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

February 2016

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 2000-2012 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, either quality-wise or in the sense that the nominee does not represent the majority-preferred position.

We consider a situation in which candidates differ both "horizontally" (i.e., with respect to their policy positions) and "vertically" (i.e., with respect to their quality or valence). For example, Republican primary candidates may be either "moderates" or "conservatives", and each voter has a preference for one of these positions, which, however, is not absolute: If a 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. A problem for voters is that they do not know the candidates' valences but just receive signals about it. In this situation, 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, a weaker candidate might win in a situation where the Condorcet winner is splitting votes with a close ideological neighbor. The sequential 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 the drop-out 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 six contested U.S. Presidential primaries that took place in 2000, 2004, 2008, and 2012[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 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 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. In other words, we show that when the same set of candidates competes in two groups of states holding elections in two different dates, the within group vote share variance is higher in the group that votes first; when the set of candidates in the second group is smaller, we adjust the vote share variance appropriately to make the comparison valid. 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

[1] We do not include George W. Bush's and Barack Obama's renominations in 2004 and 2012 because they had no significant opponent.

for a strong candidate rather than a weak candidate in that position, but *relatively*, weak candidates benefit more than strong ones.

2 Literature review

It has long been recognized that coordination between voters and learning about candidate quality are important issues in presidential primaries. For example, Bartels (1987, pp.13) describes the coordination process of those Democratic voters unhappy with the establishment candidate in the 1984 Democratic primary, as follows.

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

In the empirical part of the paper, Bartels does not focus on this coordination aspect (i.e., Hart versus other non-Mondale candidates), but rather analyzes the dynamic aspects of how expectations about the candidates' winning chances influenced voters' preferences. See also Bartels (1985, 1988), and Kenny and Rice (1994) which, however, all focus on two-candidate settings.

In a very clever lab experiment, Morton and Williams (1999, 2001) analyze the trade-off between learning and coordination in simultaneous and sequential elections, and show that both effects occur in later elections in their experiment. Our paper builds upon theirs in that we take it as given that voters in later elections learn about candidate quality and try to coordinate with other voters. Our main value added is that we empirically analyze the effects of this trade-off.

Also it has been widely accepted that ideological differences are not too important in primary contests. Indeed, to our knowledge, all theoretical learning 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. This implies that empirical models that ignore position differences may mistake ideological variation between sequentially voting states for learning about candidate valence.

Most of the theoretical literature on primaries focuses on a contest between only two candidates, and therefore does not deal with the problem of vote-splitting between similar candidates that we focus on most in the present paper (Dekel and Piccione 2000; Klumpp and Polborn 2006; Callander 2007; Schwabe (2010)).

An exception is Knight and Schiff (2010) who develop a model of voter learning about candidate quality in which voters in later states receive some imperfect information about the signal that voters in earlier states observed. Voters update, taking all pieces of information into account. Using poll data from the 2004 US Democratic Presidential Primaries, Knight and Schiff 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]

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 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 follows Krasa and Polborn (2010) and 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 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 candi-

[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.

date j, as in probabilistic voting models.[3] A possible interpretation is that candidates 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.

Voters are uncertain about the candidates' valences, which are assumed to be independent draws 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. 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 are not individual-specific. 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 that 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]

Since we focus on the implications of voters' learning behavior, the specific rules for who wins

[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.

[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.

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.

Proposition 1 provides the total number of votes for candidate j in this setting.

Proposition 1 *The total vote share of candidate $j \in J_0^s$ is*

$$\begin{aligned}(1-\mu^s)\int_{-\infty}^{\infty}\prod_{J_0^s\setminus\{j\}}\Phi\left(\frac{\hat{v}_j^s-\hat{v}_{j'}^s+\varepsilon_j}{\sigma_\varepsilon}\right)\cdot\prod_{J_1}\Phi\left(\frac{\lambda+\varepsilon_j+\hat{v}_j^s-\hat{v}_{j'}^s}{\sigma_\varepsilon}\right)\cdot\phi_\varepsilon(\varepsilon_j)d\varepsilon_j+\\ \mu^s\int_{-\infty}^{\infty}\prod_{J_0^s\setminus\{j\}}\Phi\left(\frac{\hat{v}_j^s-\hat{v}_{j'}^s+\varepsilon_j}{\sigma_\varepsilon}\right)\cdot\prod_{J_1^s}\Phi\left(\frac{-\lambda+\varepsilon_j+\hat{v}_j^s-\hat{v}_{j'}^s}{\sigma_\varepsilon}\right)\cdot\phi_\varepsilon(\varepsilon_j)d\varepsilon_j\end{aligned} \tag{3}$$

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

$$\begin{aligned}(1-\mu^s)\int_{-\infty}^{\infty}\prod_{J_0^s}\Phi\left(\frac{-\lambda+\hat{v}_j^s-\hat{v}_{j'}^s+\varepsilon_j}{\sigma_\varepsilon}\right)\cdot\prod_{J_1^s\setminus\{j\}}\Phi\left(\frac{\varepsilon_j+\hat{v}_j^s-\hat{v}_{j'}^s}{\sigma_\varepsilon}\right)\cdot\phi_\varepsilon(\varepsilon_j)d\varepsilon_j+\\ \mu^s\int_{-\infty}^{\infty}\prod_{J_0^s}\Phi\left(\frac{\lambda+\hat{v}_j^s-\hat{v}_{j'}^s+\varepsilon_j}{\sigma_\varepsilon}\right)\cdot\prod_{J_1^s\setminus\{j\}}\Phi\left(\frac{\varepsilon_j+\hat{v}_j^s-\hat{v}_{j'}^s}{\sigma_\varepsilon}\right)\cdot\phi_\varepsilon(\varepsilon_j)d\varepsilon_j\end{aligned} \tag{4}$$

Proof. See Appendix. ■

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

$$\begin{aligned} R_{AB} = (1-\mu) \int_{-\infty}^{\infty} \left[\Phi\left(\frac{\hat{v}_A - \hat{v}_B - \varepsilon}{\sigma_\varepsilon}\right) \Phi\left(\frac{\hat{v}_B - \hat{v}_C + \lambda + \varepsilon}{\sigma_\varepsilon}\right)\right] \phi_\varepsilon(\varepsilon) d\varepsilon + \\ \mu \int_{-\infty}^{\infty} \left[\Phi\left(\frac{\hat{v}_A - \hat{v}_B - \varepsilon}{\sigma_\varepsilon}\right) \Phi\left(\frac{\hat{v}_B - \hat{v}_C - \lambda + \varepsilon}{\sigma_\varepsilon}\right)\right] \phi_\varepsilon(\varepsilon) d\varepsilon \end{aligned} \tag{5}$$

and

$$\begin{aligned} R_{AC} = (1-\mu) \int_{-\infty}^{\infty} \left[\Phi\left(\frac{\hat{v}_A - \hat{v}_C + \lambda + \varepsilon}{\sigma_\varepsilon}\right) \Phi\left(\frac{\hat{v}_C - \hat{v}_B - \lambda - \varepsilon}{\sigma_\varepsilon}\right)\right] \phi_\varepsilon(\varepsilon) d\varepsilon + \\ \mu \int_{-\infty}^{\infty} \left[\Phi\left(\frac{\hat{v}_A - \hat{v}_C - \lambda + \varepsilon}{\sigma_\varepsilon}\right) \Phi\left(\frac{\hat{v}_C - \hat{v}_B + \lambda - \varepsilon}{\sigma_\varepsilon}\right)\right] \phi_\varepsilon(\varepsilon) d\varepsilon \end{aligned} \tag{6}$$

Whenever $R_{AB}/R_{AC} > 1$, 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. Figure 1 suggests that the ratio is monotonically increasing 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 (6) go to zero, while (5) 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 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

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

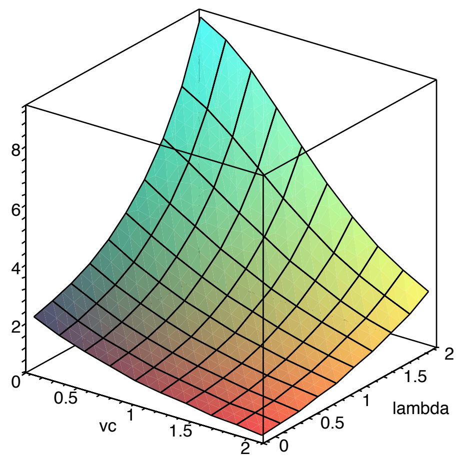


Figure 1: Ratio R_{AB}/R_{AC} for $\mu = 1/2$, $\hat{v}_A = 0$, $\hat{v}_B = 1$

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 of C. As argued above, in the case that $\hat{v}_B \geq \hat{v}_C$, our conclusion that B benefits more than his surviving competitor, if a candidate from 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.

