



Candidate competition and voter learning in the 2000–2012 US presidential primaries

George Deltas^{1,2}  · Mattias K. Polborn³

Received: 14 December 2017 / Accepted: 9 October 2018 / Published online: 25 October 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

When candidates in primary elections are ideologically differentiated (e.g., conservatives and moderates in the Republican Party), then candidates with similar positions affect each others' vote shares more strongly than candidates with different ideological positions. We measure this effect in US presidential primaries and show that it is of first-order importance. We also show that voters' beliefs about the candidates harden over the course of the primary, as manifested in the variability of candidate vote shares. We discuss models of sequential voting that cannot yield that pattern of results, and propose an explanation based on a model with horizontally and vertically differentiated candidates and incompletely informed voters. Consistent with the predictions of this model, we also show that, in more conservative states, low-quality conservative candidates do better relative to high-quality conservatives, and vice versa.

Keywords Voting · Primary elections · Simultaneous versus sequential elections

JEL Classification D72 · D60

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11127-018-0620-7>) contains supplementary material, which is available to authorized users.

✉ George Deltas
deltas@illinois.edu

Mattias K. Polborn
mattias.polborn@vanderbilt.edu

¹ Department of Economics, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA

² Lancaster University, Lancaster, UK

³ Department of Economics and Department of Political Science, Vanderbilt University, Nashville, TN 37204, USA

1 Introduction

Candidates for the US presidential election are determined by the results of a sequence of elections within each political party, the primaries, which are managed by the two major parties in collaboration with the states. We address one key feature of these primaries: at the beginning of the process, often more than just two candidates 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 qualities, often called “valences”). For example, Republican primary candidates may be either “moderates” or “conservatives”, and each voter has a preference for one of those positions, which, however, is not absolute: If a voter considers a candidate at 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 receive only imperfect and idiosyncratic signals about the candidates’ valences so that candidates occupying the same policy position may split the votes of voters with preferences 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 that appeal was unsuccessful, withdrew himself.

Such a 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 splits 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 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, we improve on our understanding of learning and inference in all elections.

We measure the extent to which candidate competition is stronger among those who occupy the same political position, compared to competition across political positions. Our analysis uses data from the six contested US presidential primaries that took place between 2000 and 2012,¹ and relies on the observation that, for those years, a dichotomous partition of (serious) candidates into a set of “conservatives” and “moderates” for the Republican Party, and “establishment” and “outsider” candidates for the Democratic Party, does well in capturing the most salient cleavages in each party. As a robustness check, we also analyze partitioning Democrats into “liberals” and “moderates.”

The empirical evidence confirms that, if a candidate drops out, exit benefits the remaining candidates in the drop-out’s position more than it benefits candidates in the opposite position. The effect is quantitatively very large: on average, a candidate will take about

¹ We exclude George W. Bush’s and Barack Obama’s essentially unopposed renominations in 2004 and 2012.

three times as many votes from competing candidates of the same position than he/she will from other candidates. The effect is robust across different specifications and highlights a crucial problem in multi-candidate primaries: candidates who are ideologically close substitutes largely “steal” votes from each other, which ultimately may lead to the nomination of the “wrong” candidate.

We then show that electoral variability declines over time. Variability captures voter learning over time, facilitated by observation of previous election results. That effect is measured without making parametric assumptions, by utilizing the fact that many state contests take place on the same date. We show that the variability of vote shares, controlling for other factors by entering election round fixed effects, declines with the number of states that voted before a particular contest. In other words, if the same set of candidates competes in two groups of states holding elections on two different dates, the within group vote share variance is higher in the states that vote first; when the set of candidates in the second group is smaller, we adjust the vote-share variance 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 any additional information that emerges.

We discuss models with predictions that are not fully consistent with the observed pattern of results. We then propose an explanation based on the model in Deltas et al. (2016). In addition to explaining the observed substitutability and volatility patterns, that model also predicts that an increase of the share of voters who prefer a particular political position leads to a larger 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. We show that that prediction bears out in our data.

Differential competition between candidates is directly relevant to (tactical) coordination between voters. Such coordination, along with learning about candidate quality, has long been known to be an important issue in presidential primaries. For example, Bartels (1987, p. 13) provides a clear description of the coordination process of those Democratic voters unhappy with the establishment candidate in the 1984 Democratic primary.²

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 2015). Those contributions focus on voter learning about valence when voters care only about valence and not about political positions. Our empirical results strongly suggest that ideological differences between candidates matter substantially—voters view some candidates as closer substitutes than others, implying that empirical models that ignore position differences may mistake ideological variation between sequentially voting states for learning about candidate valence.

To our knowledge, the only models of dynamic primaries as contests between more than two candidates are Knight and Schiff (2010), Knight and Hummel (2015) and Deltas et al. (2016). Knight and Schiff (2010) and Knight and Hummel (2015) develop a model of voter learning about candidate quality in which voters in later primary states receive some imperfect information about the signal that voters in earlier states observed. Voters update, taking all pieces of information into account, and vote for their preferred candidate given such information. In their estimation based on the 2004 Democratic presidential primaries, they find that voters attach substantial weight to the outcomes of early elections, but

² See also Bartels (1985, 1988) and Kenny and Rice (1994), which all focus on two-candidate settings.

a much smaller weight after the fourth primary. Thus, in their framework, predicted vote-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 sent by the voting outcome in that state. However, our results suggest that much of the signal is observed directly (as in our model), given that share volatility falls throughout the primary season, and not only after the first few election contests.

Deltas et al. (2016) analyze a structural model of learning about candidate quality in which candidates also are differentiated with respect to their political positions. They estimate their model using the same data as the present paper, but focus on simulating the effects of different institutional setups, such as moving to a simultaneous primary system. The analysis in the current paper focuses on the measurement of the substitution and learning effects; it does not impose the model in Deltas et al. (2016), but shows that the model can explain the empirical findings.³

2 Data

Our dataset consists of information from six of the 2000, 2004, 2008 and 2012 US presidential primaries—we exclude the 2004 Republican and the 2012 Democratic primaries because the incumbent presidents were effectively unopposed. We focus on candidates who are initially considered viable candidates in the sense that some chance exists that they will win their party's nomination. In practice, some of the candidates do not fall into that category because they are too far away from their party's mainstream, and run to represent a particular energized constituency in order to demonstrate that the party needs to pay attention to its preferences. These candidacies are unlikely to be well-captured by any theoretical model, and so we exclude them from our dataset.

The most successful excluded candidates are Dennis Kucinich (Democratic primary 2004 and 2008) and Ron Paul (Republican primary 2008). Their vote shares usually are 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 because of their relatively poor performances, and eventually drop out once it becomes clear that they have no chance of winning the nomination. Tables 6 and 7 in “Appendix” list the candidates we include for each primary, along with the states in which they competed and the vote shares they obtained. The tables also report the number of different election dates (rounds) up to the given elections in each state.

A key component of our empirical analysis (also incorporated in the model in Sect. 4 below) is that candidates from each party are characterized by one of two political positions, representing the main cleavage in the party. The classification of candidates into positions, to which we turn next, is listed in the bottom of Tables 6 and 7 in the Appendix.

³ Among theoretical static models of primaries (i.e., those where only one vote is taken at the primary stage), our paper is most related to Adams and Merrill (2008) and Serra (2011), which point out that holding primaries allows a party to select, on average, higher quality candidates than with a direct nomination of a candidate by party insiders.

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 to use NOMINATE scores based on roll-call votes (see Poole and Rosenthal 1985). However, such scores are available only for legislators, and the majority of candidates has an executive-branch 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 those contests usually are 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 than with those who said they were moderate, and vice versa for John McCain.⁴ For that 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 that 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% of Republican primary voters identify as strongly or somewhat conservative, while only 11% identify as moderates. Romney receives about the same vote share from conservatives and moderates (25% vs. 26%). However, in states like Michigan or Florida where the share of conservatives is around 60%, Romney does substantially better with conservatives than with moderates. Moreover, in the later stages of the campaign, Romney was perceived as fighting with Huckabee for the conservative vote.⁵ For that reason, we classify Romney as conservative. In the 2012 primary, however, Romney was the moderates’ standard-bearer, facing Gingrich and Santorum who were supported by conservatives.⁶ Their splitting of the conservative vote helped him win the nomination. For that year, Romney is classified as a moderate.

Before we turn to the classification of Democratic candidates, we want to note that no a priori reason exists for why party cleavages, or even the dimensions along which parties are *internally* split, should necessarily be the same for both parties. If a party is more ideologically homogeneous than its opponents, we would expect voter ideology to be less predictive for primary elections in the former, and cleavages might then arise along other dimensions. For example, Grossmann and Hopkins (2016) argue convincingly that the Republican Party and the Democratic Party are radically asymmetric organizations: “While the Democratic party is fundamentally a group coalition, the Republican party can be most accurately characterized as a vehicle of an ideological movement.”

⁴ 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.

⁵ See, e.g., <http://www.cnn.com/2008/POLITICS/02/05/super.exit/>.

⁶ The average ideological self-placement of 2012 Romney primary voters in the American National Election Survey on a seven-point scale is 5.13 (5 is “slightly conservative”, 6 is “conservative”). In contrast, the average self-placement of Santorum voters is 5.57, and 5.56 for Gingrich voters. Unfortunately, previous waves of the NES did not ask respondents about their votes in the presidential primary.

For Democrats, we would argue that two logically classification approaches are defensible, and we use both of them. First, in analogy to the Republicans, we could try to classify candidates as “liberal” or “moderate.” Note, however, that for Democrats, the ideological position of the typical voter appears to have much less predictive power than it does for Republicans. In fact, for some states, the differences in candidate support from the liberal and moderate segments seem to disappear entirely. For example, in Nevada, self-declared liberals voted 48-39-9 for Clinton, Obama and Edwards, while moderates voted 46-43-8 for those candidates. That difference between liberals and moderates is well within the margin of error.⁷

We categorize the Democratic candidates along the liberal-moderate axis as follows. In 2000, we categorize Bradley as liberal and Gore as moderate. In 2004, Howard Dean clearly is the standard-bearer of the liberal wing of the Democratic Party. In addition, we classify John Kerry, whom the non-partisan *National Journal* famously selected in 2004 as the most liberal senator, as liberal. In contrast, Clark, Edwards, Gebhardt and Lieberman are classified as moderates. In 2008, we classify Obama as liberal and Clinton as moderate; Edwards likely is located between his competitors, but arguably closer to the liberal end, as his 2008 campaign centered around poverty in the United States (“Two Americas”).

