than just 5 districts.

Unconditional expected valence, as well as all conditional expected valences increase. This is intuitive for Candidate S, as his expected opponent is now stronger and so, if S manages to win nevertheless, he must be pretty good. Also, expected valence conditional on W winning increases because winning is relatively hard for W in the NASS structure: To have a chance of winning, it must be true that W's valence realization is very close to B's (so that he is wrongly perceived as stronger even after 14 signals), and W's valence must be substantially higher than S's, because otherwise S would be able to capitalize on B and W splitting votes for 14 districts.

[26] In reality, voting was more spread out after Supertuesday, but there are computational savings in assuming that all remaining states after Supertuesday vote simultaneously, and the disadvantage is very small, because voters' valence estimates are already very precise after 5 + 22 = 27 signals have been observed.

Finally, expected valence conditional on B winning increases. For B, winning becomes both easier and harder under the NASS proposal. A positive effect for B is that his probability of being (wrongly) eliminated in favor of candidate W decreases from 29.8% in Regime IV to 21.8% under the NASS structure. Yet, the increased vote splitting under the NASS structure means that winning *conditional on not being eliminated* becomes slightly harder for B, which increases B's expected valence conditional on winning.

Our second welfare measure, the probability that the Condorcet winner is selected as nominee also increases substantially under the NASS proposal relative to the status quo, from 59.9% to 73.4%. Interestingly, this increase is driven by a relatively uniform increase in both S and B's winning probability conditional on being the Condorcet winner.

7.2 Robustness

As argued above, the main purpose of the empirical analysis was to provide reasonable starting values for the simulations in the section above. However, since the parameter values are derived only from one primary (the 2008 Democratic race), it is useful to analyze whether our main qualitative results change when the parameters change relative to the baseline case. Specifically, we will analyze an increase or decrease of one parameter by one standard deviation, respectively, while fixing the other three parameters at their level in the baseline case.

Table 3 provides the results for the baseline case and the eight parameter changes.[27] We analyze the relative performance of three systems from the previous section: A completely simultaneous primary, a system that follows the 2008 setup (with dropout of the third candidate after five elections, just before Super-Tuesday), and the NASS proposal (with dropout of the third candidate after the first round of regional primaries).

Clearly, the numeric values of expected valence or the probability that the Condorcet winner wins the nomination change significantly as the parameters change. However, the relative ranking of the three systems remains the same as in the baseline case for all eight cases: The NASS proposal is the best, followed by the 2008 system and a simultaneous primary would do worst.

Changes in λ and S_μ have only a minimal effect on the probability that the Condorcet winner wins the election in each primary system. Partly, this is due to the fact that the deviations considered are relatively small (about a 10 percent change in λ, and a 6 percent change in S_μ) because these parameters are very well determined by our empirical estimation. However, from the size of the effects is clear that

[27]Note that the results of the baseline case differ slightly from those reported in Tables 1 and 2 in the previous section. The reason is that those results were based on 25,000 different parameter draws, while (for computational reasons), we restricted each of the simulations reported in Table 3 to 5000 draws. In order to keep the results comparable, we report the results for the baseline case for the same 5000 draws.

	I: Simultaneous elections	IV: 2008 primary sequence	V: NASS plan w/ dropout after first region
Baseline case			
expected valence	0.0533	0.4841	0.6430
CW wins	35.1%	60.6%	73.3%
$\lambda \uparrow$ ($\lambda = 1.69$)			
expected valence	0.0203	0.4775	0.6384
CW wins	33.9%	60.4%	73.3%
$\lambda \downarrow$ ($\lambda = 1.35$)			
expected valence	0.1318	0.4893	0.6467
CW wins	38.6%	60.7%	73.4%
$\sigma_v \uparrow$ ($\sigma_v = 1.21$)			
expected valence	0.5100	0.7314	0.9065
CW wins	52.0%	65.3%	78.1%
$\sigma_v \downarrow$ ($\sigma_v = 0.63$)			
expected valence	0.0058	0.2384	0.3796
CW wins	33.5%	53.5%	66.2%
$\sigma_\eta \uparrow$ ($\sigma_\eta = 4.7$)			
expected valence	0.0093	0.3287	0.5286
CW wins	33.4%	51.4%	63.4%
$\sigma_\eta \downarrow$ ($\sigma_\eta = 0.9$)			
expected valence	0.6992	0.7075	0.7518
CW wins	79.4%	81.7%	89.3%
$S_\mu \uparrow$ ($S_\mu = 0.71$)			
expected valence	0.0586	0.4825	0.6430
CW wins	35.3%	60.5%	73.3%
$S_\mu \downarrow$ ($S_\mu = 0.63$)			
expected valence	0.0490	0.4866	0.6429
CW wins	34.9%	60.7%	73.3%

Table 3: Results for different parameter values

also larger changes to λ and S_μ would not immediately change the welfare ranking of the three primary systems. (Clearly, as $\lambda \to 0$, the negative effects of vote-splitting disappear, and so for sufficiently small λ, simultaneous elections are an optimal system. But this limit result is almost tautological: If the setup is such that coordination does not matter, systems that allow for coordination do not have an advantage any more.)

As σ_v increases, the expected valence difference between candidates increases and is more likely to become decisive for voters' decisions. Thus, all systems become more likely to select the Condorcet winner as σ_v increases, and less likely to do so as σ_v decreases. Also, the winner's expected valence increases in σ_v because the winner is more likely to be the highest valence candidate as valence becomes more important for voters, and the expected realization of the highest valence draw increases as σ_v increases. Theoretical considerations indicate that, as $\sigma_v \to \infty$, all systems must deliver the same outcome (as almost always all voters agree on who is the best candidate, and almost always rank this candidate highest). Considering the probability that the Condorcet winner wins the nomination as our measure of welfare, when σ_v increases, simultaneous elections reduce their disadvantage relative to the other two systems while the difference between the 2008 system and the NASS proposal remains pretty much unchanged.