[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.

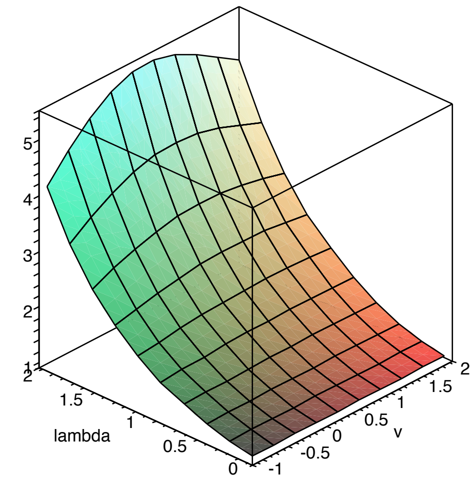


Figure 2: Ratio R_{AB}/R_{AC} for $\mu = 1/2$, $\hat{v}_A = 0$, $\hat{v}_B = \hat{v}_C = v$

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 before seeing the state-s-specific signal is distributed according to $N(\hat{v}_{j0}, \sigma_{j0}^2)$. If the state-specific signal is 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

$$\hat{v}_j^s = \frac{\sigma_\eta^2}{\sigma_{j0}^2 + \sigma_\eta^2} v_{j0} + \frac{\sigma_{j0}^2}{\sigma_{j0}^2 + \sigma_\eta^2} Z_j^s \tag{7}$$

and variance

$$(\sigma_{v_j}^s)^2 = \frac{\sigma_{j0}^2 \sigma_\eta^2}{\sigma_{j0}^2 + \sigma_\eta^2}. \tag{8}$$

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 (3) and (4), the election in state r is then captured by the following equation system:

$$(1-\mu^r)\int_{-\infty}^{\infty}\prod_{J_0^r\setminus\{j\}}\Phi\left(\frac{\hat{v}_j^r-\hat{v}_{j'}^r+\varepsilon_j}{\sigma_\varepsilon}\right)\cdot\prod_{J_1^r}\Phi\left(\frac{\lambda+\varepsilon_j+\hat{v}_j^r-\hat{v}_{j'}^r}{\sigma_\varepsilon}\right)\cdot\phi_\varepsilon(\varepsilon_j)d\varepsilon_j+$$
$$\mu^r\int_{-\infty}^{\infty}\prod_{J_0^r\setminus\{j\}}\Phi\left(\frac{\hat{v}_j^r-\hat{v}_{j'}^r+\varepsilon_j}{\sigma_\varepsilon}\right)\cdot\prod_{J_1^r}\Phi\left(\frac{-\lambda+\varepsilon_j+\hat{v}_j^r-\hat{v}_{j'}^r}{\sigma_\varepsilon}\right)\cdot\phi_\varepsilon(\varepsilon_j)d\varepsilon_j=W_j^r\ ,\forall j\in J_0^r$$
$$(1-\mu^r)\int_{-\infty}^{\infty}\prod_{J_0^r}\Phi\left(\frac{-\lambda+\hat{v}_j^r-\hat{v}_{j'}^r+\varepsilon_j}{\sigma_\varepsilon}\right)\cdot\prod_{J_1^r\setminus\{j\}}\Phi\left(\frac{\varepsilon_j+\hat{v}_j^r-\hat{v}_{j'}^r}{\sigma_\varepsilon}\right)\cdot\phi_\varepsilon(\varepsilon_j)d\varepsilon_j+$$
$$\mu^r\int_{-\infty}^{\infty}\prod_{J_0^r}\Phi\left(\frac{\lambda+\hat{v}_j^r-\hat{v}_{j'}^r+\varepsilon_j}{\sigma_\varepsilon}\right)\cdot\prod_{J_1^r\setminus\{j\}}\Phi\left(\frac{\varepsilon_j+\hat{v}_j^r-\hat{v}_{j'}^r}{\sigma_\varepsilon}\right)\cdot\phi_\varepsilon(\varepsilon_j)d\varepsilon_j=W_j^r\ ,\forall j\in J_1^r \qquad (9)$$

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 2 *There exists a unique vector* $(0, x_2, x_3, \ldots x_k)$ *such that all solutions of* (9) *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 (9) 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 3, 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 late-voting states share a lot of common information (i.e., the information inferred from states that voted 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 3 *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 3 thus suggests the following Hypothesis 2.

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).

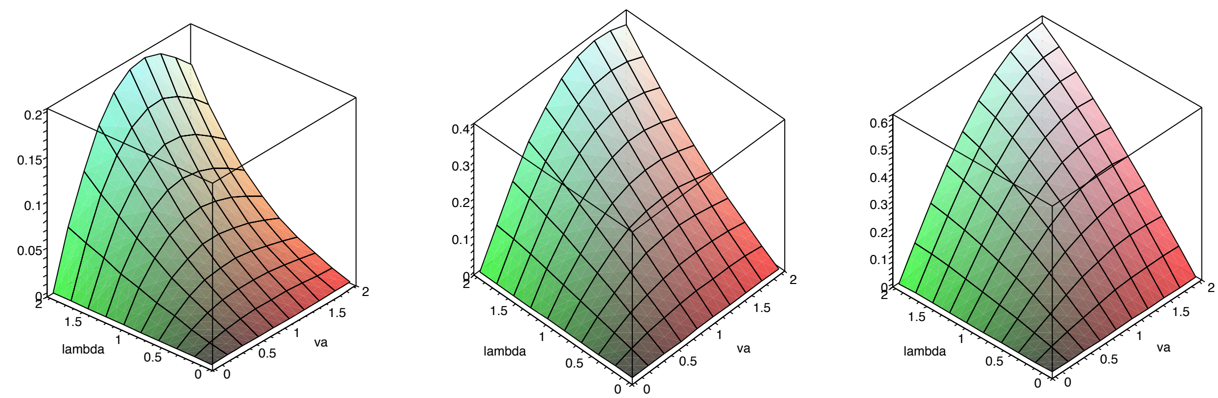
To analyze this question, let us focus on the case where there are initially three candidates, two of whom (say, A and B) have position 0, while the third one (C) has position 1. A decrease in μ benefits the vote shares of candidates A and B. Candidate A benefits at least as much as candidate B if and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

$$\begin{aligned}\int_{-\infty}^{\infty}\Phi\left(\frac{v_A-v_B+\varepsilon}{\sigma_\varepsilon}\right)\left[\Phi\left(\frac{\lambda+v_A-v_C+\varepsilon}{\sigma_\varepsilon}\right)-\Phi\left(\frac{-\lambda+v_A-v_C+\varepsilon}{\sigma_\varepsilon}\right)\right]\phi_\varepsilon(\varepsilon)d\varepsilon-\\ \int_{-\infty}^{\infty}\Phi\left(\frac{v_B-v_A+\varepsilon}{\sigma_\varepsilon}\right)\left[\Phi\left(\frac{\lambda+v_B-v_C+\varepsilon}{\sigma_\varepsilon}\right)-\Phi\left(\frac{-\lambda+v_B-v_C+\varepsilon}{\sigma_\varepsilon}\right)\right]\phi_\varepsilon(\varepsilon)d\varepsilon\geq 0.\end{aligned} \tag{10}$$

Without loss of generality, suppose that $v_A > v_B$. Whether (10) holds in general is difficult to determine. However, for $\lambda = 0$, (10) 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 (10) is satisfied as strict inequality.

Figure 3 displays the left-hand side of (10), 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 (10) is increasing in λ, so that it appears plausible that (10) holds for any $\lambda > 0$.

We now focus on relative changes. Proposition 4 shows that, if λ is sufficiently large, then the weaker candidate benefits *proportionately more* than the strong candidate (i.e., relative to previous



(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 $v_C = -1, 0, 1$

vote share) from a favorable ideological shift of the electorate.

Proposition 4 *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 4 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 4 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 a favorable ideological shift in the electorate.

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 six of the 2000, 2004, 2008, and 2012 United States Presidential primaries (we exclude the 2004 Republican and the 2012 Democratic primaries because the incumbent Presidents were 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 do not fall into this category because they are too far away from their party's mainstream, and they rather run to represent a particular energized constituency in order to demonstrate that the party needs to pay attention to its preferences. These candidacies do not fit our theoretical model well, as there is no meaningful way in which voters update their beliefs about them, and we exclude these cases from our data set.