Our preferred classification for the Democratic Party is, however, not the one with respect to ideology because a considerably better sorting of voters is achieved by an exit poll 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 it mostly is concerned with the horse-race 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 primary states, we classify Clinton as insider and Edwards and Obama as outsiders in 2008. In 2004, Kerry receives the largest share from voters who name “experience” as the most important quality,⁸ 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.

For those 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. The vote shares are reported in Tables 6 and 7 in “Appendix”. However, the shares do not sum up to 100% as they include votes for candidates whom we dropped from our analysis, for candidates who have already withdrawn, or for “uncommitted” delegates. We treat such votes as equivalent to abstention from weighing in on the choice of the party nominee. To

⁷ The 2008 Democratic primary is not an aberration in this respect—“liberal” and “moderate” candidates (where this classification is based on roll-call votes or expert judgments of their positions) often receive rather similar percentages of their votes from liberal and moderate voters.

⁸ For example, see <http://www.cnn.com/ELECTION/2004/primaries/pages/epolls/IA/index.html>.

ensure that vote shares representing serious votes sum up to 100% (as also assumed by the model), we rescale all the vote shares accordingly for the purpose of econometric analysis. We supplement the data on the presidential primaries with data from the 1992 presidential election.⁹ The vote shares of the presidential candidates Clinton, Perot and Bush 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 three-way race is expected to be associated with liberal preferences, whereas a high Bush percentage is expected to be associated with conservative preferences. The three-way nature of that presidential election permits an identification of all components of political preferences that are relevant for our analysis.

3 Results

3.1 Non-parametric mean-variance analysis

We start our analysis by pooling all data and comparing the candidates' average vote shares as a function of the distribution of candidates in political positions. In this simple analysis, we do not distinguish between parties, political positions within parties and the position of a state within the primary election sequence, but rather treat symmetrically all primary elections in which κ candidates in one position and κ' candidates in the opposite position compete. The advantage of such an approach is that it is not based on any specification assumptions. That advantage comes at the cost that the analysis in this section 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, and so on). We discuss the limitations in more detail at the end of this subsection, before proceeding to more formal analysis.

Let $\text{VoteShare}_{j,y}^{\kappa,\kappa'}$ be the vote share of candidate j (measured on a 0–100% scale) who occupies his position $a(j)$ with $\kappa - 1$ other candidates, while κ' competitors occupy the opposite position $|1 - a(j)|$. Formally, let

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

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

We report the average $\text{VoteShare}^{\kappa,\kappa'}$ (i.e., the mean over all candidates) in Table 1 for all different candidate configurations that appear in our data. If $\kappa' = 0$ (i.e., all κ candidates are at the same position), then the mean share of a candidate is, by definition, $1/\kappa$. It never happens that all participants in a primary occupy the same political position, and thus those configurations are not listed in Table 1. If $\kappa = \kappa'$, then (again by definition) the mean share of each candidate is equal to $1/2\kappa$. Because such values are not “data” but their values are

⁹ The 1992 general election results were obtained from Dave Leip's Atlas of U.S. Presidential Elections, available at <http://www.uselectionatlas.org/>.

driven by the formula, they are underlined in the table. All other reported values are the realized averages in the data. Most of them are based on more than 10 candidate-share observations, and are reported in bold. When the two classifications yield different values for the mean shares corresponding to a particular configuration of candidates, they are labeled accordingly. In most cases, though, the two classifications give the same figures. The results of Table 1 underpin much of the parametric analysis described in the subsequent sections.

We initially focus on the mean shares corresponding to our first classification. From our discussion of differential substitutability between candidates, we have the following expectations: First, a reduction in the number of candidates at 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, partial, but not complete “crowding out” occurs among candidates at the same position: A reduction in the number of candidates at the same position reduces the total vote share of the candidates at that position because some cross-over voters switch to a candidate at 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 the foregoing 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 at different positions, each attracting 50% 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. Such 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 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%.¹⁰

Holding the total number of candidates fixed, the total vote share of all candidates at a specific position is always increasing in the number of candidates at that position. For example, consider all contests involving five candidates: Here, $4 \times 17.9\% = 71.6\% > 3 \times 19.0\% = 57.0\% > 2 \times 21.5\% = 43.0\% > 28.3\%$. Thus, votes clearly divert from one candidate to another candidate at the same location, but the more candidates that are located at a position, the larger is their combined share. The same pattern holds for contests with three and four candidates.¹¹

The only case that contradicts our expectations is going from a 4-1 constellation to a 3-1 constellation, in which case the average vote share of a candidate at the crowded position falls from 17.9 to 17.3%. That result probably is explained by the small number of cases: only two state elections with a 4-1 constellation and six with a 3-1 constellation

¹⁰ Of course, $\kappa \cdot \text{VoteShare}^{\kappa, \kappa'} + \kappa' \cdot \text{VoteShare}^{\kappa', \kappa} = 100$ holds as an identity. Deviations from that outcome in Table 1, such as here where $3 \times 17.3\% + 48.2\% = 100.1\%$, are explained by rounding.

¹¹ The precise implications of the theory are for expected vote share comparisons between κ candidates at one position and κ' candidates at the other, versus $\kappa - 1$ at one position and κ' at the other. But comparisons between κ candidates at one position and κ' at the other versus $\kappa - 1$ at one position and $\kappa' + 1$ at the other can be obtained by applying our theoretical result iteratively.

Table 1 Key statistics for various candidate configurations

	Mean vote share (%)	Observations	Number of distinct candidates	Mean Obs per candidate
2 Candidates in the election				
1 Candidate in same location	<i>50.0</i>	<i>168</i>	<i>8</i>	<i>21.0</i>
3 Candidates in the election				
2 Candidates in same location	29.1	150	10	15.0
1 Candidate in same location	41.9	75	5	15.0
4 Candidates in the election (insider vs. outsider)				
3 Candidates in same location	17.3	18	6	3.0
2 Candidates in same location	<i>25.0</i>	<i>4</i>	<i>4</i>	<i>1.0</i>
1 Candidate in same location	48.2	6	2	3.0
4 Candidates in the election (extreme vs. moderate)				
3 Candidates in same location	<i>25.0</i>	<i>3</i>	<i>3</i>	<i>1.0</i>
2 Candidates in same location	<i>25.0</i>	<i>24</i>	<i>8</i>	<i>3.0</i>
1 Candidate in same location	<i>25.0</i>	<i>1</i>	<i>1</i>	<i>1.0</i>
5 Candidates in the election				
4 Candidates in same location	17.9	8	4	2.0
3 Candidates in same location	17.6	39	6	6.5
2 Candidates in same location	23.6	26	4	6.5
1 Candidate in same location	28.3	2	1	2.0

The two classifications coincide for elections with 2, 3 or 5 candidates. Italic figures are those that follow mechanically from an algebraic formula. Bold figures are those based on 10 or more candidate share observations (summing up over both political locations in a state contest). See text for a description of these statistics

are observed (indeed, the p value for the difference is 0.93, i.e., no confidence whatsoever can be placed in the sign of the gap). The flip side of that comparison is that the lonely candidate in a 3-1 constellation is doing surprisingly well, getting on average 48% of the vote. The same is true off going from 3-1 to 2-1, which reduces the vote share of the lonely candidate from 48.2 to 42.8% (the difference likewise is not statistically significant with a p value of 0.45). The largest number of observations, and therefore the highest level of confidence in the results (and preponderance of statistically significant pairwise comparisons), obtains for the case of two and three candidates.

The foregoing discussion applies to mean vote shares when Republicans are classified as moderate or conservative and Democrats as insiders or outsiders. When using the extreme versus moderate classification for both parties, the configuration of three candidates at one location and one at the opposing location is based on a single election (four candidate-share observations), in which the lone moderate by sheer coincidence had a 25% vote share.¹² However, only minimal information can be gathered from a single election, making any interpretation unwarranted.

¹² McCain in the 2000 Delaware primary. The three remaining candidates (Bush, Forbes, Keyes) share the remaining 75% and thus have, on average, a 25% vote share as well.

To summarize, the results in Table 1 are indicative of asymmetric candidate substitutability based on their political positions. Vote shares decline with the number of candidates who share a given location, holding the total number of candidates constant. Moreover, the combined vote shares of candidates at a location increases with the number of candidates at that location, holding the total number of candidates constant.

Not taking account of primary election sequencing does not lead to any biases for the questions we address with this analysis. Treating political parties and positions as fungible does not introduce any biases, provided that the political locations do not differ systematically in voter popularity. Our analysis in the next section suggests that that is indeed the case. As will become clear below, even if locations were to differ systematically in voter popularity, no biases would result provided that no systematic differences exist across political positions in the number of candidates at that position. Though that observation essentially is true for the Democrats, it is not true for the Republicans (fewer moderates than conservatives typically enter Republican Party presidential primaries). But given that political positions do not differ much in popularity among the voters, any differences in the politicians' "popularity" would not impact the validity of our results. Overall, the main value of the analysis described here is the absence of any parametric or modeling assumptions, except for those qualitative properties listed in this paragraph.

Since information about the sequence of elections has not been used in the analysis thus far, it cannot provide any evidence regarding the possibility of voter learning. Neither can it assess which of several candidates at the same position benefit most if a state leans more toward those candidates' common position, which can be an indirect test of candidate differentiation. We address such questions in the next two sections by specifying formal econometric models.

3.2 Econometric analysis of vote shares

We now investigate the degree to which candidate vote shares depend on the field of competing candidates, their political positions, and a proxy for each state's preference distribution. We do not impose the structural assumptions of a theoretical model, but rather adopt a reduced-form estimating approach, using progressively more flexible specifications.¹³ The findings are useful when thinking about which type of theoretical model is consistent with the data. The primary benefit of a reduced-form approach is that it remains valid even if a model is somewhat misspecified, and that it allows us to derive a set of separate facts that can help guide theory, rather than test a single model in its entirety.