In contrast, when $\sigma_v \to 0$, theoretical considerations suggest that valence becomes less and less important for voters, and because vote splitting still leads to an electoral advantage of candidate S, he will almost always win in simultaneous primaries. In contrast, in both forms of sequential primaries, coordination allows for a substantial winning probability for one of the two candidates in the same position.[28]

Finally, the change in σ_η that we consider is very large (the standard deviation of σ_η is very large, because σ_η is mostly estimated from only 5 elections in our data). In order to interpret this variation, it is useful to start from equation (20) for $N = 1$, and note that the ratio between the standard deviation of perceived valence and the standard deviation of actual valence in the first district is $\sqrt{\frac{\sigma_v^2}{\sigma_v^2+\sigma_\eta^2}}$. This ratio would be 0 if the signal is completely uninformative ($\sigma_\eta \to \infty$), and equal to 1 if the signal is completely informative ($\sigma_\eta \to 0$). For $\sigma_\eta = 4.7$, the ratio is about 0.192, and for $\sigma_\eta = 0.9$, the ratio is about 0.715. Thus, our robustness test with respect to σ_η covers most of the conceivable variation.

If σ_η is increased to 4.7, then vote splitting is severe (because the quality of information about valence is very low), and candidate S wins almost always under a simultaneous system (thus, the probability that the Condorcet winner wins is about $1/3$ in that system). In contrast, both sequential systems allow for coordination. However, the NASS proposal aggregates more information than the

[28] Note that, when $\sigma_v = 0$, then valence does not matter at all for voters and position is the only decisive criterion. Candidate S therefore is the Condorcet winner in $1/2$ of the cases, and since candidate S (almost) always wins in simultaneous primaries, the probability that the Condorcet winner wins goes to $1/2$ in the simultaneous system. Among the two sequential systems, the 2008 system allows for coordination in 45 states (as dropout occurs after 5 elections), while the NASS proposal in our simulation only allows for coordination in 36 states. Therefore, as σ_v becomes very low, we would expect that the 2008 system eventually looks better than the NASS system.

2008 system, because it has more elections with all candidates competing, and thus the NASS proposal does significantly better than the 2008 system. If $\sigma_\eta \to \infty$, signals become completely uninformative. In that case, the number of states with vote-splitting is the only distinction between the 2008 system and the NASS proposal.

If σ_η is decreased to 0.9, then each signal is very informative about valence. For the two candidates in the same position, this effect diminishes the importance of vote splitting, which explains the improved performance of the simultaneous system for this case. Also, the welfare difference between the 2008 system and the NASS proposal shrinks. As $\sigma_\eta \to 0$, we would expect that the 2008 system eventually becomes better than the NASS system because it has vote splitting in fewer states, and the advantage of conditioning the decision of which candidate should drop out on more observations vanishes when already a single signal is very informative.[29]

In all simulations so far, we have assumed that the valence of candidates is drawn from the same distribution, independent of their position, and the number of candidates who compete. Alternatively, one might think that the sole candidate in the one position might be the result of some coordination among potential candidates in this position. If this is the case, then it might be more reasonable to assume that the sole candidate's valence is drawn from a better distribution. A possible formalization of this idea is that his valence is $\max(v_{S,1}, v_{S,2})$, where $v_{s,j}$ is distributed $N(0, \sigma_v)$. Effectively, this presumes that there were two proto-candidates in position 0, but that, before the start of the primaries, the sole candidate already convinced the other candidate who was located in the same position (but had a worse valence) not to run. As a consequence, the distribution of candidate S's valence is the same as the distribution of candidate B's valence, and each of them is the Condorcet winner with 50 percent probability.

In this case, the winner's expected valence in a simultaneous primary (case I) is 0.5199, in the 2008 system (case IV) is 0.6435, and in the NASS system is 0.8235. The probability that the Condorcet winner wins the nomination is 50.2 percent in case I, 60.1 percent in case IV and 74.9 percent in case V. These results show that the performance of simultaneous elections in this scenario is substantially better than in the baseline case (essentially, because S now is the Condorcet winner more often and still wins with probability close to 1), while the effect in the two sequential systems is rather small. The relative ranking of the three systems is again unaffected, and this is also unlikely to change if we were to change additional parameters while maintaining the new assumption about the distribution of S's valence.

[29]Note, however, that if σ_η is close to zero, then there are essentially no momentum-effects in sequential primaries: A candidate's win in an earlier state has no (or almost no) effects on later states, because voters in these later states do not need the earlier states' signals to update on the candidates' valences. Any variation in election results between states is purely driven by differences in preferences for positions (i.e., μ), and until the third candidate drops out, the ratio between the vote shares of candidates B and W remains more-or-less the same in all states. This prediction appears to conflict with the role of momentum which is perceived to be quite important in sequential primaries.

7.3 Discussion

In this section, we discuss some of our assumptions on which the empirical and simulation analysis is based. A possible criticism of our approach is that our parameters are derived from the 2008 Democratic primary which was unusually competitive: The eventual runner-up, Hillary Clinton, received a larger share of convention delegates than any other runner-up in the history of the modern presidential primary system. Thus, the two top candidates were likely of very similar valence. To the extent that there were a number of other races that produced considerably more lopsided results (for example, whenever one of the candidates is an incumbent President), one can certainly argue that the 2008 primary was "not representative" for the set of all primaries.

However, we would argue that focusing the empirical analysis on a competitive race is actually preferable to an analysis that includes less competitive campaigns, because our main interest is the effects of different institutional designs of the primary process. Whether primaries are held simultaneously or sequentially will not matter in races where one candidate is clearly superior. In this sense, our setup that assumes that candidates are drawn from the same distribution probably *exaggerates the size of* the impact of institutions on welfare, because there are noncompetitive scenarios where the precise institutional design is unlikely to matter, one way or the other. However, this criticism does not affect what is actually the *best* institutional setup. If a given fraction of nomination campaigns are competitive, while the remainder is non-competitive (i.e., the same candidate would win in any primary system), then a welfare analysis can focus on the competitive primaries without loss, as those are the cases where the setup of the primary system potentially matters. For this reason, picking an unusually competitive race such as 2008 as the baseline is actually quite appropriate.

A second and unavoidable simplification of our simulation approach is that, when we compare different primary organizations, we hold fixed the set of candidates and the distributions from which candidate valences and signals are drawn. In principle, the temporal setup of primaries may influence both the quality of signals and the decisions of potential primary candidates (and thus the composition of the field of candidates).

