

CANDIDATE COMPETITION AND VOTER LEARNING IN THE 2000-2012 US PRESIDENTIAL PRIMARIES

GEORGE DELTAS

Department of Economics

University of Illinois

Champaign, IL 61820

and Lancaster University

United Kingdom

deltas@illinois.edu

MATTIAS K. POLBORN

Department of Economics and

Department of Political Science

Vanderbilt University

Nashville, TN 37204

mattias.polborn@vanderbilt.edu

SEPTEMBER 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 U.S. Presidential primaries and show that it is of first order importance. We also show that voter 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 this 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.

JEL Classification Numbers: D72, D60.

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

1 Introduction

Candidates for the U.S. presidential election are determined through a sequence of elections within each political party, the primaries, 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, there are often more than just two candidates who compete with each other, and this situation generates coordination problems for voters and candidates that may result in the nomination of an inferior candidate, either quality-wise or in the sense that the nominee does not represent the majority-preferred position. We consider a situation in which candidates differ both “horizontally” (i.e., with respect to their policy positions) and “vertically” (i.e., with respect to their quality, often called “valence”). For example, Republican primary candidates may be either “moderates” or “conservatives”, and each voter has a preference for one of these positions, which, however, is not absolute: If a voter considers a candidate in the other position to have a sufficiently higher valence, he would vote for that candidate rather than an ideologically closer competitor.

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

This vote-splitting effect presents a substantial problem for the efficiency of any voting system, and not just for primaries. When more than two candidates run in an election, a weaker candidate might win in a situation where the Condorcet winner is splitting votes with a close ideological neighbor. The sequential presidential primary system provides a unique opportunity to gauge the presence and size of this vote-splitting effect, because some candidates drop out during the primaries, and those voters who would have voted for the drop-out 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 U.S. 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 in 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, this 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 three times as many votes from competing candidates of the same position than he/she will from other candidates. This 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 may ultimately lead to the nomination of the “wrong” candidate.

We then show that electoral variability declines over time. Variability is defined to capture voter learning over time, facilitated through observation of previous election results. This effect is measured without making parametric assumptions, by utilizing the fact that many state contests are taking place on the same date. We show that the variability of vote shares, controlling for other factors through the use of election round fixed effects, decreases 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 in 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 further 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, Herrera, and Polborn (2016). In addition to explaining the observed substitutability and volatility patterns, this model also predicts that an increase of the share of voters who prefer a particular political position leads to a higher increase in the *absolute* number of votes for a strong candidate rather than a weak candidate in that position, but *relatively*, weak candidates benefit more than strong ones. We show that this prediction bears out in our data.

Differential competition between candidates is directly relevant to (tactical) coordination between voters. This, along with learning about candidate quality, have long been known to be important

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

issues in presidential primaries. For example, Bartels (1987, pp.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)). These papers 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. This implies 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, Herrera, and Polborn (2016). Knight and Schiff (2010) and Knight and Hummel (2015) develop a model of voter learning about candidate quality in which voters in later states receive some imperfect information about the signal that voters in earlier states observed. Voters update, taking all pieces of information into account, and vote for their preferred candidate given this information. In their estimation based on the 2004 Democratic Presidential Primaries, they find that voters attach a substantial weight on the outcomes of early elections, but a much smaller weight after the fourth primary. Thus, in their framework, predicted share volatility declines up to the fifth primary round, but is essentially constant thereafter. Our empirical strategy is agnostic about whether a voter in a state infers perfectly or noisily the signal that voters in other states have observed by the voting outcome in that state. However, our results suggest that much of this signal is directly observed (as in our model) given that share volatility falls throughout the primary season, and not only after the first few election contests.

Deltas, Herrera, and Polborn (2016) analyze a structural model of learning about candidate quality in which candidates are also 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, Herrera, and Polborn (2016), but shows that the model can explain the empirical findings.³

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

³Among theoretical static models of primaries (i.e., those where there is only one vote at the primary stage), our

2 Data

Our dataset consists of information from six of the 2000, 2004, 2008, and 2012 United States Presidential primaries – we exclude the 2004 Republican and the 2012 Democratic primaries because the incumbent Presidents were effectively unopposed. We focus on candidates who are initially considered viable candidates in the sense that there is some chance that they will win their party’s nomination. In practice, some of the candidates do not fall into this category because they are too far away from their party’s mainstream, and 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 data set.

