2009

Candidate Competition and Voter Learning in Sequential Primary Elections: Theory and Evidence

Mattias K Polborn
George Deltas, *University of Illinois at Urbana-Champaign*

Candidate Competition and Voter Learning in Sequential Primary Elections: Theory and Evidence

George Deltas
Department of Economics
University of Illinois
Champaign, IL 61820
deltas@uiuc.edu

Mattias K. Polborn
Department of Economics
University of Illinois
Champaign, IL 61820
polborn@uiuc.edu

August 2011

Abstract

We develop a model of sequential presidential primaries in which several horizontally and vertically differentiated candidates compete against each other. Voters are incompletely informed about candidate valence and learn over time from election results in previous districts. We analyze the effects of learning about candidate quality, and the effects of candidate withdrawal on the vote shares, using data from the 2000-2008 Democratic and Republican presidential primaries. Consistent with the predictions of the model, the withdrawal of a candidate has a bigger effect on the vote shares of candidates in the same political position, vote variability declines over time in a pattern consistent with learning, and a tilt of the electorate towards a particular political position disproportionately increases the vote shares of the weak candidates espousing that position (relative to the strong candidates in that position).

JEL Classification Numbers: D72, D60.
Keywords: Voting, primary elections, simultaneous versus sequential elections.

1 Introduction

Candidates for the U.S. presidential election are determined through a sequence of elections within each political party, the primaries. The nomination process is one of the most controversial institutions in the U.S. — in part, because it is the only major federal political institution that is not enshrined in the constitution, but rather managed by the two major parties in collaboration with the states. Because earlier primaries appear to be substantially more influential for the outcome of the nomination than later primaries, states compete to position their own primary early in the process. For example, in the 2008 presidential primaries, both Michigan and Florida attempted to "jump the queue" and moved their primary dates ahead to end of January; and the Democratic National Committee defended the sequential primary system by threatening to unseat the delegates elected in those states. Thus, a deeper understanding of the effects of a sequential versus a simultaneous primary system is particularly useful for guiding any attempt of an institutional reform.

We address one key feature of presidential primaries in our paper: At the beginning of the process, there are often more than just two candidates who compete with each other, and this situation generates coordination problems for voters and candidates that may result in the nomination of an inferior candidate or of a candidate who espouses a minority political view. However, in contrast to a simultaneous election in all states, the sequential primary system allows at least for some resolution of these problems over time, through relatively unsuccessful candidates dropping out.

In our model, candidates differ both "horizontally" (i.e., with respect to their policy positions) and "vertically" (i.e., with respect to their quality or valence). Specifically, there are two distinct policy positions for candidates and voters (in each party). For example, in the Republican party candidates may be either "moderates" or "conservatives", and each voter has a preference for one of these positions, which, however, is not absolute: If the voter considers a candidate in the other position to have a sufficiently higher valence, he would vote for that candidate rather than an ideologically closer competitor. While we assume that all voters know the candidates' horizontal positions from the start, valence is initially unknown, but voters in each state receive a signal about each candidate's valence. Moreover, voters in states that vote later can observe the election results in previous rounds and can use the information contained in previous election results to update their beliefs about candidates.

Since the quality of candidates is unknown and voters receive only imperfect and different signals, candidates who have the same policy position may split the votes of voters with a preference for their common position. For example, in the 2008 Republican primary, Mitt Romney felt that Mike Huckabee's presence in the competition made it impossible for him to unite the conservative wing of the Republican party behind him against John McCain. Romney first publicly called on Huckabee to drop out of the race, and, when this appeal was unsuccessful, withdrew himself.

From a theoretical perspective, this vote-splitting effect presents a substantial problem for the efficiency of any voting system, and not just for primaries. When more than two candidates run in an election, because a weaker candidate (i.e., not the Condorcet winner) might win in a situation where the Condorcet winner is splitting votes with a close ideological neighbor. The sequential U.S. presidential primary system provides a unique opportunity to gauge the presence and size of this vote-splitting effect, because some candidates drop out during the primaries, and those voters who would have voted for a dropped-out candidate need to choose which of the remaining candidates to support. Also, learning about candidate quality is just as important in simultaneous elections as in sequential ones, yet with all votes cast simultaneously, it is hard to disentangle the voters' policy preferences about candidates and their beliefs about candidate valences. By studying sequential primaries, our results inform our understanding of learning and inference in all election campaigns.

We derive several predictions from the theoretical model, and test them using data from the five contested U.S. Presidential primaries that took place in 2000, 2004 and 2008,[1] and show that the evidence is consistent with the model on a number of dimensions. For each party, a dichotomous partition of (serious) candidates on the basis of a priori positions in a set of "conservatives" and "moderates" for the Republican party, and "establishment" and "outsider" candidates for the Democratic Party, does well in predicting voter substitution patterns as candidates drop out over the course of the primaries.

The empirical evidence is broadly supportive of the following hypotheses derived from the

[1] George W. Bush was re-nominated in 2004 without any significant opponent, so that we do not include the 2004 Republican primaries in our analysis.

analysis of the theoretical model. First, if a candidate drops out, this benefits the remaining candidates who shared the drop-out's position more than it benefits candidates in the opposite position. This indicates that a crucial problem in multi-candidate primaries is that candidates who are ideologically close substitutes "steal" votes from each other, which may ultimately lead to the nomination of the "wrong" candidate; our econometric analysis allows us to gauge the size of this effect. Second, voter learning over time, facilitated through observation of previous election results, leads to reduced electoral variability over time. This effect can be measured without the use of parametric assumptions by utilizing the fact that many state contests are taking place on the same date. We show that the variability of voting shares, controlling for other factors through the use of election round fixed effects, decreases with the number of contests *prior* to a particular contest. Thus, as voters learn more about a candidate from coverage and campaigning in other states, they are less likely to be swayed by further information that emerges. Third, an increase of the share of voters who prefer a particular political position leads to a higher increase in the *absolute* number of votes for a strong candidate rather than a weak candidate in that position, but *relatively*, weak candidates benefit more than strong ones.

The paper proceeds as follows. In the next section, we review related literature. Section 3 presents the theoretical model, and in Section 4, we derive some testable hypotheses. We describe our data in Section 5, and present the results of the empirical analysis in Section 6.

2 Literature review

Several theoretical models analyze the sequential primary system. However, most papers focus on a contest between only two candidates, and therefore do not deal with the problem of vote-splitting between similar candidates that we focus on most in the present paper.

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 the voting equilibria of sequential elections are essentially the same as those in the case of simultaneous elections. Consequently, the temporal organization of elections does not matter in their model.

Callander (2007) and Ali and Kartik (2006) analyze different variations of this framework in which the outcome under simultaneous and sequential primaries can differ. Callander (2007) assumes that voters have a preference to vote for the eventual winner, and shows that this generates a bandwagon effect reminiscent of the momentum effects in U.S. primaries. Ali and Kartik (2006) stay close to Dekel and Piccione's setup, but point out that there are other equilibria in which voter herding in support of early winners can arise. Klumpp and Polborn (2006) provide a model that abstracts from the information aggregation aspect of voting and focuses purely on the candidates' actions. They demonstrate that candidates have an incentive to fight particularly hard for the first few districts, and that early wins endogenously increase a candidate's chance of winning in later elections.

Deltas, Herrera, and Polborn (2009) develop a theoretical model of primaries that with more than two candidates who are, as in our model, differentiated with respect to both their policy position and their valence. Some states vote at time $t = 0$, while others vote at $t = 1$. Voters in late states can observe the outcome of previous elections and can use results to facilitate coordination. In this setup, sequential primaries have both benefits (namely enhanced coordination possibilities) and costs (because coordination may occur on the "wrong" candidate when conditioning only on very few early primaries.

Going back to the pioneering studies of Bartels (1985, 1987, 1988), it has been widely accepted that ideological differences are not too important in primary contests. Indeed, to our knowledge, all theoretical models focus on voter learning about valence in a setting where voters care only about valence. Our empirical results strongly suggest that ideological differences between candidates matter substantially — voters view some candidates as closer substitutes than others. If this is indeed the case, then empirical models that ignore position differences may mistake ideological variation between sequentially voting states for learning about candidate valence.

Closest in many respects to our work is a recent paper by Knight and Schiff (2010) that focuses on voter learning about candidate quality in US primary style sequential elections. Like in our paper, the authors assume that voters in each state observe a state-specific signal about candidate quality. Unlike our model, they assume that a voter cannot observe the signals that voters of other states observed, even when these states vote before his own. Their assumption is based on the premise that voter preferences are not known, and thus a voter cannot back-out the signal by observing the vote outcome. Thus, the election results in a state carry only imperfect information about the signal observed by voters in that state. Using poll data from the 2004 US Democratic Presidential Primaries, they measure the extent to which voters update their beliefs about candidate quality by observing election returns in other states. They find that when this learning is assumed to be indirect, voters attach a substantial weight on the outcomes of early elections, but a much smaller weight after the fourth primary. Thus, in their framework, predicted share volatility declines up to the fifth primary round, but is essentially constant thereafter. Our empirical strategy is agnostic about whether a voter in a state infers perfectly or noisily the signal that voters in other states have observed by the voting outcome in that state. However, our results suggest that much of this signal is directly observed (as in our model) given that share volatility falls throughout the primary season, and not only after the first few election contests.

More generally, the ultimate objective of our paper is also related to a literature that compares election outcomes under different voting systems. For example, Merrill (1984) simulates multicandidate elections under plurality, runoff Borda, approval voting and under the Hare, Coombs and Black methods and compares how often the Condorcet winner is selected by the different methods (see also Chamberlin and Featherston (1986), Merrill (1985), Nurmi (1992), and Lijphart and Grofman (1984)). The comparison of voting systems in most of this literature is based on Monte-Carlo simulations of elections using the "impartial cultures" assumption which supposes that every possible preference profile over candidates is equally likely to occur. This is clearly not satisfied in our framework (and, in our opinion, would also not be desirable, because in our application, some preference profiles are much more plausible than others). In contrast, our comparison between a sequential and a counterfactual simultaneous primary system is based on our empirical analysis that uses observations from actual elections.[2]

[2]Leininger (1993) follows a conceptually similar approach to ours in a different setting. He uses observed voting behavior of legislators in the 1991 Bundestag (on the question which city should be Germany's capital) to reconstruct individuals' preferences, and finds that the final outcome would have changed if the agenda (the sequence of votes) had been different.

3 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. States vote sequentially (though some states may vote at the same time). Voters 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 binary characteristic on which candidates are exogenously fixed either to position 0 or to position 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 fixed characteristic can be thought of as arising from the candidate's history and cannot be changed at the time of the election. The assumption that it is binary is meant to capture the idea that some candidates are very similar to each other and hence close (policy) substitutes for most voters, while there is a substantial difference to some other candidates. Other issues are treated stochastically via the incorporation of a composite preference shock, as detailed below. The assumption that differences can be expressed in binary form follows Krasa and Polborn (2007), and the assumption that there is only one major issue greatly simplifies the empirical analysis.

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 fixed characteristic, and λ measures the weight of the fixed characteristic 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.[3] A possible interpretation is that candidates

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

also differ in a large number of other dimensions for which voters have different preferences. In this case, the fixed characteristic modeled explicitly ($a_j = 0$ or $a_j = 1$) should be understood as the most important policy dimension.