With respect to signal quality, it is conceivable that, in a sequential setup, residents of early-voting states receive a better signal than voters in most other states (because candidates spend a lot of time campaigning in early states). If this is the case, our simulations will overestimate the performance of a simultaneous primary system relative to a sequential one.[30]

With respect to the composition of the candidate field, the following effect may arise. If vote splitting in a simultaneous primary would be substantial when two candidates in one position compete with a sole candidate in the other position, there may be a considerable incentive to coordinate on one of the two candidates and force the second one out before the election even takes place. Moreover, even if no candidates drop out, voters may be able to use public opinion polls to effectively coordinate

[30] Of course, if we believe that early states receive on average better quality valence signals, this could also be considered in the estimation, though pinning the precise value down from very few states would be problematic.

on one of the two candidates in a simultaneous election. If this is the case, our simulations would underestimate the performance of a simultaneous primary system relative to a sequential one.

While the argument concerning the endogeneity of the candidate set with respect to the temporal organization of primaries is theoretically valid, we believe that its impact on our qualitative results is limited. Our first argument is that coordination in simultaneous primaries may be non-trivial to achieve in practice. In simultaneous party primaries (for state offices or U.S. Congress) in which no incumbent is running, there are often contests with several serious candidates who all receive substantial vote shares, and where the winner's vote share is often below 50 percent, indicating the potential importance of vote splitting. For example, in the 2010 Republican primary for Governor of Illinois, five of the seven candidates received more than 14 percent of the votes each, and Bill Brady won with a vote share of just 20.3 percent. Moreover, only Brady came from "downstate", while the remaining (serious) candidates all came from Chicago and its suburbs, and there appears to have been considerable region-based vote-splitting. For example, Brady received only 7 percent in Chicago and its suburbs, but won nevertheless because of his strong showing downstate and since the Chicago-based candidates split the vote there very evenly. This example suggests that coordination facilitated by either candidates dropping out before the election or based on opinion polls cannot be taken for granted even in high-profile races.

Our second argument focuses on the quality of coordination in simultaneous versus sequential primaries. Suppose that voters who prefer the same horizontal characteristic are actually able to solve the coordination problem in simultaneous primaries in some informal way (say, using straw polls or opinion polls). The random event that voters can utilize for coordination is likely to be of substantially worse information quality than the outcome of an actual primary election in a state because, for example, the sample of people who participate in the straw poll or opinion poll is unlikely to be perfectly representative of the population. Also, attempts by the candidates to influence the coordination criterion in a way that is not reflective of true valence are more likely to be successful in straw polls than in statewide elections.[31] So, it is true that informal coordination in simultaneous primaries might have the effect that the outcome in this system is not quite as bad as our simulations suggest, but informal coordination is unlikely to change the qualitative result that simultaneous primaries are worse than sequential primaries, because informal coordination would be an imperfect copy of coordination through early primaries.

[31]Consider the Iowa Straw Poll, which is organized by the Republican party in the summer of the year before presidential nomination contests. A poor showing in the Iowa Straw Poll is often very problematic for a candidate and may effectively end his campaign (for example, in 2008, Tommy Thompson and Sam Brownback were effectively eliminated by this straw poll). For this reason, candidates often spend substantial resources in order to provide transportation or buy tickets for their supporters, diminishing the informational content of the voting outcome.

8 Conclusion

At the beginning of presidential primaries, there are often several serious contenders. Some of them may be ideologically close substitutes for voters, while the difference to other candidates may be more significant. In a simultaneous election with a large set of candidates, the candidate who would come out on top is not necessarily the Condorcet winner. In contrast, sequential elections allow voters to narrow down the field of contenders as a way of avoiding vote-splitting among ideologically similar candidates. The sequential nature of the primaries therefore likely has facilitated the victory of candidates who were not the frontrunner at the beginning of the primary season, such as Obama (and possibly McCain) in 2008, and the very strong showing of Gary Hart in 1984.

In this paper, we have presented a model of voting in sequential primaries based on the ideas of coordination and learning about candidate quality. From a theoretical perspective, the coordination afforded by sequential elections may be beneficial or detrimental. While sequential elections have the advantage of allowing voters to coordinate (and thus avoid that a candidate wins just because his ideological opponents split the votes of their supporters among each other), the disadvantage of sequential elections is that, once coordination has occurred, there is no possibility to correct an error made in early elections. Moreover, our empirical results show that the probability of the wrong candidate dropping out after the first few primaries is substantial.

Sequential elections are likely to dominate simultaneous ones if valence differences between candidates are small; if the signal quality in early states is high; and if there is a lot of vote-splitting between ideologically similar candidates. In contrast, when valence differences are important, vote-splitting is not too important and the signal quality is bad, then a simultaneous primary system is superior.

We estimate the model using data from the 2008 Democratic primaries, and use the parameter estimates to evaluate the relative performance of different temporal organizations of the primaries. Our results suggest that vote-splitting would be a severe problem in a simultaneous primary system. However, sequential institutions in which one of the candidate is forced out (in which therefore avoid the vote-splitting problem for most districts) are also not optimal, as a too early drop-out date induces a high probability that the better candidate drops out.

A current proposal by the National Association of Secretaries of State does very well from a welfare point of view in our simulations. According to this proposal, Iowa and New Hampshire would always vote first, followed by four regional primaries (for the East, Midwest, South and West regions) scheduled on the first Tuesday in March, April, May or June of presidential election years. Assuming that all candidates stay in the race until after the first large regional contest, there are sufficiently many early elections to be relatively confident that the strongest candidates survive, yet vote splitting is absent in three out of four large regional contests.

References

Ali, N. and N. Kartik (2006). Social learning in elections. UC San Diego.

Bartels, L. M. (1985). Expectations and preferences in presidential nominating campaigns. *The American Political Science Review 79*(3), 804–815.

Bartels, L. M. (1987). Candidate choice and the dynamics of the presidential nominating process. *American Journal of Political Science 31*(1), 1–30.

Battaglini, M. (2005). Sequential voting with abstention. *Games and Economic Behavior 51*, 445–463.

Battaglini, M., R. Morton, and T. Palfrey (2007). Efficiency, equity and timing in voting mechanisms. *American Political Science Review 101*, 409–424.

Callander, S. (2007). Bandwagons and momentum in sequential voting. *Review of Economic Studies 74*, 653–684.

Coughlin, P. (1992). *Probabilistic Voting Theory*. Cambridge, Cambridge University Press.

Dal Bo, E., P. Dal Bo, and J. Snyder (2009). Political dynasties. *Review of Economic Studies 76*, 115–142.

Degan, A. and A. Merlo (2006). Do voters vote sincerely? University of Pennsylvania.

Dekel, E. and M. Piccione (2000). Sequential voting procedures in symmetric binary elections. *Journal of Political Economy 108*(1), 34–55.