The most successful excluded candidates are Dennis Kucinich (Democratic primary 2004 and 2008) and Ron Paul (Republican primary 2008). Their vote share is usually higher in low-turnout contests later in the sequence in which their energized base represents a larger fraction of the electorate. In contrast, unsuccessful but potentially “serious” candidates (for example, Joe Lieberman (D-2004) or Rudy Giuliani (R-2008)) have their best performances in early primaries, then lose voter support due to their relatively poor performance, and eventually drop out once it becomes clear that they have no chance of winning the nomination. Tables A1 and A2 list the candidates we include for each primary, along with the states in which they competed and the vote share they obtained. The tables also give the number of different election dates (rounds) up to the election in each state.

A key component of our empirical analysis (also incorporated in the model in Section 4 below) is that candidates of 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 A1 and A2.

In the Republican party, the main ideological fault line appears to be between *conservatives* (i.e., candidates and voters who often have a fundamentalist Christian background and emphasize “value-issues” such as abortion and gay marriage) and *moderates*. A standard approach to determining a candidate’s position is the use of NOMINATE scores based on roll-call votes (see Poole and Rosenthal (1985)). However, such scores are only available for legislators, and the majority of candidates has an executive background (e.g., former governors). Our classification is therefore guided by common sense and exit polls that ask voters which candidate they voted for, and whether they personally identify as conservative, moderate or liberal. We focus on exit polls in early primary or caucus

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.

states, as these are usually the only ones in which all candidates we consider are running and where each of them receives a sufficiently large vote share. For example, in the 2000 Republican contest, George W. Bush did considerably better with voters who identified as conservative rather than with those who said they were moderate, and vice versa for John McCain.⁴ For this reason, we classify Bush as conservative and McCain as moderate. In 2008, we take the MSNBC exit polls (available on <http://www.msnbc.msn.com/id/21660890>), since they ask voters to identify as conservative, moderate or liberal, while CNN has dropped this question in many exit polls). McCain and Giuliani always do considerably better with voters who identify as moderates, while Huckabee and Thompson do considerably better with conservatives. Romney generally does better with conservatives than with moderates, except for states in which the Republican primary electorate is extremely conservative. For example, in Iowa, 88 percent of Republican primary voters identify as strongly or somewhat conservative, while only 11 percent identify as moderates. Romney receives about the same vote share from conservatives and moderates (25 percent versus 26 percent). However, in states like Michigan or Florida where the percentage of conservatives is around 60, Romney does substantially better with conservatives than with moderates. Moreover, in the later stages of the campaign, Romney was perceived to fight with Huckabee over the conservative vote.⁵ For this reason, we classify Romney as conservative. In the 2012 primary, however, Romney was the moderate standard-bearer, facing Gingrich and Santorum who were supported by 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 there is no a priori reason why the 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 their opponents, we would expect that voter ideology is 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 Republican party and Democratic party are radically asymmetric organizations: “While the Democratic party is fundamentally a group coalition, the Republican party

⁴See, 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.

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

⁶The 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 for their vote in the presidential primary.

can be most accurately characterized as a vehicle of an ideological movement.”

For Democrats, we would argue that there are two logically defensible classification approaches, 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 voter appears to have much less predictive power than for Republicans. In fact, for some states, the differences in candidate support from the liberal and moderate segments seems to entirely disappear. For example, in Nevada, self-declared liberals voted 48/39/9 for Clinton, Obama and Edwards, while moderates voted 46/43/8 for these candidates. This difference between liberals and moderates is well within the margin of error.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 is clearly 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 is probably 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 this is mostly 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 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

⁷The 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 judgment of their position) often receive rather similar percentages of their votes from liberal and moderate voters.

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

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 these candidates and election contests, we obtain the vote percentage in the primary or caucus of each state from the Federal Election Commission and major media sources. These vote shares are reported in Tables A1 and A2. However, these shares do not sum up to 100 percent as they include votes for candidates whom we dropped from our analysis, for candidates who have already withdrawn, or for “uncommitted” delegates. We treat these votes as equivalent to abstention from weighing on the choice of the party nominee. To 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 these 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 3-way race is expected to be associated with liberal preferences, whereas a high Bush percentage is expected to be associated with conservative preferences. The 3-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 sequence of the primary, 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 this approach is that it is not based on any specification assumptions. This 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, etc.). We discuss these limitations in more detail at the end of this subsection, before proceeding to more formal analysis.