The results are presented in Table 2. The first five columns of the table correspond to our standard classification, under which Democratic candidates are characterized on the basis of their insider versus outsider status (Republicans are classified on the conservative vs. moderate dimension). The remaining columns of the table correspond to the analogous econometric specifications using the extreme versus moderate classification, under which Democrats are characterized as liberal or moderate and Republicans as conservative

¹³ However, we utilize for the purpose of inference the minimal information that candidate shares in a party's state primary in a given year are negatively correlated (even conditional on characteristics) and that candidate-specific information available at a given time is correlated across states. That is accomplished by using White's (1980) heteroscedasticity consistent standard errors with a two-way clustering at the state primary and candidate/round levels. Doing so tends to be conservative for the purpose of testing (i.e., ignoring clustering reduces standard errors).

Table 2 (continued)

Variable	Insider vs. outsider classification					Extreme vs. moderate classification				
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 1	Model 3	Model 4	Model 5	
OutsiderDem * Perot92%				0.11	- 0.18					
				0.46	0.95					
OutsiderDem * Perot92%*MnShr				- 1.65	0.64					
				0.91	1.73					
InsiderDem * Perot92%*MnShr				- 3.23	- 2.15					
				0.59	1.31					
LiberalDemocrat * Clinton92%								0.17	- 0.56	
								0.37	0.66	
LiberalDemocrat * Clinton92% * MnShr								- 2.84	2.61	
								0.39	1.16	
ModerateDemocrat * Clinton92% * MnShr								- 2.60	2.87	
								0.44	1.23	
Other variables	Constant	Constant	Constant	FEs, '92 shares	FEs, '92 shares	Constant	Constant	FEs, '92 shares	FEs, '92 shares	
OwnLocation-Different Location	- 10.98	- 3.64				- 9.38				
	3.01	4.85				3.15				
OwnLocation-Different Location (Rep)			- 8.61				- 8.61			
			3.33				3.33			
OwnLocation-Different Location (Dem)			- 15.37				- 10.16			
			4.53				5.58			
Party effects (p value)		0.0921	0.2583				0.4976			
Moderate and "Outsider" effects (p value)		0.1942								
Bush/Clinton/Perot effects (p value)				0.0000	0.0094			0.0000	0.0051	
Candidate*election round effects (p value)				0.0000	0.0000			0.0000	0.0000	

Table 2 (continued)

Other variables	Constant	Constant	FEs, '92 shares	FEs, '92 shares	Constant	Constant	FEs, '92 shares	FEs, '92 shares
MeanShare effects (<i>p</i> value)			0.0000	0.3009			0.0000	0.0746
R-squared	0.3043	0.3323	0.3137	0.7889	0.2835	0.2877	0.7944	0.7473

N = 502 (models 1–3); N = 382 (models 4–5). Heteroskedasticity consistent standard errors, clustered by both party/state/year and candidate/round (Models 1–3) and by party/state/year (Models 4–5), are in italics below the parameter and parameter difference estimates. Bold entries indicate significance at the 5% level, underlined entries indicate significance at the 10% level. See text for details

or moderate. We discuss the results of both classifications using the same econometric specification.

Our simplest specification estimates the equation

$$\text{VoteShare}_{j,s,y} = \alpha + \beta_1 \text{CanDif}_{j,s,y} + \beta_2 \text{CanOwn}_{j,s,y} + \epsilon_{j,s,y} \quad (2)$$

where $\text{VoteShare}_{j,s,y}$ is the adjusted vote share of candidate j in state s and year y (measured on a 0–100 scale), and $\text{CanOwn}_{j,s,y}$ and $\text{CanDif}_{j,s,y}$ are the number of candidate j 's competitors with the same or opposite political location, respectively, in state election s in year y . The specification essentially parallels the nonparametric approach in the preceding section, but it uses a statistical framework and thus provides the average effect of adding another candidate at the same or a different political position and the associated standard errors. The findings, reported under Model 1, show that an additional candidate at the same political location as candidate j reduces candidate j 's vote share by three to four times as much as an additional candidate at the opposite location. Each coefficient and their difference is statistically significant.

We next investigate whether the result is affected by the relative popularity of candidates of different political positions. Model 2 suggests that that is not the case. We let Conservative_j and Outsider_j be dummy variables that take the value 1 if candidate j is a conservative Republican or Democratic “outsider” candidate, respectively, and 0 otherwise.¹⁴ We also enter the indicator variable Dem_j for Democratic candidates to complete the set of interactions. In the regression

$$\begin{aligned} \text{VoteShare}_{j,s,y} = & \alpha + \beta_1 \text{CanDif}_{j,s,y} + \beta_2 \text{CanOwn}_{j,s,y} \\ & + \gamma_0 \text{Dem}_j + \gamma_1 \text{Conservative}_j + \gamma_2 \text{Outsider}_j + \epsilon_{j,s,y} \end{aligned} \quad (3)$$

the estimate of β_1 remains smaller than that of β_2 , albeit with a difference that is not statistically significant. However, insufficient information is available to estimate this specification credibly, as manifested by examining the coefficient estimates. Neither political position has a statistically significant effect (see estimates of γ_1 and γ_2), but the coefficient for Democrat is large and statistically significant. However, it is not meaningful (in fact it is logically impossible) to interpret these results as implying that there are no position effects, but that there are party effects. The vote share of Democrats for any given number of candidates must be the same as that of Republicans (both add up to 100%). A statistically significant party dummy makes sense only in the presence of political position effects and accounts for the fact that the distribution of candidates at those positions varies across parties. That observation, along with the fact that the standard errors of the parameters are much larger than in other specifications, indicates that insufficient variation exists in our data to meaningfully estimate the model. Similar conclusions are obtained under the extreme versus moderate classification, which is omitted for brevity.

However, a slightly more parsimonious specification (Model 3), which “cuts the data” in a different way, can be estimated and yields meaningful estimates. In the new specification, we investigate whether the relevance of candidates’ political locations is present in both major parties, or is confined to one of them. We do so by estimating

$$\begin{aligned} \text{VoteShare}_{j,s,y} = & \alpha + \beta_{1R} \text{CanDif}_{j,s,y} \text{Rep}_j + \beta_{2R} \text{CanOwn}_{j,s,y} \text{Rep}_j \\ & + \beta_{1D} \text{CanDif}_{j,s,y} \text{Dem}_j + \beta_{2D} \text{CanOwn}_{j,s,y} \text{Dem}_j + \epsilon_{j,s,y} \end{aligned} \quad (4)$$

¹⁴ Note that Conservative_j is zero for all Democrats and Outsider_j is zero for all Republicans, i.e., those variables include an implicit interaction with the party dummy.

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 suggest that voter segmentation across political locations might be more pronounced for the Democratic Party, whereby a candidate's vote share is affected only negligibly by competing against one fewer candidate at the opposing political location, but is very strongly affected by one fewer candidate at the same political location. The relative effect of the location of competing candidates also is statistically significant for the Republican primaries, but smaller in quantitative terms.¹⁵ However, the difference between the Democrats and the Republicans is smaller when using the extreme versus moderate classification for both parties. One could add the Democratic dummy to the specification in Eq. (4), allowing for different intercepts for the two parties. When doing so, the parameter estimate for the dummy is not statistically significant, and the point estimates for other parameters are not materially affected.

A consequence of candidate differentiation is that as the electorate's policy preferences shift towards one political position, the candidates that share that political position are expected to benefit. In fact, it would be reasonable to expect that, among candidates sharing a position, the ones with higher valence would obtain a greater number of votes as the electorate shifts towards that position (compared to candidates with lower valence). But it also would be reasonable to expect that the vote shares of the lower valence candidates would experience a larger *proportionate* increase, because a weak candidate's voters include a disproportionately small number of people with opposing policy preferences. Those conjectures are formalized in Sect. 4. Our final set of regressions are intended to ascertain whether they also are supported by the data, thus buttressing the support of our framework over the alternatives. The conjectures are hard to test because they demand much from our limited data (we observe only proxies for voter preferences), and also because they require an 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 also should allow expected candidate shares to vary across rounds.

We first need to operationalize and test our measures of electorate political preferences, before we use them to ascertain how they affect the relative vote shares of strong and weak candidates. 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 proxies for the distribution of political preferences in a state. Instead, our proxies for electorate preferences are based on the outcome of the 1992 presidential election between Bush, Clinton and Perot. Voter preferences in states in which Bush did well plausibly are shifted to the right relative to the rest of the country, and we would therefore likewise expect conservative Republicans to do better in those states than moderates. Similarly, voter preferences in states in which Clinton did well plausibly are shifted to the left relative

¹⁵ Adding location dummies to this regression does not materially affect the estimates of the Democratic coefficients, but reduces the Republican location effect to essentially zero. However, with Moderate_{*i*} being essentially a dummy for McCain, the Republican effect would be identified solely from the gain of voters by McCain as other candidates (of opposing location) depart, relative to the gain of voters by his opponents as other candidates (at same location) depart. Not only is the effective information for this specification even more limited (only four such withdrawals), but with McCain being a higher quality candidate, the location and valence effects are confounded (McCain gets a larger than expected share of the departing candidates' voters because he is a better candidate in the vertical dimension).

to the rest of the country, and we would therefore also expect liberal Democrats to do better in these states than moderates. Finally, states in which Perot did well likely have a larger than average share of populist voters, so that we expect candidates classified as outsiders to do better.