Deltas, G. and M. Polborn (2009). Candidate competition and voter learning in sequential primary elections: Theory and evidence. Working paper, University of Illinois.

Diermeier, D. and R. Myerson (1999). Bicameralism and its Consequences for the Internal Organization of Legislatures. *American Economic Review 89*(5), 1182–1196.

Kenny, P. J. and T. W. Rice (1994). The psychology of political momentum. *Political Research Quarterly 47*(4), 923–938.

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

Knight, B. and N. Schiff (2007). Momentum and social learning in presidential primaries. NBER.

Krasa, S. and M. K. Polborn (2010). The binary policy model. *Journal of Economic Theory 145*(2), 661–688.

Lindbeck, A. and J. Weibull (1987). Balanced-budget redistribution as the outcome of political competition. *Public Choice 52*, 273–297.

Moulitsas, M. (2008a). Edwards voters' second choice, part II. Available at Daily Kos, http://www.dailykos.com/storyonly/2008/8/12/1101/20327/19/566444.

Moulitsas, M. (2008b). Wolfson's ridiculous speculation. Available at Daily Kos, http://www.dailykos.com/storyonly/2008/8/11/163149/067/156/566331.

Persson, T. and G. Tabellini (2000). *Political Economy: Explaining Economic Policy*. MIT Press.

Schwabe, R. (2010). Super Tuesday: Campaign Finance and the Dynamics of Sequential Elections. mimeo, Northwestern.

Stimson, K. (2008). The case for regional presidential primaries in 2012 and beyond. Available at http://nass.org/index.php?option=com_docman&task=doc_download&gid=165.

9 Supplemental material

9.1 A tractable special case

In this section, we provide a complete theoretical analysis of a particular case of the model that can be solved in closed form and provides some intuition for the effects of the temporal organization of primaries.

There are initially three candidates ($J = 3$). Candidate 1's position is $a_1 = 0$, while Candidates 2 and 3 have $a_2 = a_3 = 1$. Furthermore, assume that λ is sufficiently large relative to the span of the distributions of valence υ such that a difference in the policy dimension (almost) always dominates both valence difference and the idiosyncratic preference shock ε. In other words, all voters with preferred position $\theta^i = 0$ vote for Candidate 1, while those voters with $\theta^i = 1$ either vote for Candidate 2 or 3.[32] This creates a coordination problem for those voters whose preferred position is 1: If candidates 2 and 3 split the votes of those voters who prefer position 1, then Candidate 1 may win even if he is not the Condorcet winner (i.e., the candidate who would be preferred by a majority of voters to all other candidates, if valences were known).

We also assume that the proportion of the total population with preference for $a = 1$ is equal to μ in all districts ($\mu^1 = \mu^2 = \ldots = \mu^N \equiv \mu$). Clearly, if $\mu < 1/2$, then Candidate 1 is the Condorcet winner, and his supporters form a majority in each district. If $\mu > 1/2$, then either Candidate 2 or Candidate 3 is the (full information) Condorcet winner, depending on which one of them has the higher valence.

We assume that the number of states is large ($S \to \infty$), and analyze two temporal organizations of the primary system. Under simultaneous elections, all S states vote at the same time. Under sequential elections, one state votes at $t = 0$, and the remaining $S - 1$ states vote at $t = 1$, after observing the election outcome in the first state; in this case, the set of relevant candidates at $t = 1$ is formed by excluding either Candidate 2 or 3 (i.e., one of the two candidates in position 1), depending on who did worse in the first state. Proposition 2 characterizes the equilibrium for the two different primary systems. By Condorcet loser, we mean the candidate who would lose against either opponent.

Proposition 2 *Assume that Candidate 1's policy position is 0 and both Candidate 2 and 3 have policy position 1. Additionally, suppose that λ is large relative to σ_υ and σ_ε.*

If $\mu < 1/2$, Candidate 1 is the Condorcet winner. If $1/2 < \mu < 2/3$, Candidate 1 is the Condorcet loser, and the candidate with the higher valence among Candidates 2 and 3 is the Condorcet winner.

1. *If $\mu < 1/2$, Candidate 1 wins under both a simultaneous and a sequential primary system.*

2. *If $1/2 < \mu < 2/3$,*

[32] In principle, the distribution of ε is unbounded such that there are some voters with, say, type $\theta = 1$, but a very large ε_1, who thus prefer Candidate 1. However, when λ is large relative to σ_ε, such voters will be exceedingly rare, and we just ignore these cases in this section (in order to gain tractability).

(a) *In a sequential primary system, either Candidate 2 or Candidate 3 wins. The probability that the Condorcet winner wins is decreasing in σ_η and increasing in σ_v.*

(b) *In a simultaneous primary system, either the Condorcet winner or Candidate 1 wins. There exists $\mu^* \in (1/2, 2/3)$ such that Candidate 1 (the Condorcet loser) wins the nomination with positive probability for every $\mu < \mu^*$.*

3. *If $\mu > 2/3$,*

(a) *In a sequential primary system, Candidates 2 and 3 each win with positive probability, while Candidate 1 cannot win.*

(b) *In a simultaneous primary system, the Condorcet winner wins with probability 1.*

Before we proceed to a formal proof of these claims, it is useful to discuss them informally. $1-\mu > 1/2$, Candidate 1 receives an absolute majority of votes in every district, whether he competes against one or two opponents. The election system only affects whether the votes of type $\theta = 1$ voters are split or united, but even coordination cannot change that Candidate 1 wins.

If $\mu \in (1/2, 2/3)$, type 1 voters are in the majority, and thus either Candidate 2 or 3 is the Condorcet winner. However, since Candidate 1 receives more than one-third of the votes, it is possible that he receives a plurality in some or all districts. In this case, interesting differences between sequential and simultaneous primary systems arise. The advantage of a sequential system is that it avoids vote splitting and thus prevents a victory of the Condorcet loser; however, the winning candidate may be of lower quality than the candidate who dropped out. In contrast, in a simultaneous election system, the law of large numbers guarantees that the better of Candidates 2 and 3 wins more votes than the weaker one. However, since there is vote splitting Candidate 1, the Condorcet loser, may still win.