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

Let $\text{VoteShare}_{j,y}^{\kappa,\kappa'}$ be the vote share of candidate j (measured on a 0-100% scale) who shares his position $a(j)$ with $\kappa - 1$ other candidates, while there are κ' competitors in the opposite 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 in the same position), then the mean share of a candidate is, by definition, $1/\kappa$. It never happens that all participants in a primary belong to the same political position, and thus these 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 these values are not “data” but their values are driven by this 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, these 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 in the same position increases the average vote share of the remaining candidates in that position. Formally, $\text{VoteShare}^{\kappa-1,\kappa'} > \text{VoteShare}^{\kappa,\kappa'}$. Second, there is partial, but not complete “crowding out” among candidates in the same position: A reduction in the number of candidates in the same position decreases the total vote share of the candidates in that position because there are some cross-over voters who change to a candidate in the opposite position. Formally, $\kappa \cdot \text{VoteShare}^{\kappa,\kappa'} > (\kappa - 1) \cdot \text{VoteShare}^{\kappa-1,\kappa'}$.

By-and-large, the data are consistent with these expectations. For example, when going from three candidates in a 2-1 constellation to two candidates in a 1-1 constellation, the vote share of the candidate in the previously crowded position increases from 28.6% to 50%, while the vote share of the competitor increases only from 42.8% to 50% (remember that, by definition, when $\kappa = \kappa' = 1$, the average vote share of candidates is 0.5). Or, interpreted in the other direction: A very competitive race between two candidates in different positions, each attracting 50 percent of the votes, can become very

non-competitive when another candidate enters, because the lonely candidate now attracts significantly more votes than each of his competitors. This vote-splitting may lead to the victory of a candidate who would lose if he had only one competitor. Note also that, if positions were irrelevant for voters, then entry by the third candidate would instead reduce the vote share of existing candidates to $1/3$.

Similarly, going from a 3-2 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 in a specific position is always increasing in the number of candidates in that position. For example, consider all contests involving 5 candidates: Here, $4 \times 17.9\% = 71.6\% > 3 \times 19.0\% = 57.0\% > 2 \times 21.5\% = 43.0\% > 28.3\%$. Thus, there is clearly diversion of votes from one candidate to another candidate in the same location, but the more candidates are in a location, the bigger their combined share. The same pattern holds for contests with 3 and 4 candidates.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 in the crowded position decreases from 17.9% to 17.3%. This is probably due to the small number of cases: there were only two state elections with a 4-1 constellation, and six with a 3-1 constellation (indeed, the p-value of the difference is 0.93, i.e., there is no confidence whatsoever in the sign of the gap). The flip side of this comparison is that the lonely candidate in a 3-1 constellation is doing surprisingly well, getting on average 48 percent of the vote. This is also responsible for the fact that going from 3-1 to 2-1 reduces the vote share of the lonely candidate from 48.2% to 42.8% (the difference is also 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.

This concludes the discussion of mean vote shares when Republicans are classified as Moderate or Conservative and Democrats as Insiders or Outsiders. When using the Extreme vs Moderate classification for both parties, the configuration of 3 candidates in one location and one in the opposing location is based on a single election (4 candidate-share observations), in which the lone moderate by

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

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

sheer coincidence had a 25 percent vote share.¹² However, there is only minimal information in a single election, making any interpretation unwarranted.

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 location, holding the total number of candidates constant. Moreover, the combined vote shares of candidates in a location increases with the number of candidates in that location, holding the total number of candidates constant.

Not distinguishing the election sequencing does not lead to any biases for the questions we address with this analysis. Treating political parties and positions as fungible does not create any biases, provided that the political locations do not differ systematically in voter popularity. Our analysis in the next section suggests that this 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 there is no systematic difference across political positions in the number of candidates in that position. Though this is essentially true for the Democrats, it is not true for the Republicans (there are typically fewer moderates than conservatives). But given that political positions do not differ much in popularity among the voters, any differences in their “popularity” among politicians would not impact the validity of our results. Overall, the main value of the analysis described here is the absence of any parametric or modeling assumptions, except for those qualitative properties listed in this paragraph.