To test the validity of the foregoing propositions,—a prerequisite for using 1992 vote shares to investigate the effect of electorate preferences on candidates of difference valence—we estimate the equations

$$\begin{aligned} \text{VoteShare}_{j,s,y} = & \alpha_{j,t,y} + \gamma_{0B} \text{Rep}_j \text{Bush}92\%_s + \gamma_{0P} \text{Dem}_j \text{Perot}92\%_s \\ & + \gamma_{1B} \text{Conservative}_j \text{Bush}92\%_s + \gamma_{1P} \text{Outsider}_j \text{Perot}92\%_s + \epsilon_{j,s,y} \end{aligned} \quad (5)$$

and

$$\begin{aligned} \text{VoteShare}_{j,s,y} = & \alpha_{j,t,y} + \gamma_{0B} \text{Rep}_j \text{Bush}92\%_s + \gamma_{0C} \text{Dem}_j \text{Clinton}92\%_s \\ & + \gamma_{1B} \text{Conservative}_j \text{Bush}92\%_s + \gamma_{1C} \text{Liberal}_j \text{Clinton}92\%_s + \epsilon_{j,s,y} \end{aligned} \quad (6)$$

where $\text{Bush}92\%_s$, $\text{Perot}92\%_s$ and $\text{Clinton}92\%_s$ are Bush's, Perot's and Clinton's vote shares in state s in the 1992 Presidential election, respectively,¹⁶ and $\alpha_{j,t,y}$ are *candidate* \times *year* \times *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. Those dummies would predict the share of a candidate perfectly for election days in which only a single state votes, completely eliminating their influence on the model's remaining parameters. Thus, we drop observations that consist of a single state contest from the regressions in Eqs. (5) and (6), reducing the number of observations from 502 to 382 (the same also is done in all subsequent specifications that use an exhaustive set of candidate \times year \times round effects). The flexibility embodied in the entering of the dummies 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). Note that the regressions also include the 1992 vote shares interacted with the party for which they serve as ideological proxies.¹⁷

As explained above, the expected Bush '92 effect is an increase in the vote share of conservative Republicans. That is indeed the case, with a one percentage increase in the vote share of Bush increasing the vote share of conservative Republicans in the state primary by 1.47% (the effect is statistically significant).¹⁸ The Perot effect on Democratic outsider candidates is positive, with a significance that barely misses the 5% cut-off. The point estimate also is smaller, with a one percentage increase in Perot's vote share raising the vote share of outsider Democrats in primaries by 1.10%. The fact that the evidence is not as strong as for the Republicans may be a consequence of many (if not most) of Perot's voters being fiscally conservative populists, so that their influence on the Democratic primary electorate is weaker. Finally, Clinton effectively had no effect on liberal Democrats. That result casts

¹⁶ The vote share variables $\text{Bush}92\%_s$, $\text{Clinton}92\%_s$ and $\text{Perot}92\%_s$, like $\text{VoteShare}_{j,s,y}$, range from 0 to 100.

¹⁷ The use of exhaustive candidate \times year \times round effects eliminates the need to add party dummies in these regressions.

¹⁸ The Republican estimates are identical in the two regressions, as they should be since the exhaustive set of fixed effects essentially results in two distinct equations for each of the two parties. The results of Eqs. (5) and (6) are omitted from Table 2 to conserve space, but are available upon request.

doubt on using the 1992 election results to identify Democratic electoral preferences along the liberal versus moderate dimension. We nonetheless also report the results below for completeness.

We next turn to our proxies for candidate valence. Our first proxy for valence in round t is the average vote share of a candidate in that round, $MeanShr_{j,t,y}$. Clearly, it is an imperfect measure, but a reasonable one. Candidates with high relative valence, as perceived in round t , will have higher values of $MeanShr_{j,t,y}$. The number and distribution of competing candidates also will affect the values of $MeanShr_{j,t,y}$. To reduce such candidate composition effects on this valence, we enter a measure in regressions that also include candidate \times round effects, as in regressions 5 and 6.¹⁹ Moreover, averaging the vote shares for all contests in a round is meaningful because all states have the same ex ante expectations about valence that they update independently on the basis of their privately observed signals. In addition, 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 a candidate in a particular state would lead 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 (albeit not a large one when many states are holding their primaries at the same time).

A specification (Model 4) that does not suffer from such an endogeneity short-coming is

$$\begin{aligned} \text{VoteShare}_{j,s,y} = & \alpha_{j,t,y} + \gamma_{0B} \text{Rep}_j \text{Bush}92\%_s + \gamma_{0P} \text{Dem}_j \text{Perot}92\%_s \\ & + \gamma_{1B} \text{Conservative}_j \text{Bush}92\%_s + \gamma_{1P} \text{Outsider}_j \text{Perot}92\%_s \\ & + \{ \gamma_{2Bc} \text{Conservative}_j + \gamma_{2Bm} \text{ModerateRep}_j \} \text{Bush}92\%_s \text{MeanShr}_{j,t/s,y} \\ & + \{ \gamma_{2Po} \text{Outsider}_j + \gamma_{2Pi} \text{Insider}_j \} \text{Perot}92\%_s \text{MeanShr}_{j,t/s,y} + \epsilon_{j,s,y} \end{aligned} \quad (7)$$

for the outsider-insider classification and

$$\begin{aligned} \text{VoteShare}_{j,s,y} = & \alpha_{j,t,y} + \gamma_{0B} \text{Rep}_j \text{Bush}92\%_s + \gamma_{0C} \text{Dem}_j \text{Clinton}92\%_s \\ & + \gamma_{1B} \text{Conservative}_j \text{Bush}92\%_s + \gamma_{1C} \text{Liberal}_j \text{Clinton}92\%_s \\ & + \{ \gamma_{2Bc} \text{Conservative}_j + \gamma_{2Bm} \text{ModerateRep}_j \} \text{Bush}92\%_s \text{MeanShr}_{j,t/s,y} \\ & + \{ \gamma_{2Cl} \text{Liberal}_j + \gamma_{2Cm} \text{ModerateDem}_j \} \text{Clinton}92\%_s \text{MeanShr}_{j,t/s,y} + \epsilon_{j,s,y} \end{aligned} \quad (8)$$

for the extreme-moderate classification, 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 , $ModerateRep_j$ takes the value of 1 for moderate Republicans and zero for all others, and $ModerateDem_j$ takes the value of 1 for moderate Democrats and zero for all others.²⁰ This specification, too, however, raises a potential endogeneity concern, though one

¹⁹ The theoretical model shows that the effect of changes in electorate preferences on candidate vote 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 positions. The variable $MeanShr_{j,t,y}$ also adjusts for the number of competing candidates, their valence and political positions and, thus, in a qualitative way, reflects the factors that enter in the comparative statics of the theoretical model.

²⁰ Note that, by necessity, this specification uses a different proxy for every state, since the variable $MeanShr$ no longer takes the same value for all states in a given round.

of ambiguous sign (and possibly of zero magnitude).²¹ A more conservative approach is to lag the $MeanShr_{j,t,y}$ variable by one round, i.e., use it as a proxy of valence $MeanShr_{j,t-1,y}$. This yields Model 5 below

$$\begin{aligned}
 VoteShare_{j,s,y} = & \alpha_{j,t,y} + \gamma_{0B}Rep_jBush92\%_s + \gamma_{0P}Dem_jPerot92\%_s \\
 & + \gamma_{1B}Conservative_jBush92\%_s + \gamma_{1P}Outsider_jPerot92\%_s \\
 & + \{\gamma_{2Bc}Conservative_j + \gamma_{2Bm}ModerateRep_j\}Bush92\%_s MeanShr_{j,t-1,y} \\
 & + \{\gamma_{2Po}Outsider_j + \gamma_{2Pi}Insider_j\}Perot92\%_s MeanShr_{j,t-1,y} + \epsilon_{j,s,y}
 \end{aligned}
 \tag{9}$$

for the insider-outsider classification, and a similar expression for the extreme-moderate classification. Such a specification is not necessarily better than the one in (7) for two reasons. First, lagging the mean vote 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 no longer is guaranteed to be the same across round t and $t - 1$, and therefore introduces an additional source of noise in the valence proxy.

We therefore estimate Model 4 and Model 5, present both sets of results in Table 2, and discuss them together. The variable $MeanShr$ is measured as fraction of the votes received by a candidate (not the percentage), so that all parameter estimates are of the same scale; we, of course, interpret the results accordingly. For both parties, the parameter estimates generally are more precise for the first specification. To a large degree, that is because of larger standard errors for the second specification, indicative of the weaker proxy effect of using lagged values.

The results for the Republicans do not depend on the classification system, a direct consequence of using a specification wherein no parameters are estimated from information from both parties. Conservative Republicans do better than moderates in states where Bush obtained larger vote shares in 1992, and the effect is stronger for candidates of higher valence. From parameter values in columns 4 and 8 of Table 2, the effect of a 1% increase in Bush’s vote share on the vote share difference between a conservative and a moderate is equal to 0.59%, when evaluated for candidates that get a zero fraction of the votes, i.e., at the “boundary.” The gap increases for candidates with positive vote shares, since the interaction with the $MeanShr$ variable is smaller (algebraically) for moderate than for conservative Republicans (− 3.13 vs. − 2.98). For candidates with $MeanShr$ of, say, 0.4, a 1% increase in Bush’s 1992 vote share lifts conservatives by $0.59 + 0.4 * 0.15 = 0.65\%$ relative to moderates. Thus, the gap is larger for high valence candidates, as expected, but does not increase proportionately (again, as expected).

The negative estimate for γ_{2Bm} indicates that, in states where Bush obtained more votes in 1992, moderates with high vote shares experience larger vote reductions than moderates with small vote shares. However, the negative estimate for γ_{2Bc} indicates that in those states, conservatives of high valence were helped to lesser extents than conservatives of

²¹ To see this, suppose that only two states, 1 and 2, hold primaries in a given round t , and that the mean share enters directly as a regressor (rather than as an interaction). Then, the share of candidate j in state 1 is $VS_{j,1} = BX_{j,1} + \gamma VS_{j,2} + \epsilon_{j,1}$, where BX contains all other regressors and the year subscript is suppressed. The vote share for state 2 is given analogously. Solving the two-by-two system for $VS_{j,1}$ and $VS_{j,2}$ yields $VS_{j,1} = \frac{BX_{j,1}}{1-\gamma} + \frac{\epsilon_{j,1} + \gamma\epsilon_{j,2}}{1-\gamma^2}$ and $VS_{j,2} = \frac{BX_{j,2}}{1-\gamma} + \frac{\epsilon_{j,2} + \gamma\epsilon_{j,1}}{1-\gamma^2}$. It can be seen from the reduced-form expressions for the vote shares that the share in state 2 is positively correlated with the structural error of the equation for the vote share in state 1 ($\gamma < 1$).

lower valence. The last finding is somewhat surprising, perhaps indicating that conservative Republicans with high vote shares are somehow less extreme than the average conservative candidate.²²

For the Democratic candidates, the two classification systems yield qualitatively similar results, but differ somewhat quantitatively. Outsiders do better than insiders in states where Perot's vote share was large and that gap is increasing with candidate valence ($\gamma_{2Po} > \gamma_{2Pi}$). The effect for candidates who obtain 0.4 of the vote share is $0.11 + 0.4 * (3.23 - 1.65) = 0.74$. However, with the Perot effect being nearly zero at the "origin" (only 0.11), the effect of electorate preferences on the outsider versus insider gap is nearly proportional to candidate valence. Insider Democrats with large average vote shares perform worse relative to those with smaller vote shares in states that exhibited stronger support for Perot. But so do, to a smaller and far less statistically significant degree, outsider Democrats with large average vote shares; that result largely mirrors the findings for Republican candidates.