The most successful excluded candidates are probably Dennis Kucinich (Democratic primary 2004 and 2008) and Ron Paul (Republican primary 2008 and 2012). Their vote share is usually higher in low-turnout contests later in the sequence in which their energized base represents a larger fraction of the electorate. In contrast, unsuccessful but potentially "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 A1 and A2 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, the main ideological fault line appears to be 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 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 identify 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. In the 2012 primary, however, Romney was the moderate standard-bearer, facing Gingrich and Santorum who were supported by the Republican base. Their splitting of the conservative vote helped him win the nomination. For that year, Romney is classified as a moderate.

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

[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.

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

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 A1 and A2.

For these candidates and election contests, we obtain the vote percentage in the primary or caucus of each state from the Federal Election Commission and major media sources. These vote shares are reported in Tables A1 and A2. 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 A1.

6 Results

We start, in Sections 6.1 and 6.2, by analyzing the effects of different positions on the support of candidates. Specifically, we show that a candidate benefits considerably more from the withdrawal of another candidate in the same position than from the withdrawal of a candidate in the other position. Conversely, vote-splitting is a severe problem as long as several candidates in the same position are in the race. A crucial implication of this result is that any normative analysis of the optimal primary structure should take into account a setting in which both coordination and learning play a role, rather than one that focuses exclusively on learning. Finally, in Section 6.3, we analyze the evidence on the nature of learning. In particular, we show that the observed decline in vote share variability is in fact more consistent with voter learning from previous election results rather than just from the passage

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

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

of time.

6.1 Non-Parametric Mean-Variance Analysis

We start our data analysis by pooling all data and simply comparing the candidates' average vote shares 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 take all primary elections in which κ candidates in one position and κ' candidates in the opposite position compete.

The advantage of this approach is that it is not based on any structural assumptions or even econometric specifications. Thus, it can potentially provide convincing supportive evidence for our theoretical framework. This advantage comes at the cost that the analysis in Section 6.1 is informal in nature and no formal statistical tests are performed. Also, we 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, and then proceed to more formal tests of our theory in Section 6.2.

Let $\text{VoteShare}_{j,s,y}$ be the vote share of candidate j (measured on a 0-100% scale) in state s and year y, and let his political party and political position be $p(j)$ and $a(j)$, respectively. Next, denote by $\|K_{s,p,l,y}\|$ the cardinality of the set of candidates in state contest s, political party p, political location l, and year y. Then, the average vote share of candidates over the state contests where they compete against $\kappa - 1$ other candidates of the same political position and κ' candidates in the opposite position is given by

$$\text{VoteShare}^{\kappa,\kappa'} = \frac{1}{N^{\kappa,\kappa'}} \sum_j \sum_{s,y:(\left\|K_{s,p(j),a(j),y}\right\|=\kappa \wedge \left\|K_{s,p(j),|1-a(j)|,y}\right\|=\kappa')} \text{VoteShare}_{j,s,y} \tag{11}$$

where $N^{\kappa,\kappa'}$ is the number of observations such that $\left\|K_{s,p,a(j),y} = \kappa\right\|$ and $\left\|K_{s,p,|1-a(j)|,y} = \kappa'\right\|$, and the inner sum is over all the state/year combinations in which candidate j competed against $\kappa - 1$ opponents of the same political position as him/her and κ' opponents of the other political position. We report $\text{VoteShare}^{\kappa,\kappa'}$ in Table 1, for all different candidate configurations that appear in our data. These results 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, $1/\kappa$ (or $100/\kappa$%).[14] Remarkably, it never happens that all participants in a primary belong to the same political position, and thus these configurations are

[14] The double summation consists of the sum of vote shares of all candidates in those contests with κ candidates, all

not listed in Table 1. 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, 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: A reduction in the number of candidates in the same position decreases the total vote share of the candidates in that position because there are some cross-over voters who change to a candidate in the opposite position. Formally, $\kappa \cdot \text{VoteShare}^{\kappa,\kappa'} > (\kappa - 1) \cdot \text{VoteShare}^{\kappa-1,\kappa'}$.

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, the vote share of the candidate in the previously crowded position increases from 28.6% to 50%, while the vote share of the competitor increases only from 42.8% to 50% (remember that, by definition, when $\kappa = \kappa' = 1$, the average vote share of candidates is 0.5). Or, interpreted in the other direction: A very competitive race between two candidates in different positions, each attracting 50 percent of the votes, can become very non-competitive when another candidate enters, because the lonely candidate now attracts significantly more votes than each of his competitors. This vote-splitting may lead to the victory of a candidate who would lose if he had only one competitor. Note also that, if positions were irrelevant for voters, then entry by the third candidate would instead reduce the vote share of existing candidates to 1/3.

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%, while it increases the average share of the two initially less crowded candidates only from 21.5% to 25%.[15]

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.[16]

in the same position. Since vote shares sum to 100 percent, this is simply equal to the number of such contests. The corresponding number of observations in the sample, $N^{\kappa,\kappa'}$ is equal to the number of such contests times κ, since there is one observation per candidate for each state/year.

[15] Of course, $\kappa \cdot \text{VoteShare}^{\kappa,\kappa'} + \kappa' \cdot \text{VoteShare}^{\kappa',\kappa} = 100$ holds as an identity. Deviations from this in Table 1, such as here where $3 \times 17.3\% + 48.2\% = 100.1\%$, are due to rounding.

[16] The precise implications of the theory are for expected vote share comparisons between κ candidates in one position

The only case that contradicts the 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 42.8%. 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 1 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: any trends in vote shares won't be systematically related to the number of candidates in each position. Treating political parties and positions as fungible does not create any biases either, provided that the political locations do not differ systematically in voter popularity. Subsequent analysis, described in the next section, suggests that this is not far off the mark.[17] Therefore, the figures in Table 1 are reasonable ballpark figures on the validity of Hypothesis 1. The main value of this analysis 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 generating the results in this subsection, they 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. This

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 iteratively.

[17] 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.

subsequent analysis helps provide direct statistical tests for Hypothesis 1 as well, and also measures quantitatively the extent of substitutability of candidates within and across political positions.

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 can investigate whether vote shares exhibit patterns that are consistent with the theoretical model 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.

The generic specification is of the form

$$\text{VoteShare}_{j,s,y} = \mathbf{b}\mathbf{X}_{\mathbf{j},\mathbf{s},\mathbf{y}} + \epsilon_{j,s,y} \tag{12}$$

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 $\mathbf{X}_{\mathbf{j},\mathbf{s},\mathbf{y}}$ is a vector of explanatory variables. The disturbance term $\epsilon_{j,s,y}$ likely has a special correlation structure, which means that simply accounting for heteroskedasticity using White's heteroskedasticity consistent standard errors may not be sufficient. A more conservative approach recognizes that vote shares are dependent within a party/state/year pair: a larger vote share by one candidate leads to lower than average vote shares for the others in a party's primary. This negative correlation implies that there is less information in the dataset than may be apparent given the number of observations. Clustering the standard errors by party/state/year takes this into account. We do so for all regressions reported in this paper, which yields higher (but we believe more appropriate) standard errors.[18] The disturbance terms also likely have dependence within candidate/round pairs. A candidate who obtains an higher than expected vote share in one state in a given round may also obtain higher than expected vote shares in other states if some of the signal voters receive is (contrary to the model) observable nationally. Therefore, we also cluster standard errors by candidate/round, except in the specifications were we use candidate/round fixed effects which should account for any national

[18]This clustering is also implemented for the regressions of vote share variability on the premise that a larger deviation from predicted shares for one candidate is more likely to result in a larger deviation for other candidates, yielding a positive dependence within party/state/year pairs.

signals.[19] This type of clustering also controls for disturbance correlation within candidate/round observations that arises from signals received in prior rounds.

Our simplest specification estimates the equation

$$\text{VoteShare}_{j,s,y} = \alpha + \beta_1 \text{CanDif}_{j,s,y} + \beta_2 \text{CanOwn}_{j,s,y} + \epsilon_{j,s,y} \tag{13}$$

where $\text{CanOwn}_{j,s,y}$ 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 1 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 2, 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. In fact, the effect of an additional candidate in the opposite political location is small enough that when we use our conservative clustered standard errors, it is only significant at the ten percent level (and loses significance in most of the subsequent regressions).