There is uncertainty about the valence of the candidates. 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 electoral district s observe a signal $Z_j^s = v_j + \eta_j^s$ about candidate j, where the additional term, η_j^s, is an independent draw from a normal distribution $N(0, {\sigma_\eta}^2)$. Note that η_j^s is a district-specific observation error (as opposed to a 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 the errors (if any) are not individual-specific. Note that, if instead observation error terms were individual-specific, then the true valence of candidates would be known after the election results of the first district. This appears unrealistic.[4]

Also, we assume that signals are state-specific rather than national, which has the consequence that election results are informative for voters in later states. Even if information arrives from national news media, it may be interpreted differently in different states because of experience voters in these states have with politicians adopting the same rhetoric/positions. 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 beliefs. Let $\hat{v}_j^s$ denote the valence of candidate j that is expected by voters in district s. Each voter votes sincerely.[5] That is, voter i in district s (which votes at time t) votes for candidate j if and only if

$$j \in \arg\max_{j' \in \mathcal{J}^t} \hat{v}_{j'}^s - \lambda |a_{j'} - \theta^i| + \varepsilon_{j'}^i, \tag{2}$$

where $\mathcal{J}^t$ is the set of candidates in period t elections.[6]

[4]Of course, it appears plausible that, in reality, there is both a common as well as an idiosyncratic observation error. To simplify the model and gain some tractability, we focus on the state-specific observation error.

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

[6]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.

Since we focus on the implications of voters' learning behavior, the specific rules for who wins the nomination (e.g., the candidate who wins the most states or the candidate who wins the highest average vote share) do not matter and we can therefore be silent on this. In practice, Democrats mostly have a system in which average vote share matters most (since, in each state, candidates receive delegates roughly proportional to their vote share in that state), while Republicans operate primarily under "winner-take-all" rules within each state.

4 Analysis of the model

4.1 Vote Shares

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 $\hat{v}^s = (\hat{v}_1^s, \hat{v}_2^s, \ldots, \hat{v}_J^s)$. In the next subsection, we will then turn to the determination of $\hat{v}^s$.

Let $\Phi(\cdot)$ denote the cumulative distribution of the standard normal distribution $N(0,1)$, and let $\phi_\varepsilon(\cdot)$ denote the density of $N(0,\sigma_\varepsilon)$. 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.

Our objective is to find the total number of votes that a candidate receives in this setting. Let $d(j,\theta)$ measure 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). A voter of type θ votes for Candidate $j \in J_0^s$ if and only if

$$\hat{v}_j^s + \varepsilon_j - \lambda d(j,\theta) \geq \max_{j'}(\hat{v}_{j'}^s + \varepsilon_{j'} - \lambda d(j',\theta)). \tag{3}$$

For a given ε_j, (3) is satisfied if and only if

$$\varepsilon_{j'} < \hat{v}_j^s - \hat{v}_{j'}^s + \varepsilon_j - \lambda[d(j,\theta) - d(j',\theta)] \text{ for all } j' \neq j \ . \tag{4}$$

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

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

Integrating over ε_j gives that the proportion of type 0 voters who vote for candidate j is

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

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

$$\int_{-\infty}^{\infty} \prod_{J_0^s \setminus \{j\}} \Phi\left(\frac{\hat{v}_j^s - \hat{v}_{j'}^s + \varepsilon_j}{\sigma_\varepsilon}\right) \cdot \prod_{J_1^s} \Phi\left(\frac{-\lambda + \varepsilon_j + \hat{v}_j^s - \hat{v}_{j'}^s}{\sigma_\varepsilon}\right) \cdot \phi_\varepsilon(\varepsilon_j) d\varepsilon_j. \tag{7}$$

The total vote share of candidate $j \in J_0^s$ is then

$$\begin{aligned} (1-\mu^s) \int_{-\infty}^{\infty} \prod_{J_0^s \setminus \{j\}} \Phi\left(\frac{\hat{v}_j^s - \hat{v}_{j'}^s + \varepsilon_j}{\sigma_\varepsilon}\right) \cdot \prod_{J_1} \Phi\left(\frac{\lambda + \varepsilon_j + \hat{v}_j^s - \hat{v}_{j'}^s}{\sigma_\varepsilon}\right) \cdot \phi_\varepsilon(\varepsilon_j) d\varepsilon_j + \\ \mu^s \int_{-\infty}^{\infty} \prod_{J_0^s \setminus \{j\}} \Phi\left(\frac{\hat{v}_j^s - \hat{v}_{j'}^s + \varepsilon_j}{\sigma_\varepsilon}\right) \cdot \prod_{J_1^s} \Phi\left(\frac{-\lambda + \varepsilon_j + \hat{v}_j^s - \hat{v}_{j'}^s}{\sigma_\varepsilon}\right) \cdot \phi_\varepsilon(\varepsilon_j) d\varepsilon_j \end{aligned} \tag{8}$$

Similarly, the vote share of a candidate $j \in J_1^s$ is

$$\begin{aligned} (1-\mu^s) \int_{-\infty}^{\infty} \prod_{J_0^s} \Phi\left(\frac{-\lambda + \hat{v}_j^s - \hat{v}_{j'}^s + \varepsilon_j}{\sigma_\varepsilon}\right) \cdot \prod_{J_1^s \setminus \{j\}} \Phi\left(\frac{\varepsilon_j + \hat{v}_j^s - \hat{v}_{j'}^s}{\sigma_\varepsilon}\right) \cdot \phi_\varepsilon(\varepsilon_j) d\varepsilon_j + \\ \mu^s \int_{-\infty}^{\infty} \prod_{J_0^s} \Phi\left(\frac{\lambda + \hat{v}_j^s - \hat{v}_{j'}^s + \varepsilon_j}{\sigma_\varepsilon}\right) \cdot \prod_{J_1^s \setminus \{j\}} \Phi\left(\frac{\varepsilon_j + \hat{v}_j^s - \hat{v}_{j'}^s}{\sigma_\varepsilon}\right) \cdot \phi_\varepsilon(\varepsilon_j) d\varepsilon_j \end{aligned} \tag{9}$$

4.2 Effect of drop-outs

At some point after the first elections, some candidate(s) may choose to drop out of the race. In the analysis below, the reason for withdrawing from the race is immaterial, and there could be a variety of withdrawal reasons: For example, candidates who were unsuccessful in early elections may have difficulty raising campaign contributions required for competing successfully at the second stage. Also, there may be exogenous reasons (e.g., health shocks, family reasons, a change of mind as the campaign unfolds, etc.) for candidates to withdraw.

We are interested in how the withdrawal of one candidate affects the vote shares of the remaining contenders. Specifically, consider a situation in which there are initially three candidates, two of whom (say, A and B) have position 0, while the third one (C) has position 1. What happens to the support of candidates B and C, when candidate A drops out?

It is useful to define the total number of voters who rank candidate A highest and candidate B second as R_{AB}; furthermore, let R_{AC} be defined analogously. In the Appendix, we show that

$$\begin{aligned} R_{AB} = (1-\mu)\int_{-\infty}^{\infty}\left[\Phi\left(\frac{\hat{v}_A-\hat{v}_B-\varepsilon}{\sigma_\varepsilon}\right)\Phi\left(\frac{\hat{v}_B-\hat{v}_C+\lambda+\varepsilon}{\sigma_\varepsilon}\right)\right]\phi_\varepsilon(\varepsilon)d\varepsilon + \\ \mu\int_{-\infty}^{\infty}\left[\Phi\left(\frac{\hat{v}_A-\hat{v}_B-\varepsilon}{\sigma_\varepsilon}\right)\Phi\left(\frac{\hat{v}_B-\hat{v}_C-\lambda+\varepsilon}{\sigma_\varepsilon}\right)\right]\phi_\varepsilon(\varepsilon)d\varepsilon \end{aligned} \tag{10}$$

and

$$\begin{aligned} R_{AC} = (1-\mu)\int_{-\infty}^{\infty}\left[\Phi\left(\frac{\hat{v}_A-\hat{v}_C+\lambda+\varepsilon}{\sigma_\varepsilon}\right)\Phi\left(\frac{\hat{v}_C-\hat{v}_B-\lambda-\varepsilon}{\sigma_\varepsilon}\right)\right]\phi_\varepsilon(\varepsilon)d\varepsilon + \\ \mu\int_{-\infty}^{\infty}\left[\Phi\left(\frac{\hat{v}_A-\hat{v}_C-\lambda+\varepsilon}{\sigma_\varepsilon}\right)\Phi\left(\frac{\hat{v}_C-\hat{v}_B+\lambda-\varepsilon}{\sigma_\varepsilon}\right)\right]\phi_\varepsilon(\varepsilon)d\varepsilon \end{aligned} \tag{11}$$

Whenever the ratio of (10) and (11) is greater than 1, B profits more than C from A's withdrawal, and vice versa. In general, the ratio R_{AB}/R_{AC} can be larger or smaller than 1, as Figure 1 demonstrates (for parameters). Figure 1 suggests that the ratio is monotonically in-

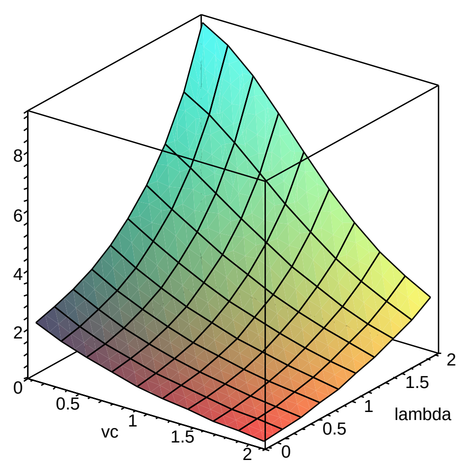


Figure 1: Ratio R_{AB}/R_{AC} for $\mu = 1/2$, $\hat{v}_A = 0$, $\hat{v}_B = 1$

creasing in λ and monotonically decreasing in $\hat{v}_C$. This makes intuitive sense: First, the more important position differences are for voters ($\lambda \uparrow$), the more candidate B (the candidate in the same position as the exiting candidate A) benefits relative to candidate C. In the limit when issue differences become very important ($\lambda \to \infty$), then both terms in (11) go to zero, while (10) goes to $(1-\mu)\int_{-\infty}^{\infty}\Phi\left(\frac{\hat{v}_A-\hat{v}_B-\varepsilon}{\sigma_\varepsilon}\right)\phi_\varepsilon(\varepsilon)d\varepsilon$, which is equal to the total number of previous voters for candidate A. Intuitively, all previous supporters of candidate A now switch to candidate B, irrespective of the valences of candidates B and C. Second, the larger is $\hat{v}_C$, the more attractive is candidate C for the previous supporters of candidate A, and thus, the smaller is the ratio of vote gains for B relative to those for C.[7]

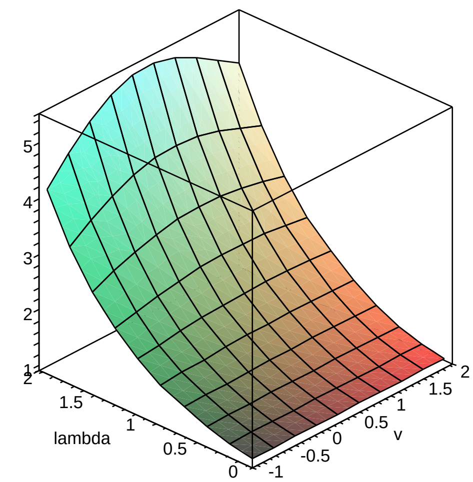


Figure 2: Ratio R_{AB}/R_{AC} for $\mu = 1/2$, $\hat{v}_A = 0$, $\hat{v}_B = \hat{v}_C = v$

Figure 2 focuses on parameters such that the two surviving candidates B and C have the same valence, while still normalizing the valence of candidate A to zero. For all values of λ and the common valence of B and C, we see that candidate B wins more votes than candidate C from a withdrawal of candidate A. Assuming that R_{AB}/R_{AC} is in fact monotone in $\hat{v}_C$, the same result would hold whenever $\hat{v}_B \geq \hat{v}_C$, and, to the extent that $\lambda > 0$, it would also be the case for some parameter values such that $\hat{v}_B < \hat{v}_C$.[8]

The assumption in Figure 2 that B and C have the same valence appears as a useful benchmark case because candidate valences are drawn from the same distribution, and under any kind of nonnegative selection (i.e. the fact that A rather than B drops out implies something positive about B), the valence of B should, on average,[9] be equal or higher than the valence

[7]While intuitively straightforward, proving that these monotonicity relations hold in general is so cumbersome that it is not worthwhile.