Moving to the liberal versus moderate classification for the Democrats, we observe that the association between Clinton's 1992 vote share and the relative vote share of liberal and moderate primary candidates essentially is zero. For both classification systems, liberal as well as outsider candidates do better in states with larger 1992 Democratic vote shares. Recall, however, that the estimates of regression 6 cast doubt on the use of Clinton's 1992 vote share as a suitable proxy for the ideological preferences of the Democratic electorate.

3.3 Econometric analysis of vote-share variability

We now analyze how vote share variability evolves over the course of the primaries, and show that its evolution is best explained by voter learning. 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 such uncertainty is resolved slowly over time, we posit that vote share variability declines over time. Moreover, since additional information moves perceptions (and thus vote shares) by progressively smaller amounts, the largest decline in variability should happen early in the primary season and be related to the level of information received (proxied by the number of states that have voted already) rather than to the simple passage of time.

Estimates of vote-share variability necessarily have to be based on an analysis of the disturbance variance of equations of the form estimated in the preceding section. The disturbance variance can be estimated from the post-estimation regression residuals or jointly with the other equation parameters by Maximum Likelihood (GLS). The former method is heteroscedasticity consistent and robust (does not depend on the specification of the variance process); the latter method is efficient under the correct specification of the variance, but inconsistent if the variance process is mis-specified. We adopt the robust approach here, and consider the Maximum Likelihood estimates of all parameters in the next section

²² The conclusions about relative effects are based on the parameters reported in Table 2. To obtain the level effect, one needs also to incorporate the effect of the Bush 1992 percentage on Republican vote shares, which is 1.38 (not reported in Table 2 to save space). Evaluated at a value of MeanShr = 0.4, a 1% increase in Bush's 1992 vote share lifts the vote share of a conservative Republican by 0.77% and reduces the vote share of the typical moderate Republican by 0.47% (the two figures are not the same because the number and typical vote shares of conservatives and moderates are not equal).

as a robustness exercise. Part of the reason for that approach is that, as will be apparent later in this section, some of the variance analysis involves aggregation of residuals which is not feasible under the GLS approach.

In choosing which equations to estimate to obtain the residuals, 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 from 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 well as possible. Our base model for obtaining the residuals includes an exhaustive set of candidate \times round \times 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 positions and numbers) in the most flexible way: with indicator variables whose coefficients vary (with no parametric constraints) over time. This regression is equivalent to Models 4 and 5 without the Bush-Perot or Bush-Clinton effects, does not rely on our classification of candidates into political locations or on any of the other aspects of our specification that involve candidate competition.

We also estimate vote share variability using the residuals of the more heavily parameterized Models 4 and 5. By their very nature, the results here would differ somewhat for each specific parametrization of the Clinton and Perot effects. Since we focus here on the time variation of the residuals, we report as a representative model (the results based on Model 4), which is one of the two most flexible specifications and uses the same valence proxy variable for all elections in a given round.²³

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

$$|\hat{\epsilon}_{j,s,y}| = a + b NumCand_{j,s,y} + c PriorSignals_{j,s,y} + u_{j,s,y} \tag{10}$$

and

$$|\hat{\epsilon}_{j,s,y}| = a + b NumCand_{j,s,y} + c PriorSignals_{j,s,y} + d PriorSignals_{j,s,y}^2 + u_{j,s,y} \tag{11}$$

where $|\hat{\epsilon}_{j,s,y}|$ is the residual from either Model 4, or from Model 4/5 without the Bush-Perot or Bush-Clinton interaction terms. In the former case, the results depend on the classification system we use; in the latter case, they do not. The number of candidates is entered as an explanatory variable in the regression because more candidates mean smaller vote shares (on average), and smaller vote shares exhibit smaller variances. We also re-estimate regressions (10) and (11) making a small sample adjustment for residuals that accounts 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 .²⁴

That procedure yields a total of 12 regressions, whose results are reported in Table 3. Consistent with our expectations, residual variance declines with the number of prior

²³ The residuals of the Model 5 give similar results.

²⁴ The adjustment is exact when no covariates are entered.

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

	Candidate*round FEs		Model 4: insider vs. outsider classification		Model 4: Extreme vs. moderate classification	
	Using the raw value of the residual	Adjusting for the number of candidates	Using the raw value of the residual	Adjusting for the number of candidates	Using the raw value of the residual	Adjusting for the number of candidates
NumberOfCandidates	-1.142	-0.749	-1.265	-1.323	-1.538	-1.370
PriorSignals	0.460	0.525	0.411	0.439	0.403	0.413
PriorSignalsSquared	-0.133	-0.077	-0.204	-0.276	-0.242	-0.232
Constant	0.043	0.049	0.035	0.127	0.034	0.036
Signal effects (p value)	0.004	0.049	0.002	0.002	0.000	0.000
R-squared	0.0351	0.0404	0.0846	0.0855	0.1236	0.1063
Observations	382	382	382	382	382	382

The dependent variable is the absolute value of the regression residual on candidate-cross-round fixed effects, or of Model 4. Adjustment 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/year, 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

contests for all specifications (an effect that is statistically significant in all but one of the regressions). 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 that expectation, we find that the coefficient of $PriorSignals_{j,s,y}^2$ is positive in all specifications. However, it generally is not statistically significant, largely because the square term is highly correlated with the linear term, as evidenced by the quadrupling of the standard errors of the latter (note that the p value for the joint test of both parameters always is statistically significant). Finally, the number of candidates has a negative effect on variance, as expected, though the effect is smaller when we adjust the dependent variable for the number of candidates.

Even though the variability reduction effect associated with 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 7% of the residual variance, on average (though the percentage can be over 12 % in some specifications). Evidently, several other determinants of vote share variability are possible, including the type of information shocks that lead to learning about candidate valence in the first place, and possible coordination of voters across states voting simultaneously.²⁵

The second of those 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 primary states can coordinate on a candidate occupying a particular political position (perhaps through local press coverage), but cannot coordinate across states. In that scenario, a candidate may obtain many votes in one state (if voters coordinate on him), but very few in another state that votes in the same round (if voters there coordinate on his opponent). Thus, candidate share variability would be relatively large in early states. Later, coordination across states increases, as voters observe who is likely to emerge as the most competitive candidate at a particular political position. Such an 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.

To test whether that alternative explanation is the driving force behind the reduction in share variability, note that it implies that vote share residuals for candidates at the same political position should be correlated strongly negatively and largely cancel out. Vote share variability at the political position level, controlling for candidates' mean vote shares, should not have a clear trend over time. We test that prediction by summing the vote share residuals of candidates at the same political position in a particular state contest. We then perform the same analysis described in Eqs. (10) and (11) using the aggregated residuals of Model 4. Note that the right-hand-side variables take the same values for candidates competing in the same state contest, so that these regressions differ only in the construction of the dependent variable (and in the number of observations). The estimates are reported in Panel A of Table 4. The pattern of coefficient estimates is unchanged: share variability, measured at the position level, declines for later contests. Statistical significance is affected when both the number of signals and the number of signals squared are entered as regressors; however, the two variables remain jointly statistically significant. The explanatory power of those regressions actually increases somewhat compared to those in Table 3. We conclude that tighter coordination of voters across states voting contemporaneously is not an explanation for the reduction in share variability.

²⁵ It is not surprising, and in fact reassuring, that when one includes the variable $PriorSignals_{j,s,y}$ in the vote share regressions it turns out to be insignificant.

Another observation supports our interpretation that the reduction in variability is explained by hardening priors as more information about the candidates becomes available. That conclusion is based on our use of a simple counter of the election round in (10) and (11), i.e., a variable that is akin to a linear time trend and does not take into consideration the number of states that vote in a given round. The election round is of course correlated with the number of signals, so when it is included in the variance regression on its own, the parameter estimate is negative.²⁶ However, when adding it in conjunction with $PriorSignals_{j,s,y}$, the coefficient on election round becomes statistically insignificant. The lack of significance is not driven by the correlation between the two variables, since the coefficient of $PriorSignals_{j,s,y}$ remains negative and statistically significant. In fact, the round coefficient is positive when entered in conjunction with $PriorSignals_{j,s,y}$. Those findings are reported in Panel B 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.

3.4 Joint analysis of vote-share variability and electorate ideology effects

The analysis in the preceding section provides evidence of a systematic relationship between candidates' vote-share variances and the timing of a state's election in the overall sequence of primaries. The empirical analysis in Sect. 3.2 is robust to any relationship between the variance of vote shares and other variables, and also accounts for the possibility of correlation in those disturbances across observations. Imposing, however, a specific skedastic function in the estimation process would improve the efficiency of the estimates, if the imposed skedastic function is indeed correct. To this end, we have re-estimated the models discussed in Sect. 3.3 by iterative Generalized Least Squares, obtaining the Maximum Likelihood parameter estimates for both the vote share equation and the skedastic function jointly.²⁷ In all estimated models, the skedastic function includes the adjustment for the number of candidates. Following the convergence of the parameter estimates, standard errors that account for clustering are computed and used for inference.

Results corresponding to Model 4 under both classifications are reported in Table 5. The first column is the counterpart of column 4 of Table 2 (for vote shares) and column 8 of Table 3 (for the skedastic function, i.e., vote-share variability). Except for the interaction of outsider Democrat mean shares with the Perot vote share, in all other cases the GLS estimates are within one standard error of the original estimates. In the only exception, the parameter estimate is negative and statistically significant in both cases. Although the coefficient of $PriorSignals_{j,s,y}$ in the skedastic function does not differ much between the two sets of estimates, it is no longer statistically significant with GLS. However, the joint-test of the linear and quadratic values is statistically significant. The second column of Table 5 uses the same vote share equation, but replaces the skedastic function with that of column 2 of Panel B of Table 4. Minimal difference exists between the GLS and OLS skedastic function estimates.

²⁶ In fact, it is larger in absolute value than those reported in Table 3 because on average multiple state contests occur in a single round.

²⁷ It is not appropriate to do so for the less parametrized models 1, 2, and 3 because the skedastic function almost surely also depends on other factors that dominate learning effects. However, we did verify that the conclusions of Table 2 are not affected when those models were estimated with GLS.