To see these effects in more detail, consider first sequential elections. Since $\mu > 1/2$, either Candidate 2 and Candidate 3 (whoever wins more votes in the first district) will win all remaining districts. Thus, in a sequential organization of primaries, it is impossible that the Condorcet loser wins. However, because the signal of first-district voters is not perfect, the Condorcet winner may fare worse in the first district than his competitor with the same position. Intuitively, a higher σ_η means that there is a larger chance that the difference of observation mistakes for the two candidates outweighs their valence difference, so that voters in the first district mistakenly perceive the worse candidate as the better one. If σ_v increases, this increases the expected valence difference between the better and the worse candidate and thus increases the probability that the Condorcet winner wins. The Condorcet winner's exact winning probability is derived in the proof.

Now consider simultaneous elections when $\mu \in (1/2, 2/3)$. Since Candidate 1's vote share, $1-\mu$, is larger than $\mu/2$, it is possible that voters with a preference for Candidate 2 or 3 split in such a way in a district that Candidate 1 wins a plurality. How often this happens depends on parameters. If there is a large difference between the perceived valences of Candidates 2 and 3, and if the idiosyncratic preference differences captured by ε are sufficiently small for most voters, then almost all of them agree on one candidate, and vote splitting is minimal. In these cases, the Condorcet winner is likely to win a plurality. In contrast, if perceived valence differences between candidates are small or idiosyncratic preference shocks are large, then both Candidate 2 and 3 receive a substantial fraction of support, and Candidate 1 may win.

In the third case where $\mu > 2/3$, type 1 voters are in the majority, and thus either Candidate 2 or 3 is the Condorcet winner. In contrast to the case that $\mu \in (1/2, 2/3)$, though, the electorate's preference distribution is sufficiently extreme for $\mu > 2/3$ to make up for any extent of vote splitting between Candidates 2 and 3. Candidate 1 cannot win if $\mu > 2/3$.

In a simultaneous elections system, the law of large numbers guarantees that the better candidate (among Candidates 2 and 3) wins a larger number of districts than his weaker competitor. Thus, when $\mu > 2/3$, the Condorcet winner always wins under simultaneous elections. In contrast, in a sequential election system, there can still be mis-coordination on the worse candidate among Candidates 2 and 3 because, depending on the outcome of the first district, the Condorcet winner may be eliminated.

Proof of Proposition 2.

1. $\mu < 1/2$. Since $1 - \mu > 1/2$, Candidate 1 receives an absolute majority of votes in every district, whether he competes against one or two opponents. The election system only affects whether the votes of type $\theta = 1$ voters are split or united.

2(a). $\mu \in (1/2, 2/3)$ **and sequential elections.** Candidate 2 gets more votes in the first district than Candidate 3 if and only if $v_2 + \eta_2^1 > v_3 + \eta_3^1$. Since $\eta_3 - \eta_2$ is distributed according to $N(0, 2\sigma_\eta^2)$, for given v_2 and v_3, Candidate 2 wins with probability $\Phi\left(\frac{v_2 - v_3}{\sqrt{2}\sigma_\eta}\right)$. Note that $v_2 - v_3$ is distributed according to $N(0, 2\sigma_v^2)$. Without loss of generality, we can focus on the case $v_2 > v_3$; conditioning on this event, the density of $v_2 - v_3$ is given by $2\phi\left(\frac{t}{\sqrt{2}\sigma_v}\right)$. Thus, the probability that the better candidate wins is given by

$$2\int_0^\infty \Phi\left(\frac{t}{\sqrt{2}\sigma_\eta}\right)\phi\left(\frac{t}{\sqrt{2}\sigma_v}\right)dt = \sqrt{2}\sigma_v\left[1 - \frac{\arctan\left(\frac{\sigma_\eta}{\sigma_v}\right)}{\pi}\right]. \tag{21}$$

Since the arc tan is an increasing function and lies between 0 and π (for positive arguments, such as here), it is easy to see that this probability is decreasing in σ_η and increasing in σ_v.

2(b). $\mu \in (1/2, 2/3)$ **and simultaneous elections.** It is useful to denote by ϕ_α, $\alpha \in \{v, \eta, \varepsilon\}$, the probability density function of the normal distribution of variable α. The voters in district s observe signal $Z_j^s = v_j + \eta_j^s$. Using Bayes' rule, the updated expected value of Candidate j's valence is

$$\hat{v}_j^s = \int_{-\infty}^{\infty} \frac{\phi_v(t)\phi_\eta(Z_j^s - t)}{\int_{-\infty}^{\infty} \phi_v(t')\phi_\eta(Z_j^s - t')dt'} t dt = \frac{{\sigma_v}^2}{{\sigma_v}^2 + {\sigma_\eta}^2} Z_j^s. \tag{22}$$

If Voter i in district s has type $\theta = 1$, he votes for Candidate 2 if $\hat{v}_2^s + \varepsilon_2^i > \hat{v}_3^s + \varepsilon_3^i$, and for Candidate 3 otherwise. Rearranging, the percentage of type $\theta = 1$ voters who vote for Candidate 2 is equal to

$$\mathrm{Prob}(\varepsilon_3 - \varepsilon_2 \leq \hat{v}_2^s - \hat{v}_3^s) = \mathrm{Prob}\left(\varepsilon_3 - \varepsilon_2 \leq \frac{{\sigma_v}^2}{{\sigma_v}^2 + {\sigma_\eta}^2}[Z_2^s - Z_3^s]\right) = \Phi\left(\frac{{\sigma_v}^2[Z_2^s - Z_3^s]}{\sqrt{2}\sigma_\varepsilon({\sigma_v}^2 + {\sigma_\eta}^2)}\right). \tag{23}$$

Similarly, Candidate 3's share of the vote of $\theta = 1$ types is equal to

$$1 - \Phi\left(\frac{{\sigma_v}^2[Z_2^s - Z_3^s]}{\sqrt{2}\sigma_\varepsilon({\sigma_v}^2 + {\sigma_\eta}^2)}\right).$$

Candidate 1, the Condorcet loser, receives all votes from $\theta = 0$ types (a proportion $1-\mu$ of the electorate) and wins a particular district s if and only if

$$1 - \mu > \mu \cdot \max\left(\Phi\left(\frac{{\sigma_v}^2[Z_2^s - Z_3^s]}{\sqrt{2}\sigma_\varepsilon({\sigma_v}^2 + {\sigma_\eta}^2)}\right), 1 - \Phi\left(\frac{{\sigma_v}^2[Z_2^s - Z_3^s]}{\sqrt{2}\sigma_\varepsilon({\sigma_v}^2 + {\sigma_\eta}^2)}\right)\right), \tag{24}$$

hence if

$$\frac{2\mu - 1}{\mu} < \Phi\left(\frac{{\sigma_v}^2[Z_2^s - Z_3^s]}{\sqrt{2}\sigma_\varepsilon({\sigma_v}^2 + {\sigma_\eta}^2)}\right) < \frac{1-\mu}{\mu}. \tag{25}$$