[8]Note, however, that if $\hat{v}_B << \hat{v}_C$, then it is possible that candidate C gains more votes from A's withdrawal than candidate B.

[9]This average refers to a number of different primary runs, each with a new draw of candidates.

of C. As argued above, in the case that $\hat{v}_B \geq \hat{v}_C$, our conclusion that candidate B benefits more than his surviving competitor, if a candidate who is located in the same political position withdraws from the race, is strengthened further.

Similar effects are likely to arise with more than three candidates. Regardless of the number of candidates in each position, as λ goes to infinity, no voter of a candidate in position 0 would switch to a candidate in position 1 if his preferred candidate were to withdraw. Similarly, as λ goes to zero, candidates in position 0 are stochastically identical to candidates in position 1 for all voters, so that, in expectation, the withdrawal of a candidate in position 0 has no systematically different effect on the vote shares of the remaining candidates with regards to their position.

In summary, the analysis in this subsection suggests the following hypothesis.

Hypothesis 1 *If a candidate in position* 0 *withdraws, the expected increase in votes is larger for the remaining candidates in position* 0 *than for those in the other position, and similarly for a withdrawal of a candidate in position* 1.

4.3 The effects of learning candidate valence over time

We now discuss voter updating about valence. Recall that voters in each state receive a normally distributed signal of candidate j's valence with expected value v_j and variance σ_η^2.

Suppose the ex-ante belief about candidate j's valence (i.e., before seeing his own state-specific signal) is distributed according to $N(\hat{v}_{j0}, \sigma_{j0}^2)$. If voters in state s receive a state-specific signal Z_j^s, one 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

$$\hat{v}_j^s = \frac{\sigma_\eta^2}{\sigma_{j0}^2 + \sigma_\eta^2} v_{j0} + \frac{\sigma_{j0}^2}{\sigma_{j0}^2 + \sigma_\eta^2} Z_j^s \tag{12}$$

and variance

$$(\sigma_{v_j}^s)^2 = \frac{\sigma_{j0}^2 \sigma_\eta^2}{\sigma_{j0}^2 + \sigma_\eta^2}. \tag{13}$$

Clearly, in the initial state(s), $\hat{v}_{j0} = 0$ and $\sigma_{j0}^2 = \sigma_v^2$. What is the information of voters in states voting later before they see their state's signal? Remember that these voters observe the vote share of each candidate j in each earlier state r, W_j^r, and know μ^r. Using (8) and (9), the election in state r is then captured by the following equation system:

$$(1-\mu^r)\int_{-\infty}^{\infty}\prod_{J_0^r\setminus\{j\}}\Phi\left(\frac{\hat{v}_j^r-\hat{v}_{j'}^r+\varepsilon_j}{\sigma_\varepsilon}\right)\cdot\prod_{J_1^r}\Phi\left(\frac{\lambda+\varepsilon_j+\hat{v}_j^r-\hat{v}_{j'}^r}{\sigma_\varepsilon}\right)\cdot\phi_\varepsilon(\varepsilon_j)d\varepsilon_j+$$

$$\mu^r\int_{-\infty}^{\infty}\prod_{J_0^r\setminus\{j\}}\Phi\left(\frac{\hat{v}_j^r-\hat{v}_{j'}^r+\varepsilon_j}{\sigma_\varepsilon}\right)\cdot\prod_{J_1^r}\Phi\left(\frac{-\lambda+\varepsilon_j+\hat{v}_j^r-\hat{v}_{j'}^r}{\sigma_\varepsilon}\right)\cdot\phi_\varepsilon(\varepsilon_j)d\varepsilon_j=W_j^r\ ,\forall j\in J_0^r$$

$$(1-\mu^r)\int_{-\infty}^{\infty}\prod_{J_0^r}\Phi\left(\frac{-\lambda+\hat{v}_j^r-\hat{v}_{j'}^r+\varepsilon_j}{\sigma_\varepsilon}\right)\cdot\prod_{J_1^r\setminus\{j\}}\Phi\left(\frac{\varepsilon_j+\hat{v}_j^r-\hat{v}_{j'}^r}{\sigma_\varepsilon}\right)\cdot\phi_\varepsilon(\varepsilon_j)d\varepsilon_j+$$

$$\mu^r\int_{-\infty}^{\infty}\prod_{J_0^r}\Phi\left(\frac{\lambda+\hat{v}_j^r-\hat{v}_{j'}^r+\varepsilon_j}{\sigma_\varepsilon}\right)\cdot\prod_{J_1^r\setminus\{j\}}\Phi\left(\frac{\varepsilon_j+\hat{v}_j^r-\hat{v}_{j'}^r}{\sigma_\varepsilon}\right)\cdot\phi_\varepsilon(\varepsilon_j)d\varepsilon_j=W_j^r\ ,\forall j\in J_1^r \quad (14)$$

The following proposition shows that observing the vote shares of all candidates in district r allows voters in later states to essentially recover the valence signal of state r.

Proposition 1 *There exists a unique vector* $(0, x_2, x_3, \ldots x_k)$ *such that all solutions of* (14) *are of the form* $(0, x_2, x_3, \ldots, x_k) + (c, c, \ldots, c)$, $c \in \mathbb{R}$.

Proof. See Appendix. ■

It is immaterial which of these possible solutions to (14) a voter in a later state takes as his ex-ante belief, as a shift of the ex-ante beliefs (about all candidates) by c translates into a shift of the ex-post beliefs by $\frac{\sigma_\eta^2}{\sigma_{j0}^2+\sigma_\eta^2}c$ for each candidate, leaving the difference between the valence estimates for the different candidates, and hence the voter's voting decision, unaffected. The vote shares are determined only by the *difference* between the candidates' valences, so we can normalize candidate 1's estimated valence to zero.

Our next result, Proposition 2, shows that, as the primaries progress, the variation of beliefs about candidate valences across those states that vote at the same time diminishes. This is intuitive since states that vote late in the process share a lot of common information (i.e., the information inferred from states that vote earlier), and thus, the differences in beliefs generated by the fact that each state receives its own state-specific signal are not as large as they are in early states.

Proposition 2 *Consider the expected variance of the valence estimates in all states that vote at time t. This variance is decreasing in t.*

Proof. See Appendix. ■

Intuitively, a lower variance of the valence estimates in later states translates into a lower variance of a candidate's vote shares in late states, relative to early states. In particular, this is clear in the limit: If there is (almost) no remaining uncertainty about candidates' valences, then vote shares in late states depend only on μ^s and are otherwise completely deterministic. Any randomness in the valence estimate across late states must increase the variance of the candidates' vote shares. Proposition 2 thus suggests Hypothesis 2 below.

Hypothesis 2 *Consider the variance of a candidate's vote shares in all those elections that occur on one date. This variance is decreasing over time.*

4.4 Effect of partisan composition

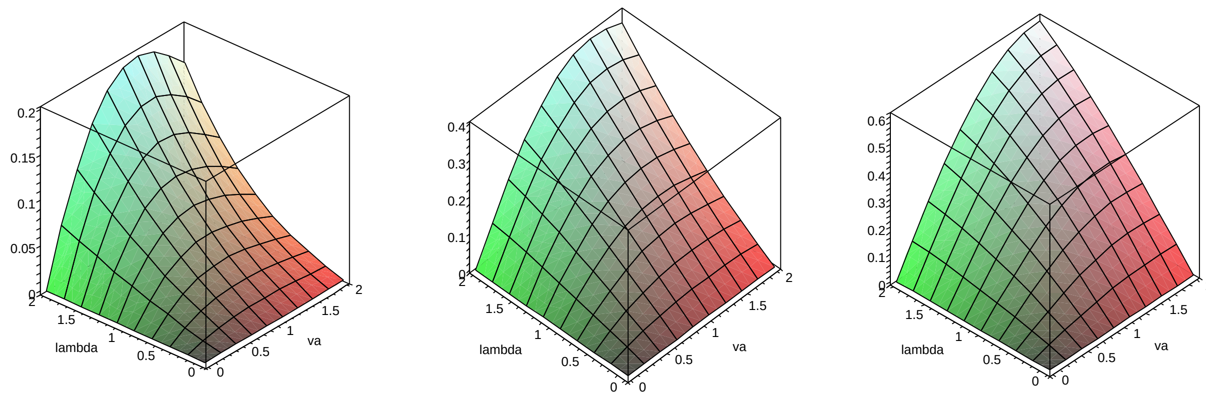
Finally, we consider the effect that the level of μ in different states has on the support of different candidates. It is quite obvious (and we skip a formal proof) that an increase in $1-\mu$, the number of voters with a preference for position $a = 0$, increases the vote share of all candidates with position $a = 0$, and decreases the vote share of all candidates with position $a = 1$. It is less obvious, though, which candidate among those with position $a = 0$ gains most, both absolutely (i.e., in terms of percentage point increase in vote share) and relatively (i.e., the increase in vote share relative to the previous level).

To analyze this question, let us focus on the case where there are initially three candidates, two of whom (say, A and B) have position 0, while the third one (C) has position 1. A decrease in μ benefits the vote shares of candidates A and B. Candidate A benefits at least as much as candidate B if and only if

$$\begin{aligned}\int_{-\infty}^{\infty}\Phi\left(\frac{v_A-v_B+\varepsilon}{\sigma_\varepsilon}\right)\left[\Phi\left(\frac{\lambda+v_A-v_C+\varepsilon}{\sigma_\varepsilon}\right)-\Phi\left(\frac{-\lambda+v_A-v_C+\varepsilon}{\sigma_\varepsilon}\right)\right]\phi_\varepsilon(\varepsilon)d\varepsilon- \\ \int_{-\infty}^{\infty}\Phi\left(\frac{v_B-v_A+\varepsilon}{\sigma_\varepsilon}\right)\left[\Phi\left(\frac{\lambda+v_B-v_C+\varepsilon}{\sigma_\varepsilon}\right)-\Phi\left(\frac{-\lambda+v_B-v_C+\varepsilon}{\sigma_\varepsilon}\right)\right]\phi_\varepsilon(\varepsilon)d\varepsilon\geq 0.\end{aligned} \tag{15}$$

Without loss of generality, suppose that $v_A > v_B$. Whether (15) holds in general is difficult to determine. However, for $\lambda = 0$, (15) obviously holds as equality, and for λ sufficiently large, the left-hand and right-hand sides go to $\int_{-\infty}^{\infty}\Phi\left(\frac{v_A-v_B+\varepsilon}{\sigma_\varepsilon}\right)\phi_\varepsilon(\varepsilon)d\varepsilon$ and $\int_{-\infty}^{\infty}\Phi\left(\frac{v_B-v_A+\varepsilon}{\sigma_\varepsilon}\right)\phi_\varepsilon(\varepsilon)d\varepsilon$, so that (15) is satisfied as strict inequality.