Table 4 Analysis of vote share variability: considering alternative explanations

	Model 4: insider vs. outsider classification				Model 4: extreme vs. moderate classification			
	Using the raw value of the residual		Adjusting for the number of candidates		Using the raw value of the residual		Adjusting for the number of candidates	
<i>Panel A: aggregate residuals by political position</i>								
NumberOfCandidates	0.054	0.024	0.473	0.435	-0.187	-0.197	0.127	0.126
	<i>0.609</i>	<i>0.617</i>	<i>0.641</i>	<i>0.647</i>	<i>0.610</i>	<i>0.623</i>	<i>0.638</i>	<i>0.651</i>
PriorSignals	-0.188	<u>-0.252</u>	-0.163	<u>-0.244</u>	-0.228	<u>-0.249</u>	-0.217	-0.218
	<i>0.035</i>	<i>0.139</i>	<i>0.037</i>	<i>0.146</i>	<i>0.034</i>	<i>0.129</i>	<i>0.035</i>	<i>0.137</i>
PriorSignalsSquared		0.002		0.002		0.000		0.000
		<i>0.003</i>		<i>0.003</i>		<i>0.003</i>		<i>0.003</i>
Constant	10.000	10.468	9.163	9.747	11.158	11.307	10.729	10.739
	<i>2.064</i>	<i>2.360</i>	<i>2.144</i>	<i>2.442</i>	<i>2.047</i>	<i>2.394</i>	<i>2.123</i>	<i>2.473</i>
Signal effects (<i>p</i> value)		0.0000		0.0000		0.0000		0.0000
R-squared	0.1107	0.1116	0.0917	0.0929	0.1531	0.1532	0.1386	0.1386
Observations	290	290	290	290	290	290	290	290
<i>Panel B: calendar time effects</i>								
NumberOfCandidates	<u>-0.738</u>	-1.121	-0.574	-0.908	-0.977	-1.408	-0.865	-1.266
	<i>0.445</i>	<i>0.442</i>	<i>0.455</i>	<i>0.454</i>	<i>0.431</i>	<i>0.430</i>	<i>0.441</i>	<i>0.441</i>
PriorSignals		-0.259		-0.227		-0.292		-0.272
		<i>0.052</i>		<i>0.055</i>		<i>0.052</i>		<i>0.056</i>
Round	-0.386	0.291	-0.356	0.237	-0.499	0.263	-0.500	0.211
	<i>0.143</i>	<i>0.221</i>	<i>0.151</i>	<i>0.235</i>	<i>0.144</i>	<i>0.216</i>	<i>0.150</i>	<i>0.229</i>
Constant	11.895	11.233	11.747	11.169	13.354	12.611	13.549	12.855
	<i>2.262</i>	<i>2.290</i>	<i>2.342</i>	<i>2.390</i>	<i>2.220</i>	<i>2.214</i>	<i>2.291</i>	<i>2.303</i>
R-squared	0.0225	0.0909	0.0176	0.0666	0.0391	0.1290	0.0362	0.1095
Observations	382	382	382	382	382	382	382	382

The dependent variable for Panel A is the absolute value of the sum of the residuals of all candidates in the same political position (in a given state contest); for Panel B it is the absolute value of the regression residual of Model 4. Adjustment 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/year, 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

Columns 3 and 4 of Table 5 are the counterparts of the first two columns, but using the moderate versus extreme classification of candidates. The difference in the estimates of the vote share equation again are within a standard error of the original values. The only exception is the interaction of the liberal Democrat indicator with Clinton's 1992 vote share: both estimates are of the same sign, but only that obtained with GLS is statistically significant. The skedastic function estimates likewise are similar in sign and generally more statistically significant with GLS. Finally, we observe that changing the skedastic function has a relatively small effect on the estimates of the mean vote-share equation. However, because a different skedastic function applies different weights to each observation, the estimates

Table 5 Analysis of vote shares and share variability-GLS

	Model 4			
	Insider vs. outsider		Moderate vs. extreme	
<i>Panel A: expected votes shares</i>				
ConservRepublican * Bush92%	<i>0.65</i>	0.57	0.64	0.54
	<u>0.33</u>	<i>0.27</i>	<i>0.31</i>	<i>0.24</i>
ConservRepublican * Bush92% * MnShr	– 3.00	– 2.93	– 2.96	– 2.89
	0.38	0.34	0.35	0.33
ModerateRepublican * Bush92% * MnShr	– 2.65	– 2.57	– 2.62	– 2.58
	0.39	0.34	0.37	0.35
OutsiderDem * Perot92%	0.19	0.37		
	0.22	0.26		
OutsiderDem * Perot92% * MnShr	– 4.77	– 4.36		
	0.16	0.24		
InsiderDem * Perot92% * MnShr	– 4.62	– 4.25		
	0.21	0.23		
LiberalDemocrat * Clinton92%			0.51	0.39
			<i>0.13</i>	<i>0.07</i>
LiberalDemocrat * Clinton92% * MnShr			– 2.36	– 2.35
			0.01	0.01
ModerateDemocrat * Clinton92% * MnShr			– 2.37	– 2.35
			0.01	0.01
Constant(s)	FES, '92 shares	FES, '92 shares	FES, '92 shares	FES, '92 shares
Bush and Perot/Clinton effects (<i>p</i> value)	0.0000	0.0000	0.0000	0.0000
Candidate*election round effects (<i>p</i> value)	0.0000	0.0000	0.0000	0.0000
MeanShare effects (<i>p</i> value)	0.0000	0.0000	0.0000	0.0000
<i>Panel B: skedastic function</i>				
NumberOfCandidates	– 0.890	– 0.766	– 1.599	– 1.267
	<u>0.526</u>	0.501	0.470	0.453
PriorSignals	– 0.188	– 0.304	– 0.347	– 0.358
	0.138	0.055	0.125	0.053
PriorSignalsSquared	– 0.001		0.002	
	0.003		0.002	
Round		0.324		<u>0.417</u>
		0.229		<u>0.222</u>
Constant	12.781	11.084	15.815	12.374
	2.461	2.453	2.344	2.340
Signal effects (<i>p</i> value)	0.0000		0.0000	
Observations	382	382	382	382

Skedastic function accounts for the number of candidates; residual scaled 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/year, 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

for the Republican candidates are no longer identical across the two classifications, as they are with the robust OLS analysis.

4 Explaining the results: formal analysis of a framework

In this section, we will argue that the model developed in Deltas et al. (2016) is consistent with the set of results presented in the preceding section. For convenience, we summarize the key features of this model here, before proceeding to formal analysis. Let $J = \{1, \dots, J\}$ denote the set of candidates who compete for their party’s nomination. The set of states is S , with typical state s . States vote sequentially, although some may vote at the same time. Voters observe the outcome in all states that voted before their own. 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, a binary characteristic fixes candidates exogenously either to position 0 or to position 1, as in Krasa and Polborn (2010). One can think of the binary characteristic as an ideological position. This setup is the simplest one for formalizing the idea that some candidates are close policy substitutes for voters, while a substantial difference exists with some other candidates. The first j_0 candidates are fixed at $a_j = 0$, while the other $j_1 = J - j_0$ candidates are fixed at $a_j = 1$.

Voter i ’s utility from victory by candidate j is

$$U_j^i = v_j - \lambda|a_j - \theta^i| + \varepsilon_j^i \tag{12}$$

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 with respect to candidate j , reflecting variation among candidates in other dimensions on which voters’ preferences differ. In that case, the fixed characteristic, modeled explicitly ($a_j = 0$ or $a_j = 1$), should be understood as the most important policy dimension. Without loss of generality, we normalize $\sigma_\varepsilon = 1$.

Voters are uncertain about the candidates’ valences, which are 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 state-specific.²⁸

Given their own signal, and possibly the election results in earlier primary states from which the signals in those 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. That is, voter i in district s who votes at time t votes for candidate j if and only if

$$j \in \arg \max_{j' \in J'} \hat{v}_{j'}^s - \lambda|a_{j'} - \theta^i| + \varepsilon_j^i, \tag{13}$$

²⁸ 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. To simplify the model and gain some tractability, we ignore nationally observed errors, although they are accounted for in the estimation.

where \mathcal{J}^t is the set of candidates in period t elections.²⁹

Deltas et al. (2016) show that, given that the beliefs of voters in state s are given by the vector $\hat{v}^s = (\hat{v}_1^s, \hat{v}_2^s, \dots, \hat{v}_j^s)$, the total number of votes for candidate $j \in J_0^s$ is

$$\begin{aligned}
 & (1 - \mu^s) \int_{-\infty}^{\infty} \prod_{J_0^s \setminus \{j\}} \Phi(\hat{v}_j^s - \hat{v}_{j'}^s + \varepsilon_j) \cdot \prod_{J_1} \Phi(\lambda + \varepsilon_j + \hat{v}_j^s - \hat{v}_{j'}^s) \cdot \phi(\varepsilon_j) d\varepsilon_j \\
 & + \mu^s \int_{-\infty}^{\infty} \prod_{J_0^s \setminus \{j\}} \Phi(\hat{v}_j^s - \hat{v}_{j'}^s + \varepsilon_j) \cdot \prod_{J_1^s} \Phi(-\lambda + \varepsilon_j + \hat{v}_j^s - \hat{v}_{j'}^s) \cdot \phi(\varepsilon_j) d\varepsilon_j
 \end{aligned}
 \tag{14}$$

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

$$\begin{aligned}
 & (1 - \mu^s) \int_{-\infty}^{\infty} \prod_{J_0^s} \Phi(-\lambda + \hat{v}_j^s - \hat{v}_{j'}^s + \varepsilon_j) \cdot \prod_{J_1^s \setminus \{j\}} \Phi(\varepsilon_j + \hat{v}_j^s - \hat{v}_{j'}^s) \cdot \phi(\varepsilon_j) d\varepsilon_j \\
 & + \mu^s \int_{-\infty}^{\infty} \prod_{J_0^s} \Phi(\lambda + \hat{v}_j^s - \hat{v}_{j'}^s + \varepsilon_j) \cdot \prod_{J_1^s \setminus \{j\}} \Phi(\varepsilon_j + \hat{v}_j^s - \hat{v}_{j'}^s) \cdot \phi(\varepsilon_j) d\varepsilon_j
 \end{aligned}
 \tag{15}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cumulative distribution and the probability density function of the standard normal distribution $N(0, 1)$, respectively (recall that ε is distributed $N(0, 1)$), and J_p^s denote the set of candidates with position $p \in \{0, 1\}$ who are running in state s .