Denoting the inverse of the cumulative distribution of the standard normal distribution by Φ^{-1}, and letting $\kappa = \frac{\sigma_v^2}{\sqrt{2}\sigma_\varepsilon({\sigma_v}^2 + {\sigma_\eta}^2)}$, we can write this as

$$\Phi^{-1}\left(\frac{2\mu - 1}{\mu}\right) < \kappa(v_2 - v_3) + \kappa(\eta_2 + \eta_3) < \Phi^{-1}\left(\frac{1-\mu}{\mu}\right) \tag{26}$$

For given v_2 and v_3, the term in the middle is normally distributed with expected value $\kappa(v_2 - v_3)$ and variance $2\kappa^2{\sigma_\eta}^2$. Thus, the percentage of districts won by Candidate 1 is given by

$$\begin{aligned}
&\mathrm{Prob}\left(\Phi^{-1}\left(\frac{2\mu - 1}{\mu}\right) - \kappa(v_2 - v_3) < \kappa(\eta_2 - \eta_3) < \Phi^{-1}\left(\frac{1-\mu}{\mu}\right) - \kappa(v_2 - v_3)\right) = \\
&\Phi\left(\frac{\Phi^{-1}\left(\frac{1-\mu}{\mu}\right) - \kappa(v_2 - v_3)}{\sqrt{2}\kappa\sigma_\eta}\right) - \Phi\left(\frac{\Phi^{-1}\left(\frac{2\mu-1}{\mu}\right) - \kappa(v_2 - v_3)}{\sqrt{2}\kappa\sigma_\eta}\right) = \\
&\Phi\left(\frac{\Phi^{-1}\left(\frac{1-\mu}{\mu}\right) - \kappa(v_2 - v_3)}{\sqrt{2}\kappa\sigma_\eta}\right) - \Phi\left(\frac{-\Phi^{-1}\left(\frac{1-\mu}{\mu}\right) - \kappa(v_2 - v_3)}{\sqrt{2}\kappa\sigma_\eta}\right),
\end{aligned} \tag{27}$$

where the last inequality uses the fact that $\Phi^{-1}\left(\frac{2\mu-1}{\mu}\right) = -\Phi^{-1}\left(\frac{1-\mu}{\mu}\right)$, because $\frac{2\mu-1}{\mu}$ and $\frac{1-\mu}{\mu}$ are symmetric around $1/2$ (i.e., add up to 1).

Again, suppose that $v_2 > v_3$, so that Candidate 2 is the toughest competitor for the nomination. The percentage of districts won by Candidate 2 is

$$\Phi\left(\frac{-\Phi^{-1}\left(\frac{1-\mu}{\mu}\right) - \kappa(v_2 - v_3)}{\sqrt{2}\kappa\sigma_\eta}\right). \tag{28}$$

Candidate 1 wins the nomination if (27) is larger than (28) he wins more districts than Candidate 2, hence if

$$\Phi\left(\frac{\Phi^{-1}\left(\frac{1-\mu}{\mu}\right)-\kappa(v_2-v_3)}{\sqrt{2}\kappa\sigma_\eta}\right) > 2\Phi\left(\frac{-\Phi^{-1}\left(\frac{1-\mu}{\mu}\right)-\kappa(v_2-v_3)}{\sqrt{2}\kappa\sigma_\eta}\right). \tag{29}$$

Note that the left hand side is decreasing in μ, while the right hand side is increasing in μ. Thus, if (29) holds for a particular level of μ, then it also holds for all smaller levels of μ (equivalently, all higher levels of $1-\mu$). This is intuitive, since $1-\mu$ is the percentage of voters who support Candidate 1. Let μ^* denote the level of μ such that (29) holds with equality.

Consider first the case of $\mu = 1/2$, such that $\frac{1-\mu}{\mu} = 1$ and hence $\Phi^{-1}\left(\frac{1-\mu}{\mu}\right) = \infty$. Clearly, (29) holds, as the left hand side goes to 1, while the right hand side goes to 0. Intuitively, if $\mu = 1/2$, then any sort of vote-splitting between Candidates 2 and 3 guarantees that Candidate 1 wins all districts. Since both sides are continuous in μ, the same result holds (for any given v_2 and v_3) for μ sufficiently close to $1/2$. Now consider the case of $\mu = 2/3$, such that $\frac{1-\mu}{\mu} = 1/2$. Since $\Phi^{-1}(1/2) = 0$, (29) is clearly violated.

Consider now the effect of changes in σ_ε, σ_η and σ_v on (29). Note first that $\kappa = \frac{\sigma_v^2}{\sqrt{2}\sigma_\varepsilon({\sigma_v}^2+{\sigma_\eta}^2)}$ is decreasing in σ_ε and increasing in σ_v. Furthermore, the left hand side of (29) is decreasing in κ (as $(1-\mu)/\mu > 1/2$, and thus $\Phi^{-1}(\frac{1-\mu}{\mu}) > 0$), while the right hand side is increasing in κ by the same argument. Thus, to preserve equality between the two sides of (29), an increase of κ needs to be balanced by a decrease of μ^*. Consequently, μ^* decreases in σ_v, and increases in σ_ε.

We now analyze the effect of σ_η. Consider the difference of the left-hand and right-hand side of (29), and substitute for κ and set the expression equal to 0 (which implicitly determines the value of μ^*); this yields

$$Z = \Phi\left(\frac{\Phi^{-1}\left(\frac{1-\mu}{\mu}\right)}{\frac{\sigma_v^2\sigma_\eta}{\sigma_\varepsilon(\sigma_v^2+\sigma_\eta^2)}} - \frac{v_2-v_3}{\sqrt{2}\sigma_\eta}\right) - 2\Phi\left(\frac{-\Phi^{-1}\left(\frac{1-\mu}{\mu}\right)}{\frac{\sigma_v^2\sigma_\eta}{\sigma_\varepsilon(\sigma_v^2+\sigma_\eta^2)}} - \frac{v_2-v_3}{\sqrt{2}\sigma_\eta}\right) = 0. \tag{30}$$