Figure 3 displays the left-hand side of (15), when we normalize $\sigma_\varepsilon = 1$ and $v_B = 0$. The three graphs show the variation for different values of λ and v_A, for $v_C = -1, 0, 1$, respectively. In all graphs, the left-hand side of (15) is increasing in λ, so that it appears plausible that (15) holds for any $\lambda > 0$.



(a) $v_C = -1$ (b) $v_C = 0$ (c) $v_C = 1$

Figure 3: Difference in vote gains (stronger candidate minus weaker candidate) for $v_B = 0$ and $v_C = -1, 0, 1$

We now focus on relative changes. Proposition 3 shows that, if λ is sufficiently large, then the weaker candidate benefits *proportionately more* than the strong candidate (i.e., relative to previous vote share) from an increase in the number of voters who prefer their joint position.

Proposition 3 *Suppose that both candidate A and B are in position 0, while candidate C is in position 1. Furthermore, suppose that $\hat{v}_A > \hat{v}_B$. There exists λ^* such that for all $\lambda \geq \lambda^*$, an increase in $1 - \mu$ increases the vote share of B by a larger percentage than the vote share of A (relative to their respective previous vote shares).*

Proof. See Appendix. ■

We conjecture that Proposition 3 holds more generally, for any λ, but again this is hard to prove. Intuitively, for $\lambda = 0$ (i.e., positions do not matter for voters), a change in μ is immaterial for vote shares, and it appears plausible that the relative effect is monotonous in λ, and since we know that Proposition 3 holds for λ sufficiently large, it would hold for all values of λ.

The following hypothesis summarizes the results regarding which candidate benefits more, both absolutely and relatively, from an increase in the number of voters who prefer their joint position.

Hypothesis 3 *Consider two candidates A and B with the same type, and suppose that $v_A > v_B$. If the percentage of voters who prefer their common position is larger in state s than in state s', then*

1. *The expected difference between A's and B's vote share is larger in state s.*
2. *The share of candidate B (the weak candidate) increases proportionately more in state s than the share of candidate A (the strong candidate).*

5 Data

Our dataset consists of information from five of the 2000, 2004, and 2008 United States Presidential primaries (we exclude the 2004 Republican primary in which George W. Bush was effectively unopposed). Our theory presumes that all candidates are initially considered viable candidates in the sense that there is some chance that they will win their party's nomination. In practice, some of the candidates who run in the primaries do not fall into this category, but rather run to represent a particular constituency in order to demonstrate that the party needs to pay attention to its preferences. Several of these candidates are "perennials" who run in many elections with typically very poor results. These candidacies do not fit our theoretical model well, as there is no meaningful way in which voters update their beliefs about them. Indeed, many of these candidates run for most if not the entire length of the campaign, and the voters have a reasonably good idea about their ability and views (especially for those who have run repeatedly). Therefore, these candidates are not included in our dataset.

The most successful candidates that we excluded are probably Dennis Kucinich (Democratic primary 2004 and 2008) and Ron Paul (Republican primary 2008). Both represent, in their respective parties, a non-trivially sized and energetic constituency, but the ideological distance to their parties' mainstream was too large for them ever to be considered viable candidates for their party's nomination. Their performance is usually better in low-turnout contests later in the sequence in which their energized base represents a larger fraction of the electorate. In contrast, unsuccessful but "serious" candidates (for example, Joe Lieberman (D-2004) or Rudy Giuliani (R-2008)) have their best performances in early primaries, then lose voter support due to their relatively poor performance, and eventually drop out once it becomes clear that they have no chance of winning the nomination. Tables 1 and 2 list the candidates we include for each primary, along with the states in which they competed and the vote share they obtained. The tables also give the number of different election dates (rounds) up to the election in each state.

Finally, a key component of the model is that candidates of each party are distinguished by their political location or position. Though there are many differences between candidates, we believe that for each party there is a single most important summary representation of these differences. In the Republican party, it appears to us that the main ideological fault line is between *conservatives* (i.e., candidates and voters who often have a fundamentalist Christian background and emphasize "value-issues" such as abortion and gay marriage) and *moderates*. A standard approach to determining a candidate's position is the use of NOMINATE scores based on roll-call votes (see Poole and Rosenthal (1985)). However, such scores are only available for legislators, and the majority of candidates has an executive background (e.g., former governors). Our classification is therefore guided by common sense and exit polls that ask voters which candidate they voted for, and whether they personally identify as conservative, moderate or liberal. We focus on exit polls in early primary or caucus states, as these are usually the only ones in which all candidates we consider are running and where each of them receives a sufficiently large vote share. For example, in the 2000 Republican contest, George W. Bush did considerably better with voters who identified as conservative rather than with those who said they were moderate, and vice versa for John McCain.[10] For this reason, we classify Bush as conservative and McCain as moderate. In 2008, we take the MSNBC exit polls (available on http://www.msnbc.msn.com/id/21660890), since they ask voters to identify as conservative, moderate or liberal, while CNN has dropped this question in many exit polls). McCain and Giuliani always do considerably better with voters who identify as moderates, while Huckabee and Thompson do considerably better with conservatives. Romney generally

[10]See, e.g., http://www.cnn.com/ELECTION/2000/primaries/NH/poll.rep.html, http://www.cnn.com/ELECTION/2000/primaries/SC/poll.rep.html, http://www.cnn.com/ELECTION/2000/primaries/IA/poll.html. In the 2000 Republican primary, we also identify Steve Forbes and Alan Keyes as conservatives, as they also do better with self-identified conservative voters.

does better with conservatives than with moderates, except for states in which the Republican primary electorate is extremely conservative. For example, in Iowa, 88 percent of Republican primary voters identify as strongly or somewhat conservative, while only 11 percent declare as moderates. Romney receives about the same vote share from conservatives and moderates (25 percent versus 26 percent). However, in states like Michigan or Florida where the percentage of conservatives in around 60, Romney does substantially better with conservatives than with moderates. Moreover, in the later stages of the campaign, Romney was perceived to fight with Huckabee over the conservative vote.[11] For this reason, we classify Romney as conservative.

It would be tempting to attempt a formally analogous classification of Democratic candidates as "liberal" or "moderate". However, for Democrats, the ideological position of the voter appears to have much less predictive power. For example, in Nevada, self-declared liberals voted 48/39/9 for Clinton, Obama and Edwards, while moderates voted 46/43/8 for these candidates. This difference between liberals and moderates is well within the margin of error. A considerably better sorting is achieved by a question that asks voters which candidate qualities matter most: "Has the necessary experience," "Can achieve the necessary change," "Cares about people like me" or "Can win in November." Leaving out the last category (since this is mostly concerned with the horserace aspect of politics, rather than policy preferences), we would argue that people who consider "experience" most important have a preference for Washington insiders, while those who appreciate "change" or "caring" candidates prefer outsiders. On the basis of this question in the MSNBC exit polls in early states, we classify Clinton as insider and Edwards and Obama as outsiders in 2008. In 2004, Kerry receives the largest share from voters who name "experience" as the most important quality,[12] while the outsider/populist categories ("cares about people like me," "takes strong stands," "can shake things up") goes predominantly to Edwards and Dean. Both Lieberman and Clark do not register at sufficiently high levels in many states to draw strong conclusions from exit polls. We use our judgment to categorize Lieberman (the 2000 Democratic vice-presidential candidate) as insider, and Clark (an anti-war general who had never run for office before) as an outsider. By a similar argument, we classify Gore as insider and Bradley as outsider in the 2000 election. A summary of our candidate partition is in bottom of Tables 1 and 2.

[11] See, e.g., http://www.cnn.com/2008/POLITICS/02/05/super.exit/.

[12] For example, see http://www.cnn.com/ELECTION/2004/primaries/pages/epolls/IA/index.html.

For these candidates and election contests, we obtain the vote percentage in the primary or caucus of each state from the Federal Election Commission. These vote shares are reported in Tables 1 and 2. However, these shares do not sum up to 100 percent as they include votes for candidates whom we dropped from our analysis, for candidates who have already withdrawn, or for "uncommitted" delegates. To ensure that vote shares representing serious votes sum up to 100% (as assumed by the model), we rescale all the vote shares accordingly for the purpose of econometric analysis. We supplement these data on the Presidential primaries with data from the 1992 Presidential election.[13] The vote shares of the Presidential candidates Clinton and Perot are used as variables that are correlated with a state's ideological position. A high Perot vote share is expected to be associated with populist preferences, while a high Clinton share in that 3-way race is expected to be associated with liberal preferences. This data is also reported in Table 1.

6 Results

6.1 Non-Parametric Mean-Variance Analysis

We start our data analysis by pooling all data and making simple comparisons of average vote shares of candidates as a function of the distribution of candidates in political positions. By pooling, we mean that we do not distinguish between parties, political positions within parties and the position of a state within the sequence of the primary, but rather analyze all primary elections in which κ candidates in one position and κ' candidates in the opposite position compete. An advantage of this approach is that it is based on minimal assumptions and thus provides important supportive evidence for our theoretical framework. However, it is informal in nature (no formal statistical testing is performed) and could be missing systematic effects (e.g., differences in mean vote shares for different locations, differences across parties, etc.). We discuss these limitations in more detail at the end of this subsection.

Let $\text{VoteShare}_{j,y}^{\kappa,\kappa'}$ be the vote share of candidate j (measured on a 0-100% scale) who shares his position $a(j)$ with $\kappa - 1$ other candidates, while there are κ' competitors in the opposite

[13]The 1992 general election results were obtained from Dave Leip's Atlas of U.S. Presidential Elections, available at http://www.uselectionatlas.org/.

position $|1-a(j)|$. Formally, let

$$\text{VoteShare}_{j,y}^{\kappa,\kappa'} = \frac{1}{N^{\kappa,\kappa'}} \sum_{j,s,y:(\left\|K_{s,p,a(j),y}\right\|=\kappa \wedge \left\|K_{s,p,|1-a(j)|,y}\right\|=\kappa')} \text{VoteShare}_{j,s,y} \tag{16}$$

where $\|K_{s,p,l,y}\|$ is the cardinality of the set of candidates in state contest s, political party p, political location l, and year y, $N^{\kappa,\kappa'}$ is the number of observations such that $\|K_{s,p,0,y} = \kappa\|$ and $\|K_{s,p,1,y} = \kappa'\|$, and $\text{VoteShare}_{j,s,y}$ is vote share of candidate j in state contest s in year y.

We report the average $\text{VoteShare}^{\kappa,\kappa'}$ (i.e., the mean over all candidates) and its standard deviation in Table 3, for all different candidate configurations that appear in our data. It is useful to go carefully over these results, as they underpin much of the parametric analysis described in the subsequent sections.

Consider the mean vote shares. If $\kappa' = 0$ (i.e., all κ candidates are in the same position), then the mean share of a candidate is (by definition) $1/\kappa$. Remarkably, it never actually happens that all participants in a primary belong to the same political position, and thus these configurations are not listed in Table 3. If $\kappa = \kappa'$, then (again by definition) the mean share of each candidate is equal to $1/(\kappa + \kappa')$. All other reported values are the realized averages in the data.

From Hypothesis 1 in the theoretical part, we have the following expectations: First, a reduction in the number of candidates in the same position increases the average vote share of the remaining candidates in that position. Formally, $\text{VoteShare}^{\kappa-1,\kappa'} > \text{VoteShare}^{\kappa,\kappa'}$. Second, there is partial, but not complete "crowding out" among candidates in the same position (unless $\lambda \to \infty$). Formally, $\kappa \cdot \text{VoteShare}^{\kappa,\kappa'} > (\kappa-1) \cdot \text{VoteShare}^{\kappa-1,\kappa'}$, i.e. a reduction in the number of candidates in the same position decreases the total vote share of the candidates in that position (in other words, there are some cross-over voters who change to a candidate in the opposite position).