Because the model is analytically complex, it is not always possible to prove that the comparative statics results hold for any arbitrary parameter combinations (though the analysis below suggests that they are in fact valid quite generally). However, we show numerically that they hold for a broad range of parameter values, including at the point estimates in Deltas et al. (2016).

4.1 Effect of drop-outs

Consider a situation in which three candidates compete initially, two of whom (say, A and B) have position 0, while the third one (C) has position 1. What happens to the support for 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} ; let R_{AC} be defined analogously. In the online Appendix, we show that

$$\begin{aligned}
 R_{AB} = & (1 - \mu) \int_{-\infty}^{\infty} [\Phi(\hat{v}_A - \hat{v}_B - \varepsilon) \Phi(\hat{v}_B - \hat{v}_C + \lambda + \varepsilon)] \phi(\varepsilon) d\varepsilon \\
 & + \mu \int_{-\infty}^{\infty} [\Phi(\hat{v}_A - \hat{v}_B - \varepsilon) \Phi(\hat{v}_B - \hat{v}_C - \lambda + \varepsilon)] \phi(\varepsilon) d\varepsilon
 \end{aligned}
 \tag{16}$$

and

$$\begin{aligned}
 R_{AC} = & (1 - \mu) \int_{-\infty}^{\infty} [\Phi(\hat{v}_A - \hat{v}_C + \lambda + \varepsilon) \Phi(\hat{v}_C - \hat{v}_B - \lambda - \varepsilon)] \phi(\varepsilon) d\varepsilon \\
 & + \mu \int_{-\infty}^{\infty} [\Phi(\hat{v}_A - \hat{v}_C - \lambda + \varepsilon) \Phi(\hat{v}_C - \hat{v}_B + \lambda - \varepsilon)] \phi(\varepsilon) d\varepsilon
 \end{aligned}
 \tag{17}$$

²⁹ Since we focus on the implications of voters' learning and preferences for vote shares, the specific rules for who wins the nomination do not matter; therefore, we are silent on that issue.

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. However, the expectation of R_{AB}/R_{AC} , taken over \hat{v}_j and μ , is positive for both the estimated Democratic and Republican parameter values from Deltas et al. (2016). Those expected values remain substantially above 1 for a range of parameter values around the point estimates, or when a mix of Republican and Democratic estimates are entered. This theoretical result corresponds well to the empirical findings reported in Sect. 3.1.

4.2 The effects of learning candidate valence over time

We now discuss voter updating of 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 that 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 normal, but now with expected value

$$\hat{v}_j^s = \frac{\sigma_\eta^2}{\sigma_{j0}^2 + \sigma_\eta^2} v_{j0} + \frac{\sigma_{j0}^2}{\sigma_{j0}^2 + \sigma_\eta^2} Z_j^s \tag{18}$$

and variance

$$(\sigma_{v_j^s}^2) = \frac{\sigma_{j0}^2 \sigma_\eta^2}{\sigma_{j0}^2 + \sigma_\eta^2}. \tag{19}$$

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 own state’s signal? Remember that those voters observe the vote share of each candidate j in each earlier state r , W_j^r , and know μ^r . Using (14) and (15), the election in state r is then captured by the following equation system:

$$\begin{aligned} & (1 - \mu^r) \int_{-\infty}^{\infty} \prod_{J_0' \setminus \{j\}} \Phi(\hat{v}_j^r - \hat{v}_j^r + \varepsilon_j) \cdot \prod_{J_1'} \Phi(\lambda + \varepsilon_j + \hat{v}_j^r - \hat{v}_j^r) \cdot \phi(\varepsilon_j) d\varepsilon_j \\ & + \mu^r \int_{-\infty}^{\infty} \prod_{J_0' \setminus \{j\}} \Phi(\hat{v}_j^r - \hat{v}_j^r + \varepsilon_j) \cdot \prod_{J_1'} \Phi(-\lambda + \varepsilon_j + \hat{v}_j^r - \hat{v}_j^r) \cdot \phi(\varepsilon_j) d\varepsilon_j = W_j^r, \quad \forall j \in J_0' \\ & (1 - \mu^r) \int_{-\infty}^{\infty} \prod_{J_0'} \Phi(-\lambda + \hat{v}_j^r - \hat{v}_j^r + \varepsilon_j) \cdot \prod_{J_1' \setminus \{j\}} \Phi(\varepsilon_j + \hat{v}_j^r - \hat{v}_j^r) \cdot \phi(\varepsilon_j) d\varepsilon_j \\ & + \mu^r \int_{-\infty}^{\infty} \prod_{J_0'} \Phi(\lambda + \hat{v}_j^r - \hat{v}_j^r + \varepsilon_j) \cdot \prod_{J_1' \setminus \{j\}} \Phi(\varepsilon_j + \hat{v}_j^r - \hat{v}_j^r) \cdot \phi(\varepsilon_j) d\varepsilon_j = W_j^r, \quad \forall j \in J_1' \end{aligned} \tag{20}$$

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

Proposition 1 *There exists a unique vector $(0, x_2, x_3, \dots, x_k)$ such that all solutions of (20) are of the form $(0, x_2, x_3, \dots, x_k) + (c, c, \dots, c)$, $c \in \mathbb{R}$.*

Proof See Appendix. □

It is immaterial which of the possible solutions to (20) a voter takes as his ex-ante belief, as a shift in ex-ante beliefs about all candidates by c translates into a shift of the ex-post beliefs by $\frac{\sigma_\eta^2}{\sigma_{\beta_0}^2 + \sigma_\eta^2} c$ 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 A's estimated valence to zero.

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

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

Proof See Appendix. □

Intuitively, a smaller variance of the valence estimates in later states translates into a lower variance of a candidate's vote shares in later states, relative to early states. In particular, that conclusion is clear in the limit: If (almost) no uncertainty about candidates' valences remains, 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. The prediction of Proposition 2 is borne out by the empirical results reported in Sect. 3.3 above.

4.3 Effect of partisan composition

To analyze the effect of the level of μ in different states on the support for different candidates, let us focus on the case wherein three candidates compete initially, 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

$$\begin{aligned} & \int_{-\infty}^{\infty} \Phi(v_A - v_B + \varepsilon) [\Phi(\lambda + v_A - v_C + \varepsilon) - \Phi(-\lambda + v_A - v_C + \varepsilon)] \phi(\varepsilon) d\varepsilon \\ & - \int_{-\infty}^{\infty} \Phi(v_B - v_A + \varepsilon) [\Phi(\lambda + v_B - v_C + \varepsilon) - \Phi(-\lambda + v_B - v_C + \varepsilon)] \phi(\varepsilon) d\varepsilon \geq 0. \end{aligned} \tag{21}$$

Without loss of generality, suppose that $v_A > v_B$. Whether (21) holds in general is difficult to determine. However, for $\lambda = 0$, (21) obviously holds as an equality and for λ sufficiently large, the left-hand and right-hand sides go to $\int_{-\infty}^{\infty} \Phi(v_A - v_B + \varepsilon) \phi(\varepsilon) d\varepsilon$ and $\int_{-\infty}^{\infty} \Phi(v_B - v_A + \varepsilon) \phi(\varepsilon) d\varepsilon$, so that (21) is satisfied as strict inequality. Moreover, the

left-hand side of (21) is positive (in expectation over valence draws) at the estimated parameter values in Deltas et al. (2016).

We now focus on relative changes. Proposition 3 shows that, if λ is sufficiently large, then the weaker candidate benefits *proportionately more* than the strong candidate (i.e., relative to previous vote share) from a favorable ideological shift of the electorate.

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

Proof See Appendix. □

We conjecture that Proposition 3 holds more generally, for any λ , but again it is hard to prove. However, as above, we can also check that Proposition 3 holds around the estimated parameter values by Deltas et al. (2016). The result is supported by the estimates in Sect. 3.2.

5 Concluding remarks

The results of this paper demonstrate that ideological differentiation between candidates leads to substantial vote-splitting among those candidates who are ideologically similar. Therefore, multi-candidate primary elections may be severely affected by coordination failures because 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 splits votes with a close ideological neighbor. The US 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 primary season and the voters who would have voted for a dropped-out candidate need to choose which of the remaining candidates to support. The sequential nature of the primaries also allows us to infer, using the pattern of decline in vote-share variability, that voters become better informed about candidate quality by observing the outcomes of earlier election rounds.

Acknowledgements We would like to acknowledge useful comments by Jeff Milyo, Jungmo Yoon, and seminar participants at Hanyang University, the University of Cyprus, University of Missouri, the University of Piraeus, SUNY Stony Brook, and at the American Economic Association Annual Meetings.

Appendix

See Tables 6 and 7.

Table 6 Election data: the Republican primaries. Sources: Federal Election Commission, George Washington University, CNN, New York Times, Dave Leip's Atlas of US Presidential Elections, USA Today, Wikipedia

State	2000 Rep. primary				2008 Republican primary				2012 Rep. primary							
	rd	Bauer	Bush	Frbs	Keyes	McC	rd	Giul	Huck	McC	Rmn	Thm	rd	Gngr	Rmn	Sntm
Alabama							7		41%	37%	18%		13	29%	29%	35%
Alaska							7		22%	15%	44%		11	14%	32%	29%
Arizona	5		53%		5%	42%	7		9%	47%	35%		8	16%	47%	27%
Arkansas							7		61%	20%	14%					
California	7		61%		4%	35%	7		12%	42%	35%					
Colorado							7		13%	18%	60%		6	13%	35%	40%
Connecticut	7		46%		3%	49%	7		7%	52%	33%					
DC							9		16%	68%						
Delaware	3		51%	20%	4%	25%	7		15%	45%	33%					
Florida							5	15%	14%	36%	31%		4	32%	46%	13%
Georgia	7		67%		5%	28%	7		34%	32%	30%		11	47%	26%	20%
Hawaii													13	11%	44%	25%
Idaho													11	2%	62%	18%
Illinois							7		17%	47%	29%		14	8%	47%	35%
Indiana																
Iowa	1	9%	41%	31%	14%	5%	1	4%	35%	13%	25%	13%	1	13%	25%	25%
Kansas							8		60%	24%			12	14%	21%	51%
Kentucky																
Louisiana							8		43%	42%			15	16%	27%	49%
Maine	7		51%		3%	44%	6		6%	22%	52%		7	6%	38%	18%
Maryland	7		56%		7%	36%	9		29%	55%			16	11%	49%	29%
Massachu.	7		32%		3%	65%	7		4%	41%	51%		11	5%	72%	12%
Michigan	5		43%		5%	51%	3	3%	16%	30%	39%	4%	8	7%	41%	38%
Minnesota	7		63%		17%	20%	7		20%	22%	41%		6	11%	17%	45%
Mississippi													13	31%	31%	33%