We next investigate whether these results are 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). In the estimation of the equation

$$\text{VoteShare}_{j,s,y} = \alpha + \beta_1 \text{CanDif}_{j,s,y} + \beta_2 \text{CanOwn}_{j,s,y} + \gamma_1 \text{Moderate}_j + \gamma_2 \text{Outsider}_j + \epsilon_{j,s,y} \tag{14}$$

the coefficients of γ_1 and γ_2 are both not statistically significant and small in numerical value (a difference of approximately one 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% versus 57% for conservatives. However, Model 2 attributes this substantial difference to the fact that there are more conservative than moderate Republican candidates, which can lead to a higher aggregate vote share for their political position. Moreover, with conservative Republicans diverting a disproportionate fraction votes

[19] We believe that any clustering must be justified on a priori grounds. Incidentally, we also note that two-way clustering in those specifications leads to non-positive definite covariance matrices.

from each other rather than from moderate Republicans, there is no residual advantage to being conservative.[20] We should also note that the coefficient of γ_1 is largely a McCain effect, as only two other Republican candidates (Giuliani and Romney in 2012) are labeled as moderate. 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

$$\begin{aligned}\text{VoteShare}_{j,s,y} \quad = \quad & \alpha + \beta_{1R}\text{CanDif}_{j,s,y}\text{Rep}_j + \beta_{2R}\text{CanOwn}_{j,s,y}\text{Rep}_j \\ & +\beta_{1D}\text{CanDif}_{j,s,y}\text{Dem}_j + \beta_{2D}\text{CanOwn}_{j,s,y}\text{Dem}_j + \epsilon_{j,s,y} \qquad (15)\end{aligned}$$

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 2, suggest that voter segmentation across political locations is quantitatively large for both parties but more pronounced for the Democratic Party. For the Democrats, a candidate's vote share is only negligibly affected by competition from 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 smaller in quantitative terms.[21]

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 plausibly correspond 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, 2008 or 2012 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

[20] The average combined share of the two political locations for Democratic primaries is much closer (47.5% versus 52.5%) and also closer 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.

[21] Adding location dummies to this regression is problematic because with Moderate_i being largely a dummy for McCain, the Republican substitutability effect is identified primarily 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).

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 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

$$\begin{aligned} \text{VoteShare}_{j,s,y} &= \alpha + \beta_1 \text{CanDif}_{j,s,y} + \beta_2 \text{CanOwn}_{j,s,y} + \gamma_1 \text{Moderate}_j + \gamma_2 \text{Outsider}_j \\ &\quad + \gamma_{1C} \text{Moderate}_j Clinton92\%_s + \gamma_{2P} \text{Outsider}_j Perot92\%_s + \epsilon_{j,s,y} \end{aligned} \tag{16}$$

and the one of the much more flexible Model 5 by

$$\text{VoteShare}_{j,s,y} = \alpha_{j,t,y} + \gamma_{1C} \text{Moderate}_j Clinton92\%_s + \gamma_{2P} \text{Outsider}_j Perot92\%_s + \epsilon_{j,s,y} \tag{17}$$

where $Perot92\%_s$ and $Clinton92\%_s$ are Perot's and Clinton's vote share in state s in the 1992 Presidential election, respectively,[22] 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 502 to 382. 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 0.75 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 the more restrictive Model 4. The fact that the evidence is not as strong as for the Republicans may be because most of Perot's voters are conservative populists, and their presence may less strongly correlated with Democratic populism.

The final set of regressions, Models 6 and 7, are intended to test Hypothesis 3, which asserts that weak candidates of an ideological position benefit disproportionately relative to strong candidates of the same ideological position from a tilt in voter preference towards that ideological position. This

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

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. 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, suggesting that any estimation approach should be based on variants of Model 5 that include candidate-round effects. The inclusion of these effects also ensures that the shift is voter preferences is measured "holding everything else constant."

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 but reasonable measure. Candidates with high relative valence, as perceived in round t, will have higher values of $MeanShr_{j,t,y}$. The theory shows that the effect of electorate preferences 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. For example, a shift of electoral preferences towards a candidate's position has a larger effect on that candidate's vote share the fewer his opponents of the same position, the fewer the overall number of opponents, and the lower the opponents' valence. These factors lead to lower values of $MeanShr_{j,t,y}$ as well. In this sense, 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 statics developed by the theoretical model. 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 based on their privately observed signal and the set of candidates is the same in all such contests.

However, adding $MeanShr_{j,t,y}$ on the right hand side of the regression suffers from a serious shortcoming: a higher than expected vote share by candidate j in any state of round t leads to a higher value of $MeanShr_{j,t,y}$. Such positive correlation leads to an upward bias in the regression coefficients of $MeanShr_{j,t,y}$ and its interactions. A specification (Model 6) that does not suffer from this endogeneity shortcoming is:

$$\begin{aligned}\text{VoteShare}_{j,s,y} &= \alpha_{j,t,y} + \gamma_{1C}\text{Moderate}_j\text{Clinton92\%}_s + \gamma_{2P}\text{Outsider}_j Perot92\%_s \\ &+ \{\gamma_{1Cs1}\text{Moderate}_j + \gamma_{1Cs0}[1-\text{Moderate}_j]\}\,\text{Clinton92\%}_s MeanShr_{j,t/s,y} \qquad (18) \\ &+ \{\gamma_{1Ps1}\text{Outsider}_j + \gamma_{1Ps0}[1-\text{Outsider}_j]\}\,\text{Perot92\%}_s MeanShr_{j,t/s,y} + \epsilon_{j,s,y}\end{aligned}$$

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] This specification, too, however, has a potential endogeneity

[23]This specification uses a different proxy for every state, since $MeanShr_{j,t/s,y}$ takes a different value for all states in a

concern if, contrary to our model, there is a common information shock observed in all states voting in a round. In that case, the $\epsilon_{j,s,y}$ would be correlated with the vote shares of the candidate in other states voting contemporaneously. A more conservative approach is to lag the $MeanShr_{j,t,y}$ variable by one round, i.e., use as a proxy of valence $MeanShr_{j,t-1,y}$. This yields Model 7 below.

$$\begin{aligned}\text{VoteShare}_{j,s,y} &= \alpha_{j,t,y} + \gamma_{1C}\text{Moderate}_j\text{Clinton92\%}_s + \gamma_{2P}\text{Outsider}_j Perot92\%_s \\ &+ \{\gamma_{1Cs1}\text{Moderate}_j + \gamma_{1Cs0}[1-\text{Moderate}_j]\}\,\text{Clinton92\%}_s MeanShr_{j,t-1,y} \qquad (19) \\ &+ \{\gamma_{1Ps1}\text{Outsider}_j + \gamma_{1Ps0}[1-\text{Outsider}_j]\}\,\text{Perot92\%}_s MeanShr_{j,t-1,y} + \epsilon_{j,s,y}\end{aligned}$$

This specification is not necessarily better than the one in (18) for two reasons. First, lagging the mean share provides a more noisy measure of perceived valence for round t because it does not include the signals received in that round. Second, the set of candidates is no longer guaranteed to be the same across round t and $t-1$, and this introduces an additional source of noise in the valence proxy. Both effects can lead to substantial attenuation bias. Nonetheless, examining the results of these specifications jointly likely gives a more balanced picture.

We therefore estimate Model 6 and Model 7 and discuss their results of these regressions together. We first point out that the two specifications yield parameter estimates that are concordant with respect to sign, except in one case where the sign flip is associated with a variable that is not statistically significant. Moreover, for both parties the estimates exhibit a higher degree of statistical significance for Model 6 than for Model 7, which is expected given the discussion above comparing the advantages and disadvantages of lagging the mean share proxy. Moderate Republicans do better than conservatives in states with strong support for Clinton in 1992, and this difference is somewhat decreasing in or broadly independent of their vote share ($\gamma_{1Cs1} < \gamma_{1Cs0}$ for Model 6, while they are similar in size in Model 7). Consistent with Hypothesis 3, high valence (high average vote share) conservatives are hurt proportionately less than low valence conservatives 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 (the effect is significant only for Model 6).[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

given round. Because $MeanShr_{j,t/s,y}$ is a proxy for perceived valence is state s, and does not have a direct causal effect, there are no endogeneity concerns from the "reduced" form of the system that consists of the vote share equations for all states in a given round. This is unlike, say, the supply-demand system where price has a causal effect on the quantity supplied and the quantity demanded.