By-and-large, the data are consistent with these expectations. For example, when going from three candidates in a 2-1 constellation to two candidates in a 1-1 constellation (in both cases, the only constellations in the data with 2 and 3 candidates), the vote share of the candidate in the previously crowded position increases from 29.6% to 50.0%, while the vote share of the competitor (who is already alone in his position in the initial situation) increases only from 40.8% to 50.0%. Similarly, going from a 3-2 constellation to a 2-2 constellation increases the average vote share of one of the initially more crowded candidates from 19.0% to 25.0%, while it increases the average share of the two initially less crowded candidates only from 21.5% to 25.0%.[14]

Holding the total number of candidates fixed, the total vote share of all candidates in a specific position is always increasing in the number of candidates in that position. For example, consider all contests involving 5 candidates: Here, $4 \times 17.9\% = 71.6\% > 3 \times 19.0\% = 57.0\% > 2 \times 21.5\% = 43.0\% > 28.3\%$. Thus, there is clearly diversion of votes from one candidate to another candidate in the same location, but the more candidates are in a location, the bigger their combined share. The same pattern holds for contests with 3 and 4 candidates, as can be easily verified from the figures reported in Table 3. [15]

The only case that contradicts both predictions of Hypothesis 1 is going from a 4-1 constellation to a 3-1 constellation, in which case the average vote share of a candidate in the crowded position decreases from 17.9% to 17.3%. This is probably due to the small number of cases (there were only two state elections with a 4-1 constellation, and six with a 3-1 constellation) and the absence of any controls. In particular, the lonely candidate in a 3-1 constellation is doing surprisingly well, getting on average 48 percent of the vote. This phenomenon is also responsible for the fact that going from 3-1 to 2-1 reduces the vote share of the lonely candidate from 48.2% to 40.8%. The largest number of observations, and therefore the highest level of confidence in the results, obtains for the case of two and three candidates.

To summarize, the results in Table 3 are indicative of the validity of Hypothesis 1. Vote shares decline with the number of candidates who share a location, holding the total number of candidates constant. Moreover, the combined vote shares of candidates in a location increases with the number of candidates in that location, holding the total number of candidates constant.

Not distinguishing the election sequencing does not lead to any biases for the questions we address with this analysis. Treating political parties and positions as fungible does not create

[14] Of course, $\kappa \cdot \text{VoteShare}^{\kappa,\kappa'} + \kappa' \cdot \text{VoteShare}^{\kappa',\kappa} = 100$ holds as an identity. Deviations from this in Table 3, such as here where $3 \times 17.3\% + 48.2\% = 100.1\%$, are due to rounding.

[15] The precise implications of the theory are for expected vote share comparisons between κ candidates in one position and κ' candidates in the other, versus $\kappa - 1$ in one position and κ' in the other. But comparisons between κ candidates in one position and κ' in the other versus $\kappa - 1$ in one position and $\kappa' + 1$ in the other can be obtained by applying our theoretical result iteritatively.

any biases, provided that the political locations do not differ systematically in voter popularity. Subsequent analysis, described in the next section, suggests that this is indeed the case. As it will become clear below, even if locations were to differ systematically in voter popularity, no biases would result provided that there is no systematic difference across political positions in the number of candidates in that position. Though this is essentially true for the Democrats, it is not true for the Republicans (there are typically fewer moderates than conservatives). But given that political positions do not differ much in popularity among the voters, any differences in their "popularity" among politicians would not impact the validity of our results. Overall, the main value of the analysis described here is the absence of any parametric or modeling assumptions, except for those qualitative properties listed in this paragraph.

Since we do not use information about the sequence of elections in the reduced form approach, it cannot provide any evidence that bears on the possibility of voter learning about candidate abilities or valence (i.e., Hypothesis 2). Neither can we assess the predictions of Hypothesis 3, since we do not use any proxies for the political leanings of the electorate in different states. We address these questions in the next two sections through the use of formal econometric specifications.

6.2 Econometric Analysis of Vote Shares

We now investigate the degree to which candidate vote shares depend on the field of competing candidates, their political position, and a proxy for each state's preference distribution. We do not impose the structural assumptions of the theoretical model, but rather adopt a reduced form approach, using progressively more flexible specifications. Thus, we are able to investigate whether voting shares exhibit patterns that are consistent with the theoretical framework and measure the salience of the phenomena the theory describes, without resting the foundations of our tests and measurements on the model itself. The primary benefit of our approach is that it remains valid even if the model itself is somewhat misspecified, and that it allows us to test separately each of the model's predictions rather than test the model in its entirety. It is also analytically and computationally convenient, as we do not need to solve a general multiperiod model for each of the US Presidential Primary competitions that would be required for a fully structural approach.

Our simplest specification estimates the equation

$$\text{VoteShare}_{j,s,y} = \alpha + \beta_1 \text{CanDif}_{j,s,y} + \beta_2 \text{CanOwn}_{j,s,y} + \epsilon_{j,s,y} \quad (17)$$

where $\text{VoteShare}_{j,s,y}$ is the adjusted vote share of candidate j in state s and year y (measured on a 0–100 scale), and $\text{CanOwn}_{j,s,y}$ and $\text{CanDif}_{j,s,y}$ is the number of candidate j's competitors with the same or opposite political location, respectively, in state election s in year y. This specification essentially parallels the nonparametric approach summarized in Table 3 and discussed above, but it uses a statistical framework and thus provides the average effect of adding another candidate of the same or a different political position and the associated standard errors. The findings, reported under Model 1 in Table 4, show that an additional candidate in the same political location as candidate j reduces candidate j's vote share by three times as much as an additional candidate in the opposite location. The difference between the two coefficients is also strongly statistically significant. Throughout, we report White's heteroskedasticity consistent standard errors and use them for all tests of significance (see White (1980)).

We next investigate whether this result is affected by the relative popularity of candidates of different political positions. As reported under Model 2, this is not the case. In fact, once we control for the number of candidates in each political location, the residual vote shares of candidates appear not to be correlated with their political position, either for the Democratic or for the Republican Party. We let Moderate_j and Outsider_j be dummy variables that take the value 1 if candidate j is a moderate Republican or Democratic "outsider" candidate, respectively, and 0 otherwise (see Section 5 for a discussion of these terms). In the estimation of the equation

$$\text{VoteShare}_{j,s,y} = \alpha + \beta_1 \text{CanDif}_{j,s,y} + \beta_2 \text{CanOwn}_{j,s,y} + \gamma_1 \text{Moderate}_j + \gamma_2 \text{Outsider}_j + \epsilon_{j,s,y} \quad (18)$$

the coefficients of γ_1 and γ_2 are both not statistically significant and small in numerical value (a difference of approximately 3 percent for either party). This does not imply that the average combined vote share of candidates in each political location does not differ. In fact, the average combined vote share of Moderate Republicans is approximately 43% while the average combined vote share of Conservative Republicans is 57%. However, Model 2 attributes this substantial difference to the fact that there are more Conservative than Moderate Republican candidates. With Conservative Republicans diverting a disproportionate fraction votes from each other rather than from Moderate Republicans, there is no residual advantage to being a Conservative candidate.[16] We should also note that the coefficient of γ_1 is basically a McCain effect, as only one other Republican candidate (Guiliani) is labeled as Moderate, and he only ran in a small number of states. To a first approximation, then, the way to interpret this coefficient is that McCain received higher percentage vote shares than the typical Republican candidate not because Moderate candidates are more popular among Republican primary voters, but because he is was by and large the only Moderate candidate running for the Republican nomination. Moreover, even this may overstate the popularity of the Moderate political position for Republicans, as the coefficient of γ_1 may be conflated with the quality (or valence) of McCain relative to other candidates. We therefore exclude the political location variables from subsequent analysis, except in the Model 4 where by they enter in interaction form (and are thus also included in levels).

In the next regression (Model 3), we investigate whether the relevance of candidate political location in vote share diversion is confined to one of the two major parties, or is present in both. We do so by estimating the regression

$$\begin{aligned} \text{VoteShare}_{j,s,y} \quad = \quad & \alpha + \beta_{1R}\text{CanDif}_{j,s,y}\text{Rep}_j + \beta_{2R}\text{CanOwn}_{j,s,y}\text{Rep}_j \\ & +\beta_{1D}\text{CanDif}_{j,s,y}\text{Dem}_j + \beta_{2D}\text{CanOwn}_{j,s,y}\text{Dem}_j + \epsilon_{j,s,y} \end{aligned} \tag{19}$$

where the variable Rep_j takes the value of one if candidate j is a Republican and zero otherwise, and the variable Dem_j takes the value of one if candidate j is a Democrat and zero otherwise. In this model, the parameters β_1 and β_2 are estimated for each party separately. The results, reported in Table 4, suggest that voter segmentation across political locations might be more pronounced for the Democratic Party, where a candidate's vote share is only negligibly affected by competing against one fewer candidate in the opposing political location, but is very strongly affected by one fewer candidate in the same political location. The relative effect of the location of competing candidates is also statistically significant for the Republican primaries, but much smaller in quantitative terms.[17]

[16]The average combined share of the two political locations for Democratic primaries is much closer (47.5% versus 52.5%) and similar to the difference implied by the regression coefficient γ_2. This is due to the fact that the Democratic candidates are more evenly distributed between the two political positions.

[17]Adding location dummies to this regression does not materially affect the estimates for the Democratic coefficients, but renders the Republican location effect to essentially zero. However, with Moderate_i being

We should note that Model 3, in attempting to identify party-specific candidate competition and substitutability effects, carries some limitations. In particular, the relative vote diversion effect for the Republican Party (i.e., the difference between β_{2R} and β_{1R} is largely based on the comparisons of differences in the vote shares of McCain across various candidate configurations as compared to differences in the vote shares of other Republican candidates across various candidate configurations.[18]

The next two regressions (Models 4 and 5) are not aimed at directly estimating the vote diversion effects but rather at evaluating whether our political location measures do indeed correspond plausibly to voter preferences. Because the winner of each party's primary was that party's candidate in the general election, we do not use the outcome of the 2000, 2004, or 2008 presidential elections as a proxy for the distribution of political preferences in a state. Instead, our proxy is the outcome of the 1992 presidential election between Bush, Clinton and Perot. Voter preferences in states in which Clinton did well are plausibly shifted to the left relative to the rest of the country, and we would therefore also expect that Moderate Republicans do better in these states than Conservatives. Similarly, states in which Perot did well likely have a a larger than average share of populist voters, so that we expect that candidates classified as Outsiders do better in the Democratic party.

The estimation equation of Model 4 is given by

$$\begin{aligned}\text{VoteShare}_{j,s,y} \;=\; & \alpha + \beta_1 \text{CanDif}_{j,s,y} + \beta_2 \text{CanOwn}_{j,s,y} + \gamma_1 \text{Moderate}_j + \gamma_2 \text{Outsider}_j \\ & + \gamma_{1C} \text{Moderate}_j Clinton92\%_s + \gamma_{2P} \text{Outsider}_j Perot92\%_s + \epsilon_{j,s,y} \qquad (20)\end{aligned}$$

and the one of the much more flexible Model 5 by

$$\text{VoteShare}_{j,s,y} \;=\; \alpha_{j,t,y} + \gamma_{1C} \text{Moderate}_j Clinton92\%_s + \gamma_{2P} \text{Outsider}_j Perot92\%_s + \epsilon_{j,s,y} \qquad (21)$$

essentially a dummy for McCain, the Republican effect would be identified solely from the gain of voters by McCain as other candidates (of opposing location) depart, relative to the gain of voters by his opponents as other candidates (of same location) depart. Not only the effective information for this specification is even more limited (only 4 such withdrawals) but with McCain being a higher quality candidate, the location and valence effects are confounded (McCain gets a bigger than expected share of the departing candidates' voters because he is a better candidate in the vertical dimension).