Since information about the sequence of elections has not been used in this analysis, it cannot provide any evidence regarding the possibility of voter learning. Neither can it assess which of several candidates in the same position benefits most if a state leans more toward these candidates’ common position, which can be an indirect test of candidate differentiation. We address these questions in the next two sections through the use of formal econometric specifications.

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 position, 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 approach, using progressively more flexible specifications.¹³ The findings are useful when thinking about which

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

¹³However, 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 informa-

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 which 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 vs 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 vs moderate classification, under which Democrats are characterized as liberal or moderate and Republicans as conservative or moderate. We discuss the results of both classifications using the same econometric specification together.

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

We next investigate whether this result is affected by the relative popularity of candidates of different political positions. Model 2 suggests this 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 add the indicator variable Dem_j for Democratic candidates to complete the set of interactions. In the regression

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

tion available in a given time is correlated across states. This is accomplished by using White’s (1980) heteroskedasticity 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).

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

the estimate of β_1 remains smaller than that of β_2 , albeit with a difference that is not statistically significant. However, there is insufficient information to credibly estimate this specification, 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 percent). 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 in these positions varies across parties. This observation, along with the fact that standard errors of these parameters are much larger than in other specifications, indicates there is insufficient variation in our data to meaningfully estimate this model. Similar conclusions are obtained under the extreme vs 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 this specification, we investigate whether the relevance of candidate political location is present in both major parties, or is confined to one of them. We do so by estimating

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

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, where a candidate’s vote share is only negligibly affected by competing against one fewer candidate in the opposing political location, but is very strongly affected by one fewer candidate in the same political location. The relative effect of the location of competing candidates is also statistically significant for the Republican primaries, but smaller in quantitative terms.¹⁵ However, the difference between

¹⁵Adding location dummies to this regression does not materially affect the estimates for the Democratic coefficients, but renders the Republican location effect to essentially zero. However, with Moderate_i being essentially a dummy for McCain, the Republican effect would be identified solely from the gain of voters by McCain as other candidates (of opposing location) depart, relative to the gain of voters by his opponents as other candidates (of same location) depart. Not only the effective information for this specification is even more limited (only 4 such withdrawals) but with McCain being a higher quality candidate, the location and valence effects are confounded (McCain gets a bigger than expected share of the departing candidates’ voters because he is a better candidate in the vertical dimension).

the Democrats and the Republicans is smaller when using the extreme vs moderate classification for both parties. One could add the Democratic dummy to the specification in equation 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 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 would also be reasonable to expect that the vote share 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. These conjectures are formalized in section 4. Our final set of regressions are intended to ascertain whether they are also supported by the data, thus buttressing the support of our framework over the alternatives. These conjectures are hard to test because they demand much from our limited data (we only observe proxies of 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 should also 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 a proxy 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 are plausibly shifted to the right relative to the rest of the country, and we would therefore also expect that conservative Republicans do better in these states than moderates. Similarly, voter preferences in states in which Clinton did well are plausibly shifted to the left relative to the rest of the country, and we would therefore also expect that liberal Democrats 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 that candidates classified as outsiders do better.

To test the validity of these propositions, – a pre-requisite to using 1992 vote shares to investigate

the effect of electorate preferences on candidates of difference valence – we estimate the equations

$$\begin{aligned} \text{VoteShare}_{j,s,y} = & \alpha_{j,t,y} + \gamma_{0B}\text{Rep}_j\text{Bush92}\%_s + \gamma_{0P}\text{Dem}_j\text{Perot92}\%_s \\ & + \gamma_{1B}\text{Conservative}_j\text{Bush92}\%_s + \gamma_{1P}\text{Outsider}_j\text{Perot92}\%_s + \epsilon_{j,s,y} \end{aligned} \quad (5)$$

and

$$\begin{aligned} \text{VoteShare}_{j,s,y} = & \alpha_{j,t,y} + \gamma_{0B}\text{Rep}_j\text{Bush92}\%_s + \gamma_{0C}\text{Dem}_j\text{Clinton92}\%_s \\ & + \gamma_{1B}\text{Conservative}_j\text{Bush92}\%_s + \gamma_{1C}\text{Liberal}_j\text{Clinton92}\%_s + \epsilon_{j,s,y} \end{aligned} \quad (6)$$