Since $\Phi(\cdot)$ is an increasing function, $\Phi^{-1}\left(\frac{1-\mu}{\mu}\right)$ is decreasing in μ, and thus $\frac{\partial Z}{\partial \mu}$. Consequently, the sign of

$$\frac{d\mu^*}{\sigma_\eta} = -\frac{\frac{\partial Z}{\partial \sigma_\eta}}{\frac{\partial Z}{\partial \mu}}$$

is the same as the sign of $\frac{\partial Z}{\partial \sigma_\eta}$. We have

$$\begin{aligned}\frac{\partial Z}{\partial \sigma_\eta} =& \left[\phi\left(\frac{\Phi^{-1}\left(\frac{1-\mu}{\mu}\right)}{\frac{\sigma_v^2\sigma_\eta}{\sigma_\varepsilon(\sigma_v^2+\sigma_\eta^2)}} - \frac{v_2-v_3}{\sqrt{2}\sigma_\eta}\right) + 2\phi\left(\frac{\Phi^{-1}\left(-\frac{1-\mu}{\mu}\right)}{\frac{\sigma_v^2\sigma_\eta}{\sigma_\varepsilon(\sigma_v^2+\sigma_\eta^2)}} - \frac{v_2-v_3}{\sqrt{2}\sigma_\eta}\right)\right] \times \left[\frac{\sigma_\varepsilon}{\sigma_v^2}\Phi^{-1}\left(\frac{1-\mu}{\mu}\right)\frac{\sigma_\eta^2-\sigma_v^2}{\sigma_\eta^2}\right] + \\ & \left[\phi\left(\frac{\Phi^{-1}\left(\frac{1-\mu}{\mu}\right)}{\frac{\sigma_v^2\sigma_\eta}{\sigma_\varepsilon(\sigma_v^2+\sigma_\eta^2)}} - \frac{v_2-v_3}{\sqrt{2}\sigma_\eta}\right) - 2\phi\left(\frac{\Phi^{-1}\left(-\frac{1-\mu}{\mu}\right)}{\frac{\sigma_v^2\sigma_\eta}{\sigma_\varepsilon(\sigma_v^2+\sigma_\eta^2)}} - \frac{v_2-v_3}{\sqrt{2}\sigma_\eta}\right)\right] \times \frac{v_2-v_3}{\sqrt{2}\sigma_\eta^2}\end{aligned} \tag{31}$$

(31) is greater than

$$2\phi\left(\frac{\Phi^{-1}\left(-\frac{1-\mu}{\mu}\right)}{\frac{\sigma_v^2\sigma_\eta}{\sigma_\varepsilon(\sigma_v^2+\sigma_\eta^2)}}-\frac{v_2-v_3}{\sqrt{2}\sigma_\eta}\right)\left[\frac{\sigma_\varepsilon}{\sigma_v^2}\Phi^{-1}\left(\frac{1-\mu}{\mu}\right)\frac{\sigma_\eta^2-\sigma_v^2}{\sigma_\eta^2}-\frac{v_2-v_3}{\sqrt{2}\sigma_\eta^2}\right]$$

Since the term in square brackets goes to $\frac{\sigma_\varepsilon}{\sigma_v^2}\Phi^{-1}\left(\frac{1-\mu}{\mu}\right) > 0$ for $\sigma_\eta \to \infty$, (31) is positive for σ_η sufficiently large. Thus, for σ_η sufficiently large, $\frac{d\mu^*}{d\sigma_\eta}$ is positive. In contrast, for $v_2 = v_3$ and $\sigma_\eta < \sigma_v$, (31) and hence $\frac{d\mu^*}{d\sigma_\eta}$ is negative.

3. $\mu > 2/3$**.** In this case, Candidate 1 receives less than a third of the votes in every district, so that he loses in every district. Without loss of generality, suppose again that $v_2 > v_3$.

Under simultaneous elections, Candidate 2 wins in district s if

$$v_2 + \eta_2^s > v_3 + \eta_3^s. \tag{32}$$

Thus, for a given $v_2 > v_3$, the proportion of districts won by Candidate 2 is equal to $\Phi\left(\frac{v_2-v_3}{\sqrt{2}\sigma_\eta}\right) > 1/2$. Consequently, Candidate 2 is certain to win the nomination contest.

Under sequential elections, the winner of the first district (either Candidate 2 or Candidate 3) gets a vote share μ in all following districts and thus wins the nomination. The probability that Candidate 2 is the winner of the first district is the same as in (21) in Case 2 above. Thus, the better candidate is likely to win the nomination, but there is a positive probability that the other candidate with the same policy position wins instead. ■

9.2 Estimation Algorithm Details

The estimation algorithm proceeds as follows. Consider a given set of parameter values, $\tilde{\sigma}_v$, $\tilde{\sigma}_\eta$, $\tilde{\lambda}$, and $\tilde{S}_\mu$. These parameter values will be the initial values at the start of the algorithm, or intermediate values given by the Newton-Raphson optimization routine while the algorithm is in progress. We draw a set of R normally distributed valence draws, v_0^r, with $r = 1, \ldots R$, with mean zero and standard deviation $\tilde{\sigma}_v$. These valence draws are assigned to the candidate in position 0. We next draw a set of R pairs of normally distributed valence draws with the same standard deviation and mean. The highest of the two is labeled, v_{1a}^r and the lowest v_{1b}^r, corresponding to the highest and lowest valence candidates from position 1, respectively.

We consider a primary election with seven rounds. For the first five rounds, all three candidates compete with each other. In the last two rounds, only the candidates with valence draws v_0^r and v_{1a}^r compete with each other. For each round, we evaluate vote shares for 25 different values of μ^s, that are equally spaced on a grid and are given by $\mu^s = \frac{1-\tilde{S}_\mu}{2} + \frac{(g-1)\tilde{S}_\mu}{24}$, for $g = 1, \ldots, 25$. These values essentially discretize the distribution of μ^s and are used to compute expectations with respect to that distribution. For each value of μ^s and each set of valence draws, we compute vote shares on the basis of equation system 7.[33] Perceived valences are obtained on the basis of equations 10 and 11 (and their initial period variants) with signals drawn from the normal distribution centered around the true valence and with standard deviation $\tilde{\sigma}_\eta$. Each set of valences gets an independent set of signal histories for each of the 25 values of μ^s in the grid of μ. For the seventh round, perceived valences are assumed to be equal to the true valences.