[18]This is because McCain has always been in the race for all of our observations, and he shared his political location with another candidate for only six states. Therefore, different Republican configurations largely differ from each other in the number of conservative candidates who have withdrawn. Thus, β_{1R} is obtained by differences in McCain's vote share across different candidate configurations, and β_{2R} is obtained by differences in candidates other than McCain across different candidate configurations.

where $Perot92\%_s$ and $Clinton92\%_s$ are Perot's and Clinton's vote share in state s in the 1992 Presidential election, respectively,[19] and $\alpha_{j,t,y}$ are candidate-year-round effects, i.e., coefficients on a set of dummies that take the value of 1 for a particular candidate for all state elections taking place on a particular day (round) in a given year, and zero otherwise. These dummies would perfectly predict the share of a candidate for election days in which only a single state votes, completely eliminating their influence on the remaining model parameters. Thus, we drop observations that consist of a single state contest from the regression in Model 5, reducing the number of observations from 415 to 328. The more flexible specification of Model 5 allows us to test the vote shifting effect across political positions without relying on any parametric assumptions on substitutability between candidates and controlling for any other variables that vary across election rounds (including perceived candidate valence).

As explained above, the expected Clinton effect is an increase in the vote share of moderate Republicans. This appears indeed to be the case, as the coefficients γ_{1C} are positive and statistically significant for both models. Each percentage point won by Clinton in 1992 translates into approximately a 0.9 percentage point gain for Moderate Republican candidates.

The Perot effect on Democratic outsider candidates is (marginally) significant only in Model 5 (i.e., the point estimate of γ_{2P} is positive), while it is essentially zero for Model 4. The fact that the evidence is not as strong as for the Republicans may be a consequence of most of Perot's voters being conservative populists, so that the influence on the Democratic primary electorate is rather weak.

The final set of regressions, Models 6 and 7, are intended to test Hypothesis 3, which asserts that weak candidates of an ideological position are likely benefit disproportionately relative to strong candidates of the same ideological position from a tilt in voter preference towards that ideological position. This hypothesis is harder to test because it demands much from our limited data (we identify this effect from a differential impact of our voter preference proxies on candidates of the same ideology), and also because it requires some operative measure of candidate valence.[20] It is important to recall that valence, as perceived by the voters, is not constant throughout the sequence of elections, but rather changes from round to round,

[19]The vote share variables $Clinton92\%_s$ and $Perot92\%_s$, like $\text{VoteShare}_{j,s,y}$ range from 0 to 100.

[20]The theory suggests that not only a candidate's valence, but the number of candidates in each position and their valence are important in determining the effect of voter preferences on voting shares.

suggesting that any estimation approach should be based on variants of Model 5.

We adopt as our proxy for valence in round t the vote average share of a candidate in that round, $MeanShr_{j,t,y}$. Clearly this is an imperfect measure, but a reasonable one. Candidates with high relative valence, as perceived in round t, will have higher values of $MeanShr_{j,t,y}$. The number and distribution of competing candidates will also affect the values of $MeanShr_{j,t,y}$. To reduce these candidate composition effects on this measure, we always employee this measure in regressions that include candidate-round effects, as in Model 5.[21] Moreover, averaging vote shares of all contests in a round is meaningful because all states have the same ex ante expectations about valence which they update independently on the basis of their privately observed signal. In addition, the set of candidates is the same in all such contests.[22]

One specification consistent with the use of this proxy (denoted by Model 6) is:

$$\begin{aligned} \text{VoteShare}_{j,s,y} &= \alpha_{j,t,y} + \gamma_{1C}\text{Moderate}_j\text{Clinton92\%}_s + \gamma_{2P}\text{Outsider}_j Perot92\%_s \\ &+ \{\gamma_{1Cs1}\text{Moderate}_j + \gamma_{1Cs0}[1 - \text{Moderate}_j]\}\,\text{Clinton92\%}_s MeanShr_{j,t,y} \\ &+ \{\gamma_{1Ps1}\text{Outsider}_j + \gamma_{1Ps0}[1 - \text{Outsider}_j]\}\,\text{Perot92\%}_s MeanShr_{j,t,y} + \epsilon_{j,s,y} \end{aligned} \tag{22}$$

This specification suffers from the short-coming that a higher than expected vote share in a particular state would lead to a higher value of $MeanShr_{j,t,y}$. Such a correlation would lead to a bias in the regression coefficients (albeit not a large one when many states are holding their primary in the same time).

A second specification that does not suffer from this short-coming (denoted by Model 7) is:

$$\begin{aligned} \text{VoteShare}_{j,s,y} &= \alpha_{j,t,y} + \gamma_{1C}\text{Moderate}_j\text{Clinton92\%}_s + \gamma_{2P}\text{Outsider}_j Perot92\%_s \\ &+ \{\gamma_{1Cs1}\text{Moderate}_j + \gamma_{1Cs0}[1 - \text{Moderate}_j]\}\,\text{Clinton92\%}_s MeanShr_{j,t-s,y} \\ &+ \{\gamma_{1Ps1}\text{Outsider}_j + \gamma_{1Ps0}[1 - \text{Outsider}_j]\}\,\text{Perot92\%}_s MeanShr_{j,t-s,y} + \epsilon_{j,s,y} \end{aligned} \tag{23}$$

[21] The theoretical model shows that the effect of changes in electorate preferecens on candidate shares depends not only on the candidate's valence and political position but also on the number of competing candidates, their valence, and their political position. The variable $MeanShr_{j,t,y}$ also adjusts for the number of competing candidates, their valence and political position, and thus in a qualitative way reflects the factors that enter in the comparative static developed by the theoretical model.

[22] Averaging vote shares across rounds or using average shares in previous rounds would not be meaningful because the set of candidates is not, in general, the same as that of the current round.

where $MeanShr_{j,t-s,y}$ is the average vote share of candidate j in the contests taking place in round t in year y, *excluding* the contest in state s.[23]

We discuss the results of both of these regressions together. For the Republican candidates, the two specifications yield the same conclusions, though standard errors are substantially lower in Model 7 (and thus statistical significance is substantially higher). Moderate Republicans do better than Conservative Republicans in states with strong support for Clinton in 1992, and this difference is broadly independent of their vote share (the coefficients γ_{1Cs1} and γ_{1Cs0} are similar in size). Consistent with Hypothesis 3, high valence (high average vote share) Conservative Republican candidates are hurt proportionately less than low valence Conservative Republicans in states with strong Clinton support.[24] Somewhat surprisingly, high valence Moderates not only benefit proportionately less than low valence Moderates in states with strong Clinton support, but also benefit less in absolute terms.[25] Thus, for Republican candidates, the data support the theoretical prediction that relative weak candidates are more sensitive (in relative terms) to shifts in electorate preferences, and in fact in one instance they appear to be more sensitive even in absolute terms.

For the Democratic candidates, the two models differ somewhat in their results. Model 6 suggests that Perot's strength in 1992 is associated with essentially proportional effects on candidates for different average voting shares: the Outsiders do better in states which voted heavily for Perot and the Insiders do better in states that did not, but weak candidates do not get a proportionately bigger boost than strong candidates. Model 7, however, does provide strong support for the theoretical predictions. In fact, as in the case with the Republican candidates, the data indicate that not only the relative weak candidates are more sensitive (in relative terms) to shifts in electorate preferences, but that in the case of Outsiders they appear to be more sensitive even in absolute terms.

The results of the last set of models are also useful in our analysis of trends in vote share variance as the primary election progresses. Trends in this variance would be indicative of voter learning about candidates, as discussed in the theoretical section of the paper. We discuss this

[23]Note that, by necessity, this specification uses a different proxy for every state, since the variable $MeanShr$ no longer takes the same value for all states in a given round.

[24]The effect for Conservative is equal to $-\gamma_{1C} - \gamma_{1Cs0} MeanShr_{j,t-s,y}$.

[25]The coefficient γ_{1Cs1} is negative. Since mean vote shares for Moderates are rarely above sixty percent, the combined effect $\gamma_{1C} + \gamma_{1Cs1} * MeanShr_{j,t-s,y}$ is positive almost everywhere.

in more detail in the next section.

6.3 Econometric Analysis of Share Variability

The theoretical model posits that voters initially have less precise information about candidate quality (valence), and that their perceptions of candidate quality stabilizes over the course of the primary election contests. Even with complete information about candidate attributes, the vote shares of candidates would vary across states because voter preferences for positions differ. Uncertainty about candidate quality that is slowly resolved over time provides an additional component of vote share variability. As voters learn about about candidates, additional information moves perceptions (and thus vote shares) by a progressively smaller amount. Unless voter preferences in states voting late differ among them by more than voter preferences in early voting states, it follows that vote share variability declines in later state contests.

Estimates of vote share variability necessarily have to be based on the analysis of the residuals of equations of the form estimated in the preceding section. However, care must be taken so that the greatest proportion of systematic variation in vote shares be removed, without removing any component of the residuals that would help identify learning effects or introducing any biases in the estimation of such effects. Note that, from the point of estimating the reduction in variability due to learning, all parameters associated with systematic differences in the expected vote shares are nuisance parameters: We do not care about their values here, except that they are accounted for as best as possible. Our base model to obtain the residuals has an exhaustive set of candidate-round-year dummies. The residuals indicate whether a candidate did better or worse in a state relative how he did in other states that voted on the same date. It controls for the very identity of competing candidates (rather than merely their political position and number) in the most flexible way: with indicator variables whose coefficients vary (with no parametric constraints) over time. This regression is equivalent to Models 5, 6, and 7 without the Clinton and Perot effects, does not rely on our classification of candidates into political locations or any of the other aspects of our modeling that involve the competition between candidates of different political positions.

We also estimate vote share variability using the residuals of the more heavily parameterized Models 5, 6, and 7. By their very nature, the results here would differ somewhat for each specific parametrization of the Clinton and Perot effects. Given that our focus here is in the time variation of the residuals, we report as a representative model the results based on Model 6, which is one the two most flexible specifications and uses the same valence proxy variable for all elections in a given round.[26]

Let $NumCand_{j,s,y}$ be the number of candidates contesting the state s in year y for the party of candidate j, and let $PriorSignals_{j,s,y}$ be the number of state contests for the party of candidate j prior to the contest in state s in year y. We estimate the regressions

$$|\hat{\epsilon}_{j,s,y}| = a + b\ NumCand_{j,s,y} + c\ PriorSignals_{j,s,y} + u_{j,s,y} \tag{24}$$

and

$$|\hat{\epsilon}_{j,s,y}| = a + b\ NumCand_{j,s,y} + c\ PriorSignals_{j,s,y} + d\ PriorSignals^2_{j,s,y} + u_{j,s,y} \tag{25}$$

where $|\hat{\epsilon}_{j,s,y}|$ is the residual from either Model 6, or from Model 5/6/7 without the Perot and Clinton interaction terms. The number of candidates is included as a variable in the regression because a higher number of candidates means smaller vote shares (on average), and smaller vote shares exhibit smaller variances. We also re-estimate regressions (24) and (25) making a small sample adjustment for residuals that controls for the fact that OLS residuals are a biased estimate of disturbance variance when computed from small samples. In particular, we use $\left(\frac{m_{j,s,y}}{m_{j,s,y}-1}\right)^{0.5} |\hat{\epsilon}_{j,s,y}|$ as the dependent variable, where $m_{j,s,y}$ is the number of candidates in the party of candidate j for state s in year y.[27]

This yields a total of eight regressions, whose results are reported in in the first 8 columns of Table 5. Consistent with Hypothesis 2, residual variance is decreasing with the number of prior contests for all specifications (an effect that is statistically significant in all regressions at the 10 percent level (in 6 out of 8 regressions at the 5 percent level). Moreover, we would expect that learning about candidate quality reduces variability more in the early than the late contests: in the early rounds, voters have weaker priors about candidates and information can move their opinions substantially; in the later rounds, voters have stronger priors and the same amount of information can only move their opinion marginally. Consistent with this, we find that the coefficient of $PriorSignals^2_{j,s,y}$ is positive and statistically significant in all specifications. In

[26]The residuals of the other two models give similar results (except that when using the residuals of Model 7, the quadratic term in (25) is no longer statistically significant, though it remains of the same sign).