where $\text{Bush92}\%_s$, $\text{Perot92}\%_s$ and $\text{Clinton92}\%_s$ are Bush’s, Perot’s and Clinton’s vote share 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. These dummies would perfectly predict the share of a candidate for election days in which only a single state votes, completely eliminating their influence on the remaining model parameters. Thus, we drop observations that consist of a single state contest from the regression in equation 5 and equation 6, reducing the number of observations from 502 to 382 (this is also done in all subsequent specifications that use an exhaustive set of candidate \times year \times round effects). The flexibility embodied in the use of these 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 these 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. This is indeed the case, with an 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 strongly statistically significant).¹⁸ The Perot effect on Democratic outsider candidates is positive, with a significance that barely misses the five percent cut-off. The point estimate is also smaller, with a percentage increase in Perot’s vote share increase 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 conservative populists, so that the influence on the Democratic primary electorate is weaker. Finally, there is effectively no Clinton effect on liberal

¹⁶The vote share variables $\text{Bush92}\%_s$, $\text{Clinton92}\%_s$ and $\text{Perot92}\%_s$, like $\text{VoteShare}_{j,s,y}$ range from 0 to 100.

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

¹⁸The 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 equation 5 and equation 6 are omitted from Table 2 to conserve space, but are available upon request.

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

We next turn to our proxies for candidate valence. Our first proxy for valence in round t is the vote average share of a candidate in that round, $MeanShr_{j,t,y}$. Clearly this is an imperfect measure, but a reasonable one. Candidates with high relative valence, as perceived in round t , will have higher values of $MeanShr_{j,t,y}$. The number and distribution of competing candidates will also affect the values of $MeanShr_{j,t,y}$. To reduce these candidate composition effects on this measure, we employ this measure in regressions that also include candidate \times round effects, as in regressions 5 and 6.¹⁹ Moreover, averaging vote shares of all contests in a round is meaningful because all states have the same ex ante expectations about valence which they update independently on the basis of their privately observed signal. In addition, the set of candidates is the same in all such contests. However, adding $MeanShr_{j,t,y}$ on the right hand 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 primary at the same time).

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

$$\begin{aligned}
VoteShare_{j,s,y} = & \alpha_{j,t,y} + \gamma_{0B}Rep_jBush92\%_s + \gamma_{0P}Dem_jPerot92\%_s \\
& + \gamma_{1B}Conservative_jBush92\%_s + \gamma_{1P}Outsider_jPerot92\%_s \\
& + \{\gamma_{2Bc}Conservative_j + \gamma_{2Bm}ModerateRep_j\} Bush92\%_s MeanShr_{j,t/s,y} \\
& + \{\gamma_{2Po}Outsider_j + \gamma_{2Pi}Insider_j\} Perot92\%_s MeanShr_{j,t/s,y} + \epsilon_{j,s,y}
\end{aligned} \tag{7}$$

for the Outsider/Insider classification and

$$\begin{aligned}
VoteShare_{j,s,y} = & \alpha_{j,t,y} + \gamma_{0B}Rep_jBush92\%_s + \gamma_{0C}Dem_jClinton92\%_s \\
& + \gamma_{1B}Conservative_jBush92\%_s + \gamma_{1C}Liberal_jClinton92\%_s \\
& + \{\gamma_{2Bc}Conservative_j + \gamma_{2Bm}ModerateRep_j\} Bush92\%_s MeanShr_{j,t/s,y} \\
& + \{\gamma_{2Cl}Liberal_j + \gamma_{2Cm}ModerateDem_j\} Clinton92\%_s MeanShr_{j,t/s,y} + \epsilon_{j,s,y}
\end{aligned} \tag{8}$$

¹⁹The theoretical model shows that the effect of changes in electorate preferences on candidate shares depends not only on the candidate's valence and political position but also on the number of competing candidates, their valence, and their political position. The variable $MeanShr_{j,t,y}$ also adjusts for the number of competing candidates, their valence and political position, and thus, in a qualitative way, reflects the factors that enter in the comparative statics of the theoretical model.