[24]The effect for conservatives 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 fact in one instance they appear to be more sensitive even in absolute terms. Similar conclusions are obtained for the Democratic candidates, even though the two specifications differ somewhat in their results. 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 primaries progress. Trends in this variance would be indicative of voter learning about candidates, as discussed in Section 4. We turn to this analysis next.

6.3 Econometric Analysis of Share Variability

We now analyze the nature of voter learning using of vote share variability as our main focus. 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 provides an additional component of vote share variability, and since this uncertainty is slowly resolved over time, the theoretical model posits that vote share variability becomes declines over time as well. Moreover, since additional information moves perceptions (and thus vote shares) by a progressively smaller amount, the largest decline in variability should happen early.

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. We need to ensure that the greatest proportion of systematic variation in vote shares is removed, without removing any component of the residuals that helps identify learning effects or introducing any biases in the estimation of such effects. With respect to 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. Therefore, 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 to 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. Since we focus here is on the time variation of the residuals, we report as a representative model the results based on Model 6, which is one the most flexible specifications, and where the Perot/Clinton covariates have the largest effects.[26]

Let $NumCand_{j,s,y}$ be the number of candidates contesting 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 state s. We estimate the regressions

$$|\hat{\epsilon}_{j,s,y}| = a + b\ NumCand_{j,s,y} + c\ PriorSignals_{j,s,y} + u_{j,s,y} \tag{20}$$

and

$$|\hat{\epsilon}_{j,s,y}| = a + b\ NumCand_{j,s,y} + c\ PriorSignals_{j,s,y} + d\ PriorSignals^2_{j,s,y} + u_{j,s,y} \tag{21}$$

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 (20) and (21) 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 Table 3. Consistent with Hypothesis 2, residual variance is decreasing with the number of prior contests for all specifications, an effect that is statistically significant at the 5 percent level in 6 out of 8 regressions, and significant at the 10 percent level in one of the remaining two (when quadratic effects are included, significance is based on the joint test). Moreover, since voters initially have weaker priors about candidates, new information can move their opinions more easily, which implies that vote share variability should decline fastest in the early rounds. Consistent with this expectation, we find that the coefficient of $PriorSignals^2_{j,s,y}$ is positive in all specifications, and of a value that implies that the marginal effect of more contests is near 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. However, $PriorSignals_{j,s,y}$ and $PriorSignals^2_{j,s,y}$ are highly correlated and thus statistical significance of each variable suffers when they are jointly included in the regressions, even though the two variables are always jointly

[26]The residuals of the other two models give similar results.

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

significant (for example, standard errors of $PriorSignals_{j,s,y}$ triple when $PriorSignals^2_{j,s,y}$ is added).[28] Finally, the number of candidates has a negative effect on variance, as expected, though the effect is typically not statistically significant when we adjust the dependent variable for the number of candidates.

Even though this variability reduction effect due to learning from earlier election results is statistically significant and exhibits the expected diminishing pattern, it is quantitatively small relative to other factors: it explains only about 2 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.[29]

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 local press coverage) 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 in early states. Later, 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 this explanation implies that vote share residuals for candidates in the same political position are strongly negatively correlated and largely cancel out. This suggests a test for whether this alternative explanation is the driving force behind the reduction of share variability. In particular, under it vote share variability at the political position level, controlling for candidate mean shares, should not have a clear trend over time. We implement this test 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 equations (20) and (21) 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 in the number of observations). The estimates are reported in the first 4 columns of Table 4. The pattern of coefficient

[28]Statistical significance is higher if we forgo clustering.

[29]It is not surprising, and in fact reassuring, that when one includes the variable $PriorSignals_{j,s,y}$ in the vote share regressions in Table 2 it comes out uniformly insignificant.

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.[30] 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 (20) and (21), i.e., a variable that is akin to a time trend and does not take into consideration the number of states that vote in a given round, the coefficients on that variable are not statistically significant. This is reported in the last four columns of Table 4 for the counterparts of the regressions in Table 3 (omitting the quadratic models). Thus, it is not the passage of time that is associated with reduced variability, but rather 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 multi-candidate elections is that candidates whose political philosophy is very similar may "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. 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

[30] Results using models 5/6/7 without any covariates are similar.

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, 2008 and 2012. Moreover, the first of these three effects is quantitatively very important, indicating the importance of intra-party divisions.

Overall, our results indicate that the organization of primaries in either sequential or simultaneous form play an important role for which candidate wins the nomination. Moreover, normative analyses that have the objective to find the optimal organization form of the primaries should be based on an environment that captures both learning and vote-splitting.

References

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.

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

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

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

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

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.

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

Morton, R. and K. Williams (2001). *Learning by Voting: Sequential Choices in Presidential Primaries and Other Elections*. University of Michigan Press.

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.

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

8 Appendix (not for publication)

Proof of Proposition 1. 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

$$\hat{v}_j^s + \varepsilon_j - \lambda d(j,\theta) \geq \max_{j'}(\hat{v}_{j'}^s + \varepsilon_{j'} - \lambda d(j',\theta)). \tag{22}$$

For a given ε_j, (22) is satisfied if and only if

$$\varepsilon_{j'} < \hat{v}_j^s - \hat{v}_{j'}^s + \varepsilon_j - \lambda[d(j,\theta) - d(j',\theta)] \text{ for all } j' \neq j \ . \tag{23}$$

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

$$\prod_{J_0^s \backslash \{j\}} \Phi\left(\frac{\hat{v}_j^s - \hat{v}_{j'}^s + \varepsilon_j}{\sigma_\varepsilon}\right) \cdot \prod_{J_1^s} \Phi\left(\frac{\lambda + \varepsilon_j + \hat{v}_j^s - \hat{v}_{j'}^s}{\sigma_\varepsilon}\right). \tag{24}$$

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

$$\int_{-\infty}^{\infty} \prod_{J_0^s \backslash \{j\}} \Phi\left(\frac{\hat{v}_j^s - \hat{v}_{j'}^s + \varepsilon_j}{\sigma_\varepsilon}\right) \cdot \prod_{J_1} \Phi\left(\frac{\lambda + \varepsilon_j + \hat{v}_j^s - \hat{v}_{j'}^s}{\sigma_\varepsilon}\right) \cdot \phi_\varepsilon(\varepsilon_j) d\varepsilon_j. \tag{25}$$

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

$$\int_{-\infty}^{\infty} \prod_{J_0^s \backslash \{j\}} \Phi\left(\frac{\hat{v}_j^s - \hat{v}_{j'}^s + \varepsilon_j}{\sigma_\varepsilon}\right) \cdot \prod_{J_1^s} \Phi\left(\frac{-\lambda + \varepsilon_j + \hat{v}_j^s - \hat{v}_{j'}^s}{\sigma_\varepsilon}\right) \cdot \phi_\varepsilon(\varepsilon_j) d\varepsilon_j. \tag{26}$$

Multiplying (25) by $(1-\mu^s)$ and (26) by μ^s, and adding up gives (3). Equation (4) is obtained analogously. ■

Derivation of (5) **and** (6)**.** 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

$$\hat{v}_j^s + \varepsilon_j - \lambda d(j,\theta) \geq \hat{v}_{j'}^s + \varepsilon_{j'} - \lambda d(j',\theta) \geq \max_{k \neq j,j'}(\hat{v}_k^s + \varepsilon_k - \lambda d(k,\theta)). \tag{27}$$

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 (27) is satisfied if and only if

$$\varepsilon_k < \hat{v}_{j'}^s - \hat{v}_k^s + \varepsilon_{j'} - \lambda[d(j',\theta) - d(k,\theta)] \text{ for all } k \neq j, j' \ . \tag{28}$$

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

$$\prod_{J_0^s \backslash \{j,j'\}} \Phi\left(\frac{\hat{v}_{j'}^s - \hat{v}_k^s + \varepsilon_{j'}}{\sigma_\varepsilon}\right) \cdot \prod_{J_1^s} \Phi\left(\frac{\lambda + \varepsilon_{j'} + \hat{v}_{j'}^s - \hat{v}_k^s}{\sigma_\varepsilon}\right). \tag{29}$$

Turning to the first inequality in (27), 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

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

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

$$\int_{-\infty}^{\infty} \Phi\left(\frac{\varepsilon_{j'} + \hat{v}^s_j - \hat{v}^s_{j'}}{\sigma_\varepsilon}\right) \prod_{J^s_0\backslash\{j,j'\}} \Phi\left(\frac{\hat{v}^s_{j'} - \hat{v}^s_k + \varepsilon_{j'}}{\sigma_\varepsilon}\right) \cdot \prod_{J^s_1} \Phi\left(\frac{-\lambda + \varepsilon_{j'} + \hat{v}^s_{j'} - \hat{v}^s_k}{\sigma_\varepsilon}\right) \cdot \phi_\varepsilon(\varepsilon_{j'}) d\varepsilon_{j'}. \quad (31)$$