Table 6 (continued)

State	2000 Rep. primary					2008 Republican primary					2012 Rep. primary					
	rd	Bauer	Bush	Frbs	Keyes	McC	rd	Giul	Hueck	McC	Rmn	Thm	rd	Gngr	Rmn	Sntm
Missouri	7		58%		6%	35%	7		32%	33%	29%					
Montana							7		15%	22%	38%					
Nebraska								4	4%	13%	51%	8%	5	21%	50%	10%
Nevada							2	8%	11%	37%	32%	1%	2	9%	39%	9%
N. Hampshire	2	1%	30%	13%	6%	49%	7		8%	55%	28%					
New Jersey																
New Mexico																
New York	7		51%		3%	43%	7		11%	52%	28%					
N. Carolina																
N. Dakota	6		76%		5%	19%	7		20%	23%	36%		11	8%	24%	40%
Ohio	7		58%		4%	37%	11		31%	60%			11	15%	38%	37%
Oklahoma							7		33%	37%	25%		11	27%	28%	34%
Oregon																
Pennsylvania																
Rhode Island	7		36%		3%	60%	11		22%	65%						
S. Carolina	4		53%		5%	42%	4	2%	30%	33%	15%	16%	3	40%	28%	17%
S. Dakota																
Tennessee							7		34%	32%	24%		11	24%	28%	37%
Texas							11		38%	51%						
Utah							7		2%	5%	90%					
Vermont	7		34%		3%	60%	11		14%	72%			11	8%	39%	24%
Virginia	6		53%		3%	44%	9		41%	50%						
Washington	6		58%		2%	39%	10		24%	50%			10	10%	38%	24%
West Virginia																

Table 6 (continued)

State	2000 Rep. primary				2008 Republican primary					2012 Rep. primary							
	rd	Bauer	Bush	Frbs	Keyes	McC	rd	Giul	Huck	McC	Rmn	Thm	rd	Gngr	Rmn	Sntm	
Wisconsin							10		37%	55%			16	6%	44%	37%	
Wyoming								Not applicable: see text					9	8%	39%	32%	
Candidate position classification																	
Extreme vs. moderate	1	1	1	1	1	0	0	0	1	0	1	1	1	1	0	1	1

Empty cells are contests after a candidate's official withdrawal or the last competitive primary, or otherwise not used. Positions: 1, conservative; 0, moderate. See text for details

rd round, *Frbs* Forbes, *Giul* Giuliani, *Huck* Huckabee, *McC* McCain, *Rmn* Romney, *Thm* Thompson, *Gngr* Gingrich, *Sntm* Santorum

Table 7 Election data: the democratic primaries and the 1992 general election. Sources: Federal Election Commission, ABC, CBS, CNN, MSNBC, PBS, USA Today

State	2000 Dem. primary				2004 Democratic primary						2008 Dem. primary				1992 Election		
	rd	Brad	Gore	rd	Clrk	Dean	Eds	Geph	Kerry	Lieb	rd	Clint	Eds	Obm	Clinton	Perot	
Alabama											6	42%		56%	41%	11%	
Alaska											6	25%		75%	30%	28%	
Arizona				3	26%	14%	7%		43%	7%	6	50%		43%	37%	24%	
Arkansas											6	70%		26%	53%	10%	
California	5	18%	81%	9			20%		65%		6	52%		43%	46%	21%	
Colorado											6	32%		67%	40%	23%	
Connecticut	5	42%	55%	9			24%		58%		6	47%		51%	42%	22%	
DC											8	24%		76%	85%	4%	
Delaware	3			3	10%	10%	11%		50%	11%	6	43%		53%	44%	20%	
Florida											5	50%	14%	33%	39%	20%	
Georgia	5	16%	84%	9			41%		47%		6	31%		67%	43%	13%	
Hawaii				8			14%		50%		9	24%		76%	48%	14%	
Idaho				8			22%		54%		6	17%		79%	28%	27%	
Illinois											6	33%		65%	49%	17%	
Indiana											14	51%		49%	37%	20%	
Iowa	1	35%	63%	1	0%	18%	32%	11%	38%	0%	1	29%	30%	38%	43%	19%	
Kansas											6	26%		74%	34%	27%	
Kentucky											16	65%		30%	45%	14%	
Louisiana											7	36%		57%	46%	12%	
Maine	6	41%	54%	4	4%	28%	8%		45%		7	40%		59%	39%	30%	
Maryland	6	28%	67%	9			26%		60%		8	36%		61%	50%	14%	
Massachu.	6	37%	60%	9			18%		72%		6	56%		41%	48%	23%	
Michigan				4	7%	17%	13%		52%			not applicable			44%	19%	
Minnesota				9			27%		51%		6	32%		66%	43%	24%	
Mississippi											12	37%		61%	41%	9%	

Table 7 (continued)

State	2000 Dem. primary			2004 Democratic primary						2008 Dem. primary				1992 Election		
	rd	Brad	Gore	rd	Clrk	Dean	Eds	Geph	Kerry	Lieb	rd	Clint	Eds	Obm	Clinton	Perot
Missouri	6	34%	65%	3	4%	9%	25%		51%	4%	6	48%		49%	44%	22%
Montana											17	41%		57%	38%	26%
Nebraska				6		17%	10%		63%		7	32%		68%	29%	24%
Nevada				2	12%	26%	12%		38%	9%	3	51%	4%	45%	37%	26%
N. Hampshire	2	46%	50%	2							2	39%	17%	37%	39%	23%
New Jersey											6	54%		44%	43%	16%
New Mexico				3	21%	16%	11%		42%	3%	6	49%		48%	46%	16%
New York	6	33%	66%	9			20%		61%		6	57%		40%	50%	16%
N. Carolina											14	42%		56%	43%	14%
N. Dakota				3	24%	12%	10%		50%	1%	6	37%		61%	32%	23%
Ohio	6	25%	74%	9			34%		52%		10	53%		45%	40%	21%
Oklahoma				3	30%	4%	30%		27%	7%	6	55%		31%	34%	23%
Oregon											16	41%		59%	42%	24%
Pennsylvania											13	55%		45%	45%	18%
Rhode Island	6	40%	57%	9			19%		71%		10	58%		40%	47%	23%
S. Carolina				3	7%	5%	45%		30%	2%	4	27%	18%	55%	40%	12%
S. Dakota											17	55%		45%	37%	22%
Tennessee				5	23%	4%	27%		41%		6	54%		41%	47%	10%
Texas											10	51%		48%	37%	22%
Utah				8			30%		55%		6	39%		57%	25%	27%
Vermont	6	44%	54%	9			6%		32%		10	39%		59%	46%	23%
Virginia				5	9%	7%	27%		52%		8	35%		64%	41%	14%
Washington	4	32%	68%	4	3%	30%	7%		49%		7	31%		68%	43%	24%
West Virginia											15	67%		26%	48%	16%

Table 7 (continued)

State	2000 Dem. primary			2004 Democratic primary					2008 Dem. primary			1992 Election					
	rd	Brad	Gore	rd	Clrk	Dean	Eds	Geph	Kerry	Lieb	rd	Clint	Eds	Obm	Clinton	Perot	
Wisconsin				7		18%	34%		40%		9	41%		58%	41%	22%	
Wyoming											11	38%		61%	34%	26%	
Candidate position classification																	
Insider vs. outsider	1	0			1	1	1	0	0	0		0		1	1		1
Extreme vs. moderate	1	0			0	1	0	0	1	0		0		1	1		1

Empty cells are contests after a candidate's official withdrawal or the last competitive primary, or otherwise not used. Positions: 0, insider; 1, outsider; 1, conservative; 0, moderate. See text for details

rd round, *Brad* Bradley, *Clrk* Clark, *Eds* Edwards, *Geph* Gephardt, *Lieb* Lieberman, *Clint* Clinton, *Obm* Obama

References

- Adams, J., & Merrill, S. (2008). Candidate and party strategies in two-stage elections beginning with a primary. *American Journal of Political Science*, 52(2), 344–359.
- 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.
- Bartels, L. M. (1988). *Presidential primaries and the dynamics of public choice*. Princeton: Princeton University Press.
- Callander, S. (2007). Bandwagons and momentum in sequential voting. *Review of Economic Studies*, 74, 653–684.
- Dekel, E., & Piccione, M. (2000). Sequential voting procedures in symmetric binary elections. *Journal of Political Economy*, 108(1), 34–55.
- Deltas, G., Herrera, H., & Polborn, M. K. (2016). Learning and coordination in the presidential primary system. *Review of Economic Studies*, 83(4), 1544–1578.
- Grossmann, M., & Hopkins, D. A. (2016). *Asymmetric politics: Ideological Republicans and group Interest Democrats*. Oxford: Oxford University Press.
- Kenny, P. J., & Rice, T. W. (1994). The psychology of political momentum. *Political Research Quarterly*, 47(4), 923–938.
- Klump, T., & Polborn, M. K. (2006). Primaries and the New Hampshire effect. *Journal of Public Economics*, 90, 1073–1114.
- Knight, B., & Hummel, P. (2015). Sequential or simultaneous elections? An empirical welfare analysis. *International Economic Review*, 56, 851–887.
- Knight, B., & Schiff, N. (2010). Momentum and social learning in presidential primaries. *Journal of Political Economy*, 118, 1110–1150.
- Krasa, S., & Polborn, M. K. (2010). The binary policy model. *Journal of Economic Theory*, 145(2), 661–688.
- Poole, K., & Rosenthal, H. (1985). A spatial model for legislative roll call analysis. *American Journal of Political Science*, 29(2), 357–384.
- Schwabe, R. (2015). Super Tuesday: Campaign finance and the dynamics of sequential elections. *Social Choice and Welfare*, 44(4), 927–951.
- Serra, G. (2011). Why primaries? The party's tradeoff between policy and valence. *Journal of Theoretical Politics*, 23(1), 21–51.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(3), 817–838.