for the Extreme/Moderate classification, where $\text{MeanShr}_{j,t/s,y}$ 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, has a potential endogeneity concern, though one of ambiguous sign (and possibly of zero magnitude).²¹ A more conservative approach is to lag the $\text{MeanShr}_{j,t,y}$ variable by one round, i.e., use as a proxy of valence $\text{MeanShr}_{j,t-1,y}$. This yields Model 5 below

$$\begin{aligned} \text{VoteShare}_{j,s,y} = & \alpha_{j,t,y} + \gamma_{0B}\text{Rep}_j\text{Bush92}\%_s + \gamma_{0P}\text{Dem}_j\text{Perot92}\%_s \\ & + \gamma_{1B}\text{Conservative}_j\text{Bush92}\%_s + \gamma_{1P}\text{Outsider}_j\text{Perot92}\%_s \\ & + \{\gamma_{2Bc}\text{Conservative}_j + \gamma_{2Bm}\text{ModerateRep}_j\} \text{Bush92}\%_s \text{MeanShr}_{j,t-1,y} \quad (9) \\ & + \{\gamma_{2Po}\text{Outsider}_j + \gamma_{2Pi}\text{Insider}_j\} \text{Perot92}\%_s \text{MeanShr}_{j,t-1,y} + \epsilon_{j,s,y} \end{aligned}$$

for the Insider/Outsider classification, and a similar expression for the Extreme/Moderate classification. This specification is not necessarily better than the one in (7) for two reasons. First, lagging the mean share provides a more noisy measure of perceived valence for round t because it does not include the signals received in that round. Second, the set of candidates is no longer guaranteed to be the same across round t and $t - 1$, and this introduces an additional source of noise in the valence proxy.

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 given to 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 are generally more statistically significant for the first specification. To a large degree, this 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 where no parameters are estimated from information from both parties. Conservative Republicans do better than moderates in states where Bush obtained a higher vote share in

²⁰Note 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.

²¹To see this, suppose there are only two states, 1 and 2 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 equation for the vote share in state 1 ($\gamma < 1$).

1992, and the effect is larger for candidates of higher valence. From parameter values in columns 4 and 8 of Table 2, the effect of a one percent 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.” This gap increases for candidates with positive 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 one percent increase in Bush 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 bigger vote reductions than moderates with low vote shares. However, the negative estimate for γ_{2Bc} indicates that in those states, conservatives of high valence were helped by less compared with conservatives of lower valence. This 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 got a big vote share, and this gap is increasing with candidate valence ($\gamma_{2Po} > \gamma_{2Pi}$). The effect for candidates who obtain a fraction of 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 vs 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 had exhibited stronger support for Perot. But so do, to a smaller and far less statistically significant degree, Outsider Democrats with large average vote shares; this largely mirrors the findings for Republican candidates.

Moving the Liberal vs 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 is essentially zero. For both classification systems, liberal as well as outsider candidates do better in states with a higher 1992 Democratic vote share. Recall, however, that the estimates of

²²These conclusions about relative effects are based on the parameters reported in Table 2. To obtain the level effect, one needs to also incorporate the effect of the Bush 1992 percentage on Republican vote shares, which is 1.38 (not reported in Table 2 to save space). Evaluating at a value of MeanShr = 0.4, a one percent 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).

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 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 this uncertainty is slowly resolved over time, we posit that vote share variability declines over time. Moreover, since additional information moves perceptions (and thus vote shares) by a progressively smaller amount, the largest decline in variability should happen early, and be related to the level of information received (proxied by the number of states that have already voted) rather than to the simple passage of time.

Estimates of vote share variability necessarily have to be based on the 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 via Maximum Likelihood (GLS). The former method is consistent under heteroskedasticity 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 as a robustness exercise. Part of the reason for this 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 use 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 due to learning, all parameters associated with systematic differences in the expected vote shares are nuisance parameters: We do not care about their values here, except that they are accounted for as best as possible. Our base model to obtain the residuals has an exhaustive set of candidate \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

position and number) in the most flexible way: with indicator variables whose coefficients vary (with no parametric constraints) over time. This regression is equivalent to Models 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

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

and

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

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 included as a variable in the regression because a higher number of candidates means smaller vote shares (on average), and smaller vote shares exhibit smaller variances. We also re-estimate regressions (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

This yields a total of twelve regressions, whose results are reported in Table 3. Consistent with our expectations, residual variance is decreasing with the number of prior 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 this expectation, we find that the coefficient of $PriorSignals_{j,s,y}^2$ is positive in all specifications. However,

²³The residuals of the other Model 5 give similar results.