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

$$(1-\mu^s)\int_{-\infty}^{\infty} \Phi\left(\frac{\hat{v}^s_j - \hat{v}^s_{j'} - \varepsilon_{j'}}{\sigma_\varepsilon}\right) \prod_{J^s_0\backslash\{j,j'\}} \Phi\left(\frac{\hat{v}^s_{j'} - \hat{v}^s_k + \varepsilon_{j'}}{\sigma_\varepsilon}\right) \cdot \prod_{J^s_1} \Phi\left(\frac{\varepsilon_{j'} + \hat{v}^s_{j'} - \hat{v}^s_k + \lambda}{\sigma_\varepsilon}\right) \cdot \phi_\varepsilon(\varepsilon_{j'}) d\varepsilon_{j'} +$$
$$\mu^s\int_{-\infty}^{\infty} \Phi\left(\frac{\hat{v}^s_j - \hat{v}^s_{j'} - \varepsilon_{j'}}{\sigma_\varepsilon}\right) \prod_{J^s_0\backslash\{j,j'\}} \Phi\left(\frac{\hat{v}^s_{j'} - \hat{v}^s_k + \varepsilon_{j'}}{\sigma_\varepsilon}\right) \cdot \prod_{J^s_1} \Phi\left(\frac{\varepsilon_{j'} + \hat{v}^s_{j'} - \hat{v}^s_k - \lambda}{\sigma_\varepsilon}\right) \cdot \phi_\varepsilon(\varepsilon_{j'}) d\varepsilon_{j'} \quad (32)$$

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 $j \in J^s_0$ and $j' \in J^s_1$) is then $R_{01}(j, j') =$

$$(1-\mu^s)\int_{-\infty}^{\infty} \Phi\left(\frac{\hat{v}^s_j - \hat{v}^s_{j'} + \lambda - \varepsilon_{j'}}{\sigma_\varepsilon}\right) \prod_{J^s_0\backslash\{j,j'\}} \Phi\left(\frac{\hat{v}^s_{j'} - \hat{v}^s_k + \varepsilon_{j'} - \lambda}{\sigma_\varepsilon}\right) \cdot \prod_{J^s_1} \Phi\left(\frac{\varepsilon_{j'} + \hat{v}^s_{j'} - \hat{v}^s_k}{\sigma_\varepsilon}\right) \cdot \phi_\varepsilon(\varepsilon_{j'}) d\varepsilon_{j'} +$$
$$\mu^s\int_{-\infty}^{\infty} \Phi\left(\frac{\hat{v}^s_j - \hat{v}^s_{j'} - \lambda - \varepsilon_{j'}}{\sigma_\varepsilon}\right) \prod_{J^s_0\backslash\{j,j'\}} \Phi\left(\frac{\hat{v}^s_{j'} - \hat{v}^s_k + \varepsilon_{j'} + \lambda}{\sigma_\varepsilon}\right) \cdot \prod_{J^s_1} \Phi\left(\frac{\varepsilon_{j'} + \hat{v}^s_{j'} - \hat{v}^s_k}{\sigma_\varepsilon}\right) \cdot \phi_\varepsilon(\varepsilon_{j'}) d\varepsilon_{j'}. \quad (33)$$

Analogous conditions to (32) and (33) can be derived for candidate j being located at $a = 1$.

Equations (5) and (6) are special cases of (32) and (33). ■

Proof of Proposition 2. Existence follows by construction: Since the vector W^r is generated using the realized vector of estimated valences $(\hat{v}^r_j)_{j=1,\ldots,k}$, a solution to (9) exists. Furthermore, it is clear that any vector of the form $(0, x_2, x_3, \ldots, x_k)+(c, c, \ldots, c)$ also satisfies (9). It remains to be shown that there cannot be a solution of the form $(0, y_2, y_3, \ldots, y_k)$ with $(0, y_2, y_3, \ldots, y_k) \neq (0, x_2, x_3, \ldots, x_k)$. Assume to the contrary, and let $\bar{k}$ be the candidate for whom $y_j - x_j$ is maximal. If $y_{\bar{k}} - x_{\bar{k}} > 0$, then substituting in the corresponding equation of (9) shows that candidate $\bar{k}$ receives a strictly higher vote share than $W^r_{\bar{k}}$, a contradiction. Similarly, let $\underline{k}$ be the candidate for whom $y_j - x_j$ is minimal. If $y_{\underline{k}} - x_{\underline{k}} < 0$, then substituting in the corresponding equation of (9) shows that candidate $\underline{k}$ receives a strictly smaller vote share than $W^r_{\underline{k}}$, a contradiction. But then, it must be true that $y_j = x_j$ for all $j = 2, \ldots, k$. ■

Proof of Proposition 3. 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

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

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

$$\left(\frac{\sigma_v^2}{\sigma_v^2+\sigma_\eta^2}\right)^2 Var(Z_j) = \left(\frac{\sigma_v^2}{\sigma_v^2+\sigma_\eta^2}\right)^2 \sigma_\eta^2 \tag{35}$$

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

$$\hat{v}_{j0,late} = \frac{\frac{\sigma_\eta^2}{R}}{\sigma_v^2 + \frac{\sigma_\eta^2}{R}} \cdot 0 + \frac{\sigma_v^2}{\sigma_v^2 + \frac{\sigma_\eta^2}{R}} \frac{\sum Z_j^r}{R} \tag{36}$$

with a variance of (using (8))

$$\frac{\sigma_v^2 \frac{\sigma_\eta^2}{R}}{\sigma_v^2 + \frac{\sigma_\eta^2}{R}}. \tag{37}$$

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

$$\hat{v}_j^s = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \frac{\sigma_v^2 \frac{\sigma_\eta^2}{R}}{\sigma_v^2 + \frac{\sigma_\eta^2}{R}}} \cdot \hat{v}_{j0,late} + \frac{\frac{\sigma_v^2 \frac{\sigma_\eta^2}{R}}{\sigma_v^2 + \frac{\sigma_\eta^2}{R}}}{\sigma_\eta^2 + \frac{\sigma_v^2 \frac{\sigma_\eta^2}{R}}{\sigma_v^2 + \frac{\sigma_\eta^2}{R}}} \cdot Z_j^s \tag{38}$$

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

$$\left(\frac{\frac{\sigma_v^2 \frac{\sigma_\eta^2}{R}}{\sigma_v^2 + \frac{\sigma_\eta^2}{R}}}{\sigma_\eta^2 + \frac{\sigma_v^2 \frac{\sigma_\eta^2}{R}}{\sigma_v^2 + \frac{\sigma_\eta^2}{R}}}\right)^2 \cdot Var(Z_j^s) = \left(\frac{1}{1 + \frac{\sigma_\eta^2\left(\sigma_v^2 + \frac{\sigma_\eta^2}{R}\right)}{\sigma_v^2 \frac{\sigma_\eta^2}{R}}}\right)^2 \cdot Var(Z_j^s) = \left(\frac{1}{1 + \frac{\sigma_\eta^2}{\sigma_v^2} + R}\right)^2 \cdot Var(Z_j^s), \tag{39}$$

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 4. 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

$$Z_A = \frac{\frac{dVS_A}{d(1-\mu)}}{VS_A} = \frac{VS_{A,0} - VS_{A,1}}{(1-\mu)VS_{A,0} + \mu VS_{A,1}} \tag{40}$$

and, similarly,

$$Z_B = \frac{\frac{dVS_B}{d(1-\mu)}}{VS_B} = \frac{VS_{B,0} - VS_{B,1}}{(1-\mu)VS_{B,0} + \mu VS_{B,1}} \tag{41}$$

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

$$\frac{VS_{A,1}}{VS_{B,1}} > \frac{VS_{A,0}}{VS_{B,0}}. \tag{42}$$

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 (42) 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 (42). 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

$$\frac{VS_{A,1}}{VS_{B,1}} = \frac{\int_{-\infty}^{\frac{v_A - v_C - \lambda}{\sqrt{2}\sigma_\varepsilon}} \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)dt}{\int_{-\infty}^{\frac{v_B - v_C - \lambda}{\sqrt{2}\sigma_\varepsilon}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}dt} = \frac{\int_{-\infty}^{\frac{v_A - v_C - \lambda}{\sqrt{2}\sigma_\varepsilon}} \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)dt}{\int_{-\infty}^{\frac{v_A - v_C - \lambda}{\sqrt{2}\sigma_\varepsilon}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\left(t - \frac{v_A - v_B}{\sqrt{2}\sigma_\varepsilon}\right)^2}{2}\right) dt} \tag{43}$$

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, (43) goes to infinity, which proves that (42) holds for λ sufficiently large. This proves the claim. ■

Table 1. Key statistics for various candidate configurations.