[27]This adjustment is exact when no covariates are used.

fact, the marginal effect of more contests approaches zero towards the end of the primaries: it appears that there is no further reduction in voter uncertainty about candidates towards the end of the typical primary run. Finally, the number of candidates has a negative effect on variance, as expected (though the effect is often not statistically significant when we adjust the dependent variable for the number of candidates).

It is worth noting that, even though this variability reduction effect that we attribute to firmer beliefs due to learning is statistically significant and exhibits the expected diminishing pattern, it is quantitatively small relative to other factors: it tends to explain only about 3 to 5 percent of the residual variance. Evidently, there are several other determinants of vote share variability, including the type of information shocks that lead to learning about candidate valence in the first place, and possible co-ordination of voters across states voting simultaneously.[28]

The second of these two possibilities is of special concern, because it could lead to a systematic relationship between variance and number of signals or rounds. Suppose that voters in early states can coordinate on a candidate of a particular political position (perhaps through social interaction or coverage in the local press) but cannot coordinate across states. In this scenario, a candidate may obtain many votes in one state (if voters coordinate on him) but very few on another state that votes at the same round (if voters there coordinate on his opponent). Thus, candidate share variability would be relatively high ex ante, after controlling for mean candidate vote shares. In later rounds, coordination across states increases, as voters observe who is likely to emerge as the most competitive candidate in a particular political position. This effect would lead to a reduction in share variability, even in the absence of any firming of priors about quality, based only on coordination across states.

Note that, if this effect were indeed the driving force behind the reduction of share variability, then vote share residuals for candidates in the same political position should be strongly negatively correlated and largely cancel out. Vote share variability at the political position level, controlling for candidate mean shares, should not have a clear trend over time. We test for this possibility by summing the vote share residuals of candidates in the same political position in a particular state contest. We then perform the same analysis described in equa-

[28] It is not surprising, and in fact reassuring, that when one includes the variable $PriorSignals_{j,s,y}$ in the vote share regressions it comes out uniformly insignificant.

tions (24) and (25) using the (aggregated) residuals of Model 6. Note that the right-hand side variables take the same values for candidates competing in the same state contest, so that these regressions only differ in the construction of the dependent variable (and of course in the number of observations). The estimates are reported in the last 4 columns of Table 5. The pattern of coefficient estimates is unchanged: share variability, measured at the position level, declines for later contests. Statistical significance is affected when both the number of signals and the number of signals squared are used as regressors; however, the two variables remain jointly statistically significant. We conclude that increased coordination of voters across states voting contemporaneously is not an explanation for the reduction of share variability.

There is also another observation that supports our interpretation that the reduction in variability is due to hardening priors as more information about the candidates becomes available. If one were to use a simple counter of the election round in (24) and (25), i.e., a variable that is akin to a trend and does not take into consideration the number of states that compete in a round, the coefficients on that variable would generally not be statistically significant. Thus, it is not just the passage of time that is associated with reduced variability, but the number of states that voted previously.

7 Concluding Remarks

In this paper, we have developed a model of sequential primaries featuring coordination problems for voters as well as voter learning about the valences of politically differentiated candidates. A substantial problem in our framework is that candidates whose political philosophy is very similar "steal" votes mainly from each other, so that the candidate who ends up with a plurality of votes is not necessarily preferred by a majority of the electorate to all of his competitors. This vote-splitting effect presents a substantial problem for the efficiency of any voting system when more than two candidates run in an election, because a weaker candidate (i.e., not the Condorcet winner) might win in a situation where the Condorcet winner is splitting votes with a close ideological neighbor. However, the U.S. presidential primary system provides a unique opportunity to gauge the presence and size of this vote-splitting effect, because some candidates drop out during the primaries, and the voters that would have voted for a dropped-out candidate need to choose which of the remaining candidates to support.

We derive several predictions from the theoretical model. First, if a candidate drops out, this benefits the remaining candidates who shared the drop-out's position more than it benefits candidates in the opposite position. Second, voter learning over time, facilitated through observation of previous election results, leads to a reduction over time of the variance of a candidate's vote share, and does so at a progressively slower rate as uncertainty about candidates gets resolved. Third, an increase of the share of voters who prefer a particular political position leads to a higher increase in the *absolute* number of votes for a strong candidate rather than a weak candidate in that position, but *relatively*, weak candidates benefit more than strong ones. These hypotheses are broadly supported in a dataset that contains the election results from the five contested U.S. presidential primaries in 2000, 2004 and 2008. Moreover, the first of these three effects is quantitatively very important.

8 Appendix

Derivation of (10) **and** (11)**.** To calculate the number of voters who rank candidate j highest and candidate j' second, consider first the case that both candidates j and j' have the same position, say, $a = 0$ (i.e., $j, j' \in J_0$). A voter of type θ ranks j highest and j' second if and only if

$$\hat{v}^s_j + \varepsilon_j - \lambda d(j,\theta) \geq \hat{v}^s_{j'} + \varepsilon_{j'} - \lambda d(j',\theta) \geq \max_{k \neq j,j'} (\hat{v}^s_k + \varepsilon_k - \lambda d(k,\theta)). \tag{26}$$

Consider first the second inequality (i.e., the one that secures that j' is preferred to every candidate except j). For a given $\varepsilon_{j'}$, the second inequality in (26) is satisfied if and only if

$$\varepsilon_k < \hat{v}^s_{j'} - \hat{v}^s_k + \varepsilon_{j'} - \lambda[d(j',\theta) - d(k,\theta)] \text{ for all } k \neq j, j' \ . \tag{27}$$

Since the ε_k's are distributed independently, the probability that a voter of type $\theta = 0$ ranks candidate j' higher than any other candidate (except j) is

$$\prod_{J^s_0 \backslash \{j,j'\}} \Phi\left(\frac{\hat{v}^s_{j'} - \hat{v}^s_k + \varepsilon_{j'}}{\sigma_\varepsilon}\right) \cdot \prod_{J^s_1} \Phi\left(\frac{\lambda + \varepsilon_{j'} + \hat{v}^s_{j'} - \hat{v}^s_k}{\sigma_\varepsilon}\right). \tag{28}$$

Turning to the first inequality in (26), it must also be true that $\varepsilon_j \geq \varepsilon_{j'} + \hat{v}^s_{j'} - \hat{v}^s_j$, which, for given $\varepsilon_{j'}$, has probability $\left[1 - \Phi\left(\frac{\varepsilon_{j'} + \hat{v}^s_{j'} - \hat{v}^s_j}{\sigma_\varepsilon}\right)\right] = \Phi\left(\frac{\hat{v}^s_j - \hat{v}^s_{j'} - \varepsilon_{j'}}{\sigma_\varepsilon}\right)$, where the equality uses the identity $1 - \Phi(x) = \Phi(-x)$ for the cdf of the normal distribution.

Integrating over the possible realizations of $\varepsilon_{j'}$ gives that the proportion of type 0 voters who rank candidate j highest and candidate j' second, is

$$\int_{-\infty}^{\infty} \Phi\left(\frac{\hat{v}^s_j - \hat{v}^s_{j'} - \varepsilon_{j'}}{\sigma_\varepsilon}\right) \prod_{J^s_0 \backslash \{j,j'\}} \Phi\left(\frac{\hat{v}^s_{j'} - \hat{v}^s_k + \varepsilon_{j'}}{\sigma_\varepsilon}\right) \cdot \prod_{J^s_1} \Phi\left(\frac{\lambda + \varepsilon_{j'} + \hat{v}^s_{j'} - \hat{v}^s_k}{\sigma_\varepsilon}\right) \cdot \phi_\varepsilon(\varepsilon_{j'}) d\varepsilon_{j'}. \tag{29}$$

Similarly, the share of type 1 voters who rank candidate j highest and candidate j' second, is

$$\int_{-\infty}^{\infty} \Phi\left(\frac{\varepsilon_{j'} + \hat{v}^s_j - \hat{v}^s_{j'}}{\sigma_\varepsilon}\right) \prod_{J^s_0 \backslash \{j,j'\}} \Phi\left(\frac{\hat{v}^s_{j'} - \hat{v}^s_k + \varepsilon_{j'}}{\sigma_\varepsilon}\right) \cdot \prod_{J^s_1} \Phi\left(\frac{-\lambda + \varepsilon_{j'} + \hat{v}^s_{j'} - \hat{v}^s_k}{\sigma_\varepsilon}\right) \cdot \phi_\varepsilon(\varepsilon_{j'}) d\varepsilon_{j'}. \tag{30}$$

The total proportion of voters who rank candidate j highest and candidate j' second (where both $j, j' \in J_0$) is then $R_{00}(j, j') =$

$$(1-\mu^s)\int_{-\infty}^{\infty}\Phi\left(\frac{\hat{v}_j^s-\hat{v}_{j'}^s-\varepsilon_{j'}}{\sigma_\varepsilon}\right)\prod_{J_0^s\setminus\{j,j'\}}\Phi\left(\frac{\hat{v}_{j'}^s-\hat{v}_k^s+\varepsilon_{j'}}{\sigma_\varepsilon}\right)\cdot\prod_{J_1^s}\Phi\left(\frac{\varepsilon_{j'}+\hat{v}_{j'}^s-\hat{v}_k^s+\lambda}{\sigma_\varepsilon}\right)\cdot\phi_\varepsilon(\varepsilon_{j'})d\varepsilon_{j'}+$$
$$\mu^s\int_{-\infty}^{\infty}\Phi\left(\frac{\hat{v}_j^s-\hat{v}_{j'}^s-\varepsilon_{j'}}{\sigma_\varepsilon}\right)\prod_{J_0^s\setminus\{j,j'\}}\Phi\left(\frac{\hat{v}_{j'}^s-\hat{v}_k^s+\varepsilon_{j'}}{\sigma_\varepsilon}\right)\cdot\prod_{J_1^s}\Phi\left(\frac{\varepsilon_{j'}+\hat{v}_{j'}^s-\hat{v}_k^s-\lambda}{\sigma_\varepsilon}\right)\cdot\phi_\varepsilon(\varepsilon_{j'})d\varepsilon_{j'} \quad (31)$$

We now turn to the case that the position of candidate j' is $a = 1$. Proceeding as above, with the necessary adjustments, one can show that the total proportion of voters who rank candidate j highest and candidate j' second (where $j \in J_0^s$ and $j' \in J_1^s$) is then $R_{01}(j, j') =$

$$(1-\mu^s)\int_{-\infty}^{\infty}\Phi\left(\frac{\hat{v}_j^s-\hat{v}_{j'}^s+\lambda-\varepsilon_{j'}}{\sigma_\varepsilon}\right)\prod_{J_0^s\setminus\{j,j'\}}\Phi\left(\frac{\hat{v}_{j'}^s-\hat{v}_k^s+\varepsilon_{j'}-\lambda}{\sigma_\varepsilon}\right)\cdot\prod_{J_1^s}\Phi\left(\frac{\varepsilon_{j'}+\hat{v}_{j'}^s-\hat{v}_k^s}{\sigma_\varepsilon}\right)\cdot\phi_\varepsilon(\varepsilon_{j'})d\varepsilon_{j'}+$$
$$\mu^s\int_{-\infty}^{\infty}\Phi\left(\frac{\hat{v}_j^s-\hat{v}_{j'}^s-\lambda-\varepsilon_{j'}}{\sigma_\varepsilon}\right)\prod_{J_0^s\setminus\{j,j'\}}\Phi\left(\frac{\hat{v}_{j'}^s-\hat{v}_k^s+\varepsilon_{j'}+\lambda}{\sigma_\varepsilon}\right)\cdot\prod_{J_1^s}\Phi\left(\frac{\varepsilon_{j'}+\hat{v}_{j'}^s-\hat{v}_k^s}{\sigma_\varepsilon}\right)\cdot\phi_\varepsilon(\varepsilon_{j'})d\varepsilon_{j'}. \quad (32)$$

Analogous conditions to (31) and (32) can be derived for candidate j being located at $a = 1$.