This procedure returns seven matrices, each containing the vote shares of the candidate in political position 0 for each of the seven rounds. The rows of the matrix index different μ^s draws and the columns different valence draws. Index each of the seven matrices by $\rho = 1, \ldots, 7$, and their typical element by $w_\rho^{\iota,\nu}$. Note that our method of constructing vote shares for each round fixes the signal history for each value of μ^s. In other words, in each signal history and valence draw, the value of μ^s is held fixed. Therefore, the vote share paths are not representative of the actual vote share paths, for which the value of μ differs across states, and we cannot use any moments based on correlations or differences of vote shares across rounds. Our approach is valid for computing moments within a round since, as we pointed out in the text, vote shares in a particular state do not depend on voter preferences in preceding states, but only on the signals on the preceding states.

The value of $m1(\tilde{\sigma}_\upsilon, \tilde{\sigma}_\eta, \tilde{\lambda}, \tilde{S}_\mu)$ is computed by

$$m1(\tilde{\sigma}_\upsilon, \tilde{\sigma}_\eta, \tilde{\lambda}, \tilde{S}_\mu) = \frac{1}{5R} \sum_{\iota,\nu} \left\{ w_1^{\iota,\nu} + w_2^{\iota,\nu} + w_3^{\iota,\nu} + w_4^{\iota,\nu} + w_5^{\iota,\nu} \right\} \tag{33}$$

The value of $m2(\tilde{\sigma}_\upsilon, \tilde{\sigma}_\eta, \tilde{\lambda}, \tilde{S}_\mu)$ is computed by

$$m2(\tilde{\sigma}_\upsilon, \tilde{\sigma}_\eta, \tilde{\lambda}, \tilde{S}_\mu) = \frac{1}{45R} \sum_{\iota,\nu} \left\{ 22 w_6^{\iota,\nu} + 23 w_7^{\iota,\nu} \right\} \tag{34}$$

where the weights reflect the fact that there are 22 states in round 6 (Super Tuesday) and 23 states voting in round 7 (after Super Tuesday). The value of $m3(\tilde{\sigma}_\upsilon, \tilde{\sigma}_\eta, \tilde{\lambda}, \tilde{S}_\mu)$ is computed by

$$m3(\tilde{\sigma}_\upsilon, \tilde{\sigma}_\eta, \tilde{\lambda}, \tilde{S}_\mu) = \frac{1}{R} \sum_{\iota,\nu} \left\{ |w_6^{\iota,\nu} - \bar{w}_6^{\nu}| \right\} \tag{35}$$

where $\bar{w}_6^{\nu}$ is the average vote share in round 6 for a given set of valence draws and signals observed by voters in prior rounds, with the average taken over the different values of μ in the μ grid and signals observed by voters in the current round. In other words, in computing this average, the candidates as perceived by voters at the start of the round are held "fixed," but the voter preferences and signals in round 6 vary. This mimics the vote share process during Super Tuesday. Finally, the value of $m4(\tilde{\sigma}_\upsilon, \tilde{\lambda}, \tilde{S}_\mu)$ is computed by

$$m4(\tilde{\sigma}_\upsilon, \tilde{\lambda}, \tilde{S}_\mu) = \frac{1}{R} \sum_{\iota,\nu} \left\{ |w_7^{\iota,\nu} - \bar{w}_7^{\nu}| \right\} \tag{36}$$

[33] The distribution of ϵ is discretized and evaluated at 70 equally spaced points between -3.5 and 3.5, and the the sum of the probabilities adjusted to sum to unity.

where $\bar{w}_7^v$ is the average vote share in round 7 for a given set of valence draws with the average taken over the different values of μ in μ grid (recall that in round 7 we assume that valences are perfectly observed, which allows us to collapse all rounds following Super Tuesday into a single round; this assumption yields substantial computational savings). These moment values are used to calculate deviations the corresponding observed moment values in the data (reported in the equations 16 to 19). Parameter values are updated using the Newton-Raphson method until these deviations vanish (given exact identification, values for the four parameter values are found to exactly satisfy the four equation system). For the estimation, we use $R = 18,000$ resulting in very small sampling errors (on average about 5% of the standard error). This sampling error has been estimated by repeating the estimation for 26 different replications of the algorithm with different random seeds and $R = 3,000$. We calculated the standard deviation of the resulting estimates and used the fact that increasing the simulation draws by a factor of 6 decreases simulation error by a factor of $\sqrt{6} \approx 2.45$. The point estimates of the large run are within two standard deviations of the average estimates of the 26 short estimation results (generally within 5 percent of the standard error of the point estimates). Thus, there appears to be negligible simulation bias at this number of replications. Estimation time of all runs is in the order of three weeks in a personal computer using GAUSS.

State	round	Clinton	Edwards	Obama	State	round	Clinton	Edwards	Obama
Iowa	1	29	30	38	Utah	6	39	0	57
New Hampshire	2	39	17	37	Louisiana	7	36	0	57
Nevada	3	51	4	45	Maine	7	40	0	59
South Carolina	4	27	18	55	Nebraska	7	32	0	68
Florida	5	50	14	33	Washington	7	31	0	68
Alabama	6	42	0	56	DC	8	24	0	76
Alaska	6	25	0	75	Maryland	8	36	0	61
Arizona	6	50	0	43	Virginia	8	35	0	64
Arkansas	6	70	0	26	Hawaii	9	24	0	76
California	6	52	0	43	Wisconsin	9	41	0	58
Colorado	6	32	0	67	Ohio	10	53	0	45
Connecticut	6	47	0	51	Rhode Island	10	58	0	40
Delaware	6	43	0	53	Texas	10	51	0	48
Georgia	6	31	0	67	Vermont	10	39	0	59
Idaho	6	17	0	79	Wyoming	11	38	0	61
Illinois	6	33	0	65	Mississippi	12	37	0	61
Kansas	6	26	0	74	Pennsylvania	13	55	0	45
Massachusetts	6	56	0	41	Indiana	14	51	0	49
Minnesota	6	32	0	66	North Carolina	14	42	0	56
Missouri	6	48	0	49	West Virginia	15	67	0	26
New Jersey	6	54	0	44	Kentucky	16	65	0	30
New Mexico	6	49	0	48	Oregon	16	41	0	59
New York	6	57	0	40	Montana	17	41	0	57
North Dakota	6	37	0	61	South Dakota	17	55	0	45
Oklahoma	6	55	0	31					
Tennessee	6	54	0	41					

Table 4: 2008 Democratic primary election results