	Mean Share	Observations	Number of Distinct Candidates	Mean Obs per Candidate
2 Candidates in the election:				
1 candidate in same location	50.0%	168	8	21.0
3 Candidates in the election:				
2 candidates in same location	28.6%	150	10	15.0
1 candidate in same location	42.8%	75	5	15.0
4 Candidates in the election:				
3 candidates in same location	17.3%	18	6	3.0
2 candidates in same location	25.0%	4	4	1.0
1 candidate in same location	48.2%	6	2	3.0
5 Candidates in the election:				
4 candidates in same location	17.9%	8	4	2.0
3 candidates in same location	19.0%	39	6	6.5
2 candidates in same location	21.5%	26	4	6.5
1 candidate in same location	28.3%	2	1	2.0

Notes: See text for a description of these statistics.

Table 2. Estimation Results.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
CandidatesDifferentLocation	-3.59	-3.34		-3.04			
	2.18	*2.83*		*2.81*			
CandidatesOwnLocation	**-14.58**	**-14.76**		**-14.78**			
	1.60	***1.94***		***1.94***			
CandidatesDifferentLocation (Rep)			**-5.28**				
			2.03				
CandidatesOwnLocation (Rep)			**-13.90**				
			2.06				
CandidatesDifferentLocation (Dem)			-0.75				
			3.31				
CandidatesOwnLocation (Dem)			**-16.12**				
			1.67				
Moderate (Rep)		-1.26		**-32.91**			
		4.44		***8.70***			
"Outsider" (Dem)		-1.32		-0.86			
		3.66		*6.60*			
Moderate (Rep) * Clinton92%				**0.75**	**0.74**	**2.48**	1.40
				0.16	***0.29***	***0.40***	*1.28*
Moderate (Rep) * Clinton92% * MeanShare						**-4.36**	-1.71
						0.83	*3.40*
Conservative (Rep) * Clinton92% * MeanShare						**-2.17**	**-1.72**
						0.34	***0.52***
"Outsider" (Dem) * Perot92%				-0.02	0.51	**1.14**	0.31
				0.33	*0.27*	***0.40***	*0.65*
"Outsider" (Dem) * Perot92% * MeanShare						-1.65	0.64
						0.90	*1.72*
"Insider" (Dem) * Perot92% * MeanShare						**-1.52**	**-1.25**
						0.47	***0.57***
Constant	**65.44**	**65.96**	**65.37**	**65.63**	na	na	na
	4.16	***4.32***	***4.14***	***4.33***	na	na	na
Tests of parameter difference and significance:							
OwnLocation - Different Location	**-10.98**	**-11.42**		**-11.74**			
	3.01	***1.94***		***4.18***			
OwnLocation - Different Location (Rep)			**-8.61**				
			3.33				
OwnLocation - Different Location (Dem)			**-15.37**				
			4.53				
Party effects (p-value)		0.9897	0.2583				
Moderate and "Outsider" effects (p-value)		0.9230		0.0007			
Perot and Clinton effects (p-value)				0.0000	0.0068	0.0000	0.0100
Candidate * election round effects (p-value)					0.0000	0.0000	0.0000
MeanShare effects (p-value)						0.0000	0.0036
R-squared	0.3043	0.3051	0.3137	0.3165	0.7338	0.7929	0.7560

Notes: N=502 for Models 1-4. N=382 for Models 5-7. Heteroskedasticity consistent standard errors, clustered by both party/state/round and candidate/round (Models 1-4) and by party/state/round (Models 5-7), are reported in italics below the parameter and parameter difference estimates. Bold entries indicate parameter significance at the 5% level, underlined entries indicate significance at the 10% level. See text for details.

Table 3. Analysis of vote share variability in each election date/round.

	Model 5/6/7 without Clinton/Perot effects				Model 6			
	using the raw value of the residual		correcting for the number of candidates		using the raw value of the residual		correcting for the number of candidates	
NumberOfCandidates	**-1.142**	**-1.283**	-0.749	-0.948	**-1.227**	**-1.269**	**-1.014**	**-1.069**
	0.460	***0.494***	*0.525*	*0.549*	***0.431***	***0.453***	***0.435***	***0.457***
PriorSignals	**-0.133**	**-0.309**	-0.077	**-0.325**	**-0.150**	-0.203	**-0.115**	-0.184
	0.043	***0.143***	*0.049*	***0.158***	***0.034***	*0.128*	***0.037***	*0.135*
PriorSignalsSquared		0.004		0.006		0.001		0.002
		0.003		*0.004*		*0.003*		*0.003*
Constant	**13.500**	**14.941**	**12.590**	**14.619**	**12.400**	**12.827**	**11.975**	**12.545**
	1.924	***2.373***	***2.070***	***2.479***	***1.769***	***2.122***	***1.798***	***2.160***
Signal Effects (p-value)		0.0054		0.0745		0.0000		0.0067
R-squared	0.0351	0.0404	0.0112	0.0205	0.0521	0.0526	0.0292	0.0301
Observations	382	382	382	382	382	382	382	382

Notes: The dependent variable is the absolute value of the regression residual of Model 6 or Model 5/6/7 without the Clinton/Perot interactions with the candidate's political position. Correction for the number of candidates involves multiplying the residual by $[m/(m-1)]^{0.5}$, where m is the number of candidates competing in a state. Heteroskedasticity consistent standard errors, clustered by party/state/round, are reported in italics below the parameter estimates. Bold entries indicate parameter significance at the 5% level, underlined entries indicate significance at the 10% level. See text for details.

Table 4. Analysis of vote share variability: Considering alternative explanations.

	Model 6 aggregate residuals by political position				Model 5/6/7 (stripped)		Model 6	
	using the raw value of the residual		correcting for the number of candidates		using raw residuals	adjusted residuals	using raw residuals	adjusted residuals
NumberOfCandidates	0.292	0.273	0.696	0.670	-0.429	-0.114	-0.731	-0.609
	0.694	*0.698*	*0.718*	*0.721*	*0.498*	*0.555*	*0.464*	*0.473*
PriorSignals	**-0.140**	-0.182	**-0.103**	-0.160				
	0.034	*0.146*	***0.037***	*0.156*				
PriorSignalsSquared		0.001		0.001				
		0.003		*0.003*				
Round					-0.012	0.137	-0.216	-0.147
					0.174	*0.191*	*0.151*	*0.158*
Constant	**9.053**	**9.360**	**8.222**	**8.642**	**9.884**	**8.798**	**10.692**	**10.465**
	2.198	***2.440***	***2.265***	***2.513***	***2.570***	***2.755***	***2.308***	***2.377***
Signal Effects (p-value)		0.0002		0.0192				
R-squared	0.0688	0.0691	0.0483	0.0489	0.0030	0.0038	0.0106	0.0060
Observations	290	290	290	290	382	382	382	382

Notes: The dependent variable for the first four regressions is the absolute value of the sum of the residuals of all candidates in the same political position (in a given state contest); in the remaining regressions it is the absolute value of the regression residual of Model 6 or Model 5/6/7 without the Clinton/Perot interactions with the candidate's political position. Correction for the number of candidates involves multiplying the residual by $[m/(m-1)]$^0.5, where m is the number of candidates competing in a state. Heteroskedasticity consistent standard errors, clustered by party/state/round, are reported in italics below the parameter estimates. Bold entries indicate parameter significance at the 5% level, underlined entries indicate significance at the 10% level. See text for details.

Table A-1. Election Data: The 1992 General Election and the 2000 and 2012 primaries.