Equations (10) and (11) are special cases of (31) and (32). ■

Proof of Proposition 1. Existence follows by construction: Since the vector W^r is generated using the realized vector of estimated valences $(\hat{v}_j^r)_{j=1,\dots,k}$, a solution to (14) exists. Furthermore, it is clear that any vector of the form $(0, x_2, x_3, \dots, x_k) + (c, c, \dots, c)$ also satisfies (14). It remains to be shown that there cannot be a solution of the form $(0, y_2, y_3, \dots, y_k)$ with $(0, y_2, y_3, \dots, y_k) \neq (0, x_2, x_3, \dots, x_k)$. 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 (14) shows that candidate $\bar{k}$ receives a strictly higher vote share than $W_{\bar{k}}^r$, 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 (14) shows that candidate $\underline{k}$ receives a strictly smaller vote share than $W_{\underline{k}}^r$, a contradiction. But then, it must be true that $y_j = x_j$ for all $j = 2, \dots, k$. ■

Proof of Proposition 2. Each state voting at time t has a different estimate of candidate j's valence. The average ex-post valence of candidate j in those states that vote at the first election date is

$$E\left(\frac{\sigma_v^2}{\sigma_v^2+\sigma_\eta^2}Z_j\right) = \frac{\sigma_v^2}{\sigma_v^2+\sigma_\eta^2}v_j \quad (33)$$

and the variance of this ex-post estimate across these states is

$$\left(\frac{\sigma_v^2}{\sigma_v^2+\sigma_\eta^2}\right)^2 Var(Z_j) = \left(\frac{\sigma_v^2}{\sigma_v^2+\sigma_\eta^2}\right)^2 \sigma_\eta^2 \tag{34}$$

Now consider the average estimated valence of candidate j in those states that vote simultaneously at some later date t, and its variance. Suppose there are R earlier elections, indexed by r. The sum of state-specific signals for candidate j is distributed $N(Rv_j, R\sigma_\eta^2)$, so that the average state-specific signal is distributed $N(v_j, \sigma_\eta^2/R)$. The ex-ante estimate in late states (i.e., before the state-specific signal is observed) is therefore

$$\hat{v}_{j0,late} = \frac{\frac{\sigma_\eta^2}{R}}{\sigma_v^2+\frac{\sigma_\eta^2}{R}} \cdot 0 + \frac{\sigma_v^2}{\sigma_v^2+\frac{\sigma_\eta^2}{R}} \frac{\sum Z_j^r}{R} \tag{35}$$

with a variance of (using (13))

$$\frac{\sigma_v^2\frac{\sigma_\eta^2}{R}}{\sigma_v^2+\frac{\sigma_\eta^2}{R}}. \tag{36}$$

In addition, each late state receives its own signal Z_j^s of variance σ_η^2. The ex-post estimate of candidate j's valence is therefore

$$\hat{v}_j^s = \frac{\sigma_\eta^2}{\sigma_\eta^2+\frac{\sigma_v^2\frac{\sigma_\eta^2}{R}}{\sigma_v^2+\frac{\sigma_\eta^2}{R}}} \cdot \hat{v}_{j0,late} + \frac{\frac{\sigma_v^2\frac{\sigma_\eta^2}{R}}{\sigma_v^2+\frac{\sigma_\eta^2}{R}}}{\sigma_\eta^2+\frac{\sigma_v^2\frac{\sigma_\eta^2}{R}}{\sigma_v^2+\frac{\sigma_\eta^2}{R}}} \cdot Z_j^s \tag{37}$$

The first term comes from the ex-ante estimate and is the same for all states that vote at time t. These states differ only by their signals Z_j^s, and the variance of the valence estimate in late states (around the mean valence estimate in late states) is therefore

$$\left(\frac{\frac{\sigma_v^2\frac{\sigma_\eta^2}{R}}{\sigma_v^2+\frac{\sigma_\eta^2}{R}}}{\sigma_\eta^2+\frac{\sigma_v^2\frac{\sigma_\eta^2}{R}}{\sigma_v^2+\frac{\sigma_\eta^2}{R}}}\right)^2 \cdot Var(Z_j^s) = \left(\frac{1}{1+\frac{\sigma_\eta^2\left(\sigma_v^2+\frac{\sigma_\eta^2}{R}\right)}{\sigma_v^2\frac{\sigma_\eta^2}{R}}}\right)^2 \cdot Var(Z_j^s) = \left(\frac{1}{1+\frac{\sigma_\eta^2}{\sigma_v^2}+R}\right)^2 \cdot Var(Z_j^s), \tag{38}$$

which is clearly decreasing in the number of states R that voted earlier. Since R is increasing in t, this proves the claim. ■

Proof of Proposition 3. Let VS_i denote the overall vote share of candidate i, and let $VS_{i,j}$ denote candidate i's vote share among voters of type j. Clearly, $VS_i = (1-\mu)VS_{i,0} +$

$\mu VS_{i,1}$. When the share of type 0 voters increases, the relative change of candidate A's vote share is then

$$Z_A = \frac{\frac{dVS_A}{d(1-\mu)}}{VS_A} = \frac{VS_{A,0} - VS_{A,1}}{(1-\mu)VS_{A,0} + \mu VS_{A,1}} \tag{39}$$

and, similarly,

$$Z_B = \frac{\frac{dVS_B}{d(1-\mu)}}{VS_B} = \frac{VS_{B,0} - VS_{B,1}}{(1-\mu)VS_{B,0} + \mu VS_{B,1}} \tag{40}$$

Cross-multiplying and simplifying, we find that $Z_B > Z_A$ if and only if

$$\frac{VS_{A,1}}{VS_{B,1}} > \frac{VS_{A,0}}{VS_{B,0}}. \tag{41}$$

A type 0 voter prefers candidate A to B if and only if $v_A + \varepsilon_A \geq v_B + \varepsilon_B$. If λ is large, so that there are almost no cross-over voters (i.e., type 0 voters who vote for C, or type 1 voters who vote for A or B), then the ratio on the right-hand side of (41) is $\Phi\left(\frac{v_A - v_B}{\sqrt{2}\sigma_\varepsilon}\right)/\Phi\left(\frac{v_B - v_A}{\sqrt{2}\sigma_\varepsilon}\right)$.

Consider now the term on the left-hand side of (41). The probability that a type 1 voter ranks *both* A and B higher than C is exceedingly small for λ large and neglected in the following. A type 1 voter prefers A to C if $v_A + \varepsilon_A - \lambda \geq v_C + \varepsilon_C$, and thus, $VS_{A,1} = \Phi\left(\frac{v_A - v_C - \lambda}{\sqrt{2}\sigma_\varepsilon}\right)$. Similarly, $VS_{B,1} = \Phi\left(\frac{v_B - v_C - \lambda}{\sqrt{2}\sigma_\varepsilon}\right)$. We therefore have

$$\frac{VS_{A,1}}{VS_{B,1}} = \frac{\int_{-\infty}^{\frac{v_A - v_C - \lambda}{\sqrt{2}\sigma_\varepsilon}} \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)dt}{\int_{-\infty}^{\frac{v_B - v_C - \lambda}{\sqrt{2}\sigma_\varepsilon}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}dt} = \frac{\int_{-\infty}^{\frac{v_A - v_C - \lambda}{\sqrt{2}\sigma_\varepsilon}} \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)dt}{\int_{-\infty}^{\frac{v_A - v_C - \lambda}{\sqrt{2}\sigma_\varepsilon}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\left(t - \frac{v_A - v_B}{\sqrt{2}\sigma_\varepsilon}\right)^2}{2}\right)dt} \tag{42}$$

Compare the integrands on the right-hand side. Note that $\left(t - \frac{v_A - v_B}{\sqrt{2}\sigma_\varepsilon}\right)^2 - t^2 = \frac{v_A - v_B}{2\sigma_\varepsilon}(v_A - v_B - 2\sqrt{2\sigma_\varepsilon}t)$ is decreasing in t, and is positive for all $t \leq \frac{v_A - v_C - \lambda}{\sqrt{2}\sigma_\varepsilon}$, provided that λ is sufficiently large (clearly, $\lambda \geq v_A - v_C$ is a sufficient condition for this). Thus, substituting the upper limit of the integral for t, the integrand in the denominator is at most $\exp\left(-\frac{v_A - v_B}{2\sigma_\varepsilon}(v_A - v_B - 2\sqrt{2\sigma_\varepsilon}\frac{v_A - v_C - \lambda}{\sqrt{2}\sigma_\varepsilon})\right)$ times the integrand in the numerator, and thus the same relation holds for the values of the two integrals. Since this factor goes to zero as λ grows, (42) goes to infinity, which proves that (41) holds for λ sufficiently large. This proves the claim. ■

References

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

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

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.

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

Chamberlin, J. and F. Featherston (1986). Selecting a voting system. *The Journal of Politics 48*, 347–369.

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

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., H. Herrera, and M. Polborn (2009). A theory of learning and coordination in the presidential primary system. University of Illinois.

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 (2010). Momentum and social learning in presidential primaries. *Journal of Political Economy 118*, 1110–1150.

Krasa, S. and M. K. Polborn (2007). The binary policy model. University of Illinois.

Leininger, W. (1993). The fatal vote: Berlin versus Bonn. *Finanzarchiv 50*(1), 1–20.

Lijphart, A. and B. Grofman (1984). *Choosing an electoral system: issues and alternatives*. Praeger Publishers.

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

Merrill, S. (1984). A comparison of efficiency of multicandidate electoral systems. *American Journal of Political Science 28*, 23–48.

Merrill, S. (1985). A statistical model for Condorcet efficiency based on simulation under spatial model assumptions. *Public Choice 47*(2), 389–403.

Nurmi, H. (1992). An assessment of voting system simulations. *Public Choice 73*(4), 459–487.

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

Poole, K. and H. Rosenthal (1985). A spatial model for legislative roll call analysis. *American Journal of Political Science 29*(2), 357–384.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica 48*(3), 817–838.