	1992 Election		2000 Dem. Primary			2000 Rep. Primary						2012 Rep. Primary			
State	Clinton	Perot	rd	Brad	Gore	rd	Bauer	Bush	Frbs	Keyes	McC	rd	Gngr	Rmn	Sntrm
Alabama	41%	11%										13	29%	29%	35%
Alaska	30%	28%										11	14%	32%	29%
Arizona	37%	24%				5		53%		5%	42%	8	16%	47%	27%
Arkansas	53%	10%													
California	46%	21%	5	18%	81%	7		61%		4%	35%				
Colorado	40%	23%										6	13%	35%	40%
Connecticut	42%	22%	5	42%	55%	7		46%		3%	49%				
DC	85%	4%													
Delaware	44%	20%	3			3		51%	20%	4%	25%				
Florida	39%	20%										4	32%	46%	13%
Georgia	43%	13%	5	16%	84%	7		67%		5%	28%	11	47%	26%	20%
Hawai	48%	14%										13	11%	44%	25%
Idaho	28%	27%										11	2%	62%	18%
Illinois	49%	17%										14	8%	47%	35%
Indiana	37%	20%													
Iowa	43%	19%	1	35%	63%	1	9%	41%	31%	14%	5%	1	13%	25%	25%
Kansas	34%	27%										12	14%	21%	51%
Kentucky	45%	14%													
Louisiana	46%	12%										15	16%	27%	49%
Maine	39%	30%	6	41%	54%	7		51%		3%	44%	7	6%	38%	18%
Maryland	50%	14%	6	28%	67%	7		56%		7%	36%	16	11%	49%	29%
Massachu.	48%	23%	6	37%	60%	7		32%		3%	65%	11	5%	72%	12%
Michigan	44%	19%				5		43%		5%	51%	8	7%	41%	38%
Minesotta	43%	24%				7		63%		17%	20%	6	11%	17%	45%
Missisippi	41%	9%										13	31%	31%	33%
Missouri	44%	22%	6	34%	65%	7		58%		6%	35%				
Montana	38%	26%													
Nebraska	29%	24%													
Nevada	37%	26%										5	21%	50%	10%
N. Hampshire	39%	23%	2	46%	50%	2	1%	30%	13%	6%	49%	2	9%	39%	9%
New Jersey	43%	16%													
New Mexico	46%	16%													
New York	50%	16%	6	33%	66%	7		51%		3%	43%				
N. Carolina	43%	14%													
N. Dakota	32%	23%				6		76%		5%	19%	11	8%	24%	40%
Ohio	40%	21%	6	25%	74%	7		58%		4%	37%	11	15%	38%	37%
Oklahoma	34%	23%										11	27%	28%	34%
Oregon	42%	24%													
Pennsylvania	45%	18%													
Rhode Island	47%	23%	6	40%	57%	7		36%		3%	60%				
S. Carolina	40%	12%				4		53%		5%	42%	3	40%	28%	17%
S. Dakota	37%	22%													
Tennessee	47%	10%										11	24%	28%	37%
Texas	37%	22%													
Utah	25%	27%													
Vermont	46%	23%	6	44%	54%	7		34%		3%	60%	11	8%	39%	24%
Virginia	41%	14%				6		53%		3%	44%				
Washington	43%	24%	4	32%	68%	6		58%		2%	39%	10	10%	38%	24%
West Virginia	48%	16%													
Wisconsin	41%	22%										16	6%	44%	37%
Wyoming	34%	26%										9	8%	39%	32%
Candidate Position				1	0		0	0	0	0	1		0	1	0

Notes: Abreviations: round (rd), Brad (Bradley), Frbs (Forbes), Gngr (Gingrich), McC (McCain), Rmn (Romney), Sntrm (Santorum). Empty cells correspond to contests after a candidate's official withdrawal, or following the end of a competitive primary, or otherwise not used in the analysis. Position: 0=insider/experienced, 1=outsider/grass roots/for change (for D), 0=conservative, 1=moderate (for R). See text for details. Sources: Federal Election Commission, George Washington University, CNN, New York Times, Dave Leip's Atlas of US Presidential Elections, USA Today, Wikipedia.

Table A-2. Election Data: The 2004 and 2008 primaries.

	2004 Democratic Primary							2008 Dem. Primary				2008 Republican Primary					
State	rd	Clrk	Dean	Eds	Geph	Kerry	Lieb	rd	Clint	Eds	Obm	rd	Giul	Huck	McC	Rmn	Thm
Alabama								6	42%		56%	7		41%	37%	18%	
Alaska								6	25%		75%	7		22%	15%	44%	
Arizona	3	26%	14%	7%		43%	7%	6	50%		43%	7		9%	47%	35%	
Arkansas								6	70%		26%	7		61%	20%	14%	
California	9			20%		65%		6	52%		43%	7		12%	42%	35%	
Colorado								6	32%		67%	7		13%	18%	60%	
Connecticut	9			24%		58%		6	47%		51%	7		7%	52%	33%	
DC			not applicable: see text					8	24%		76%	9		16%	68%		
Delaware	3	10%	10%	11%		50%	11%	6	43%		53%	7		15%	45%	33%	
Florida								5	50%	14%	33%	5	15%	14%	36%	31%	
Georgia	9			41%		47%		6	31%		67%	7		34%	32%	30%	
Hawai	8			14%		50%		9	24%		76%						
Idaho	8			22%		54%		6	17%		79%						
Illinois								6	33%		65%	7		17%	47%	29%	
Indiana								14	51%		49%						
Iowa	1	0%	18%	32%	11%	38%	0%	1	29%	30%	38%	1	4%	35%	13%	25%	13%
Kansas								6	26%		74%	8		60%	24%		
Kentucky								16	65%		30%						
Louisiana								7	36%		57%	8		43%	42%		
Maine	4	4%	28%	8%		45%		7	40%		59%	6		6%	22%	52%	
Maryland	9			26%		60%		8	36%		61%	9		29%	55%		
Massachu.	9			18%		72%		6	56%		41%	7		4%	41%	51%	
Michigan	4	7%	17%	13%		52%			not applicable			3	3%	16%	30%	39%	4%
Minesotta	9			27%		51%		6	32%		66%	7		20%	22%	41%	
Missisippi								12	37%		61%						
Missouri	3	4%	9%	25%		51%	4%	6	48%		49%	7		32%	33%	29%	
Montana								17	41%		57%	7		15%	22%	38%	
Nebraska								7	32%		68%						
Nevada	6		17%	10%		63%		3	51%	4%	45%	4	4%	8%	13%	51%	8%
N. Hampshire	2	12%	26%	12%		38%	9%	2	39%	17%	37%	2	8%	11%	37%	32%	1%
New Jersey								6	54%		44%	7		8%	55%	28%	
New Mexico	3	21%	16%	11%		42%	3%	6	49%		48%						
New York	9			20%		61%		6	57%		40%	7		11%	52%	28%	
N. Carolina								14	42%		56%						
N. Dakota	3	24%	12%	10%		50%	1%	6	37%		61%	7		20%	23%	36%	
Ohio	9			34%		52%		10	53%		45%	11		31%	60%		
Oklahoma	3	30%	4%	30%		27%	7%	6	55%		31%	7		33%	37%	25%	
Oregon								16	41%		59%						
Pennsylvania								13	55%		45%						
Rhode Island	9			19%		71%		10	58%		40%	11		22%	65%		
S. Carolina	3	7%	5%	45%		30%	2%	4	27%	18%	55%	4	2%	30%	33%	15%	16%
S. Dakota								17	55%		45%						
Tennessee	5	23%	4%	27%		41%		6	54%		41%	7		34%	32%	24%	
Texas								10	51%		48%	11		38%	51%		
Utah	8			30%		55%		6	39%		57%	7		2%	5%	90%	
Vermont	9			6%		32%		10	39%		59%	11		14%	72%		
Virginia	5	9%	7%	27%		52%		8	35%		64%	9		41%	50%		
Washington	4	3%	30%	7%		49%		7	31%		68%	10		24%	50%		
West Virginia								15	67%		26%						
Wisconsin	7		18%	34%		40%		9	41%		58%	10		37%	55%		
Wyoming								11	38%		61%			not applicable: see text			
Candidate Position		1	1	1	0	0	0		0	1	1		1	0	1	0	0

Notes: Abreviations: round (rd), Clark (Clrk), Edwards (Eds), Gephardt (Gep), Lieberman (Lieb), Clinton (Cli), Obama (Obm), Giuliani (Giul), Huckabee (Huck), McCain (McC), Romney (Rmn), Thompson (Thm). Empty cells correspond to contests after a candidate's official withdrawa, or following the end of a competitive primary, or otherwise not used in the analysis. Position: 0=insider, 1=outsider/grass roots/for change (for D), 0=conservative, 1=moderate (for R). See text for details. Sources: Federal Election Commission, ABC, CBS, CNN, MSNBC, PBS, USA Today.