²⁴This adjustment is exact when no covariates are used.

it is generally 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 of the joint test of both parameters is always 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 this variability reduction effect due to learning from earlier election results is statistically significant and exhibits the expected diminishing pattern, it is quantitatively small relative to other factors: it explains only about 7 percent of the residual variance, on average (though the percentage can be over 12 percent on some specifications). Evidently, there are several other determinants of vote share variability, including the type of information shocks that lead to learning about candidate valence in the first place, and possible co-ordination of voters across states voting simultaneously.²⁵

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

To test whether this alternative explanation is the driving force behind the reduction of share variability, note that it implies that vote share residuals for candidates in the same political position should be strongly negatively correlated and largely cancel out. Vote share variability at the political position level, controlling for candidate mean shares, should not have a clear trend over time. We test this prediction by summing the vote share residuals of candidates in the same political position in a particular state contest. We then perform the same analysis described in equations (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 only differ 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,

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

measured at the position level, declines for later contests. Statistical significance is affected when both the number of signals and the number of signals squared are used as regressors; however, the two variables remain jointly statistically significant. The explanatory power of these regressions actually increases somewhat compared to those in Table 3. We conclude that increased coordination of voters across states voting contemporaneously is not an explanation for the reduction of share variability.

There is also another observation that supports our interpretation that the reduction in variability is due to hardening priors as more information about the candidates becomes available. This 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 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 coefficient of round is positive when used in conjunction with $PriorSignals_{j,s,y}$. These 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 Share Variability and Electorate Ideology Effects

The analysis in the preceding section provides evidence of a systematic relationship between the vote share variance and the timing of a state’s election in the overall sequence of primaries. The empirical analysis in section 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 section 3.3 via 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.

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

²⁷It 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 these models were estimated via GLS.

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. Though 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 under the 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. There is minimal difference between the GLS and OLS skedastic function estimates.

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 are again 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, where both estimates are of the same sign but only that obtained via GLS is statistically significant. The skedastic function estimates are also similar in sign and generally more statistically significant under 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 for the Republican candidates are no longer identical across the two classifications, as they are under 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, Herrera, and Polborn (2016) is consistent with 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 $\mathcal{J} = \{1, \dots, J\}$ 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, though 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, there is a binary characteristic on which candidates are exogenously fixed 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 to formalize the idea that some candidates are close policy substitutes for voters, while there is a substantial difference to 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 a victory of candidate j is

$$U_j^i = v_j - \lambda|a_j - \theta^i| + \varepsilon_j^i. \quad (12)$$

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 for candidate j , reflecting variation among candidates in other dimensions for which voters have different preferences. In this case, the fixed characteristic modeled explicitly ($a_j = 0$ or $a_j = 1$) should be understood as the most important policy dimension. 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 states from which the signals in those earlier states can be inferred, voters rationally update their beliefs. Let \hat{v}_j^s denote the valence of candidate j that is expected by voters in district s . Each voter votes sincerely. That is, voter i in district s that votes at time t votes for candidate j if and only if

$$j \in \arg \max_{j' \in \mathcal{J}^t} \hat{v}_{j'}^s - \lambda|a_{j'} - \theta^i| + \varepsilon_{j'}^i, \quad (13)$$

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

²⁸The 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, though these are accounted for in the estimation.

²⁹Since 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 this.

Deltas, Herrera, and Polborn (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

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

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

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

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 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 involved, 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, Herrera, and Polborn (2016).

Effect of drop-outs. Consider a situation in which there are initially three candidates, two of whom (say, A and B) have position 0, while the third one (C) has position 1. What happens to the support of candidates B and C, when candidate A drops out? It is useful to define the total number of voters who rank candidate A highest and candidate B second as R_{AB} ; let R_{AC} be defined analogously. In the Appendix, we show that

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

and

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

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, Herrera, and Polborn (2016). These 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 used. This theoretical result corresponds well to the empirical results reported in Section 3.1.

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 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 normal but now with expected value

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

and variance

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

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

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

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

Proposition 1 *There exists a unique vector $(0, x_2, x_3, \dots, x_k)$ 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 these 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 $\frac{\sigma_\eta^2}{\sigma_{j0}^2 + \sigma_\eta^2} c$ for each candidate, leaving the difference between the valence estimates for the different candidates, and hence the voter's voting decision, unaffected. The vote shares are determined only by the *difference* between the candidates' valences, so we can normalize candidate A 's estimated valence to zero.

Our next result, Proposition 2, shows that, as the primaries progress, the variation of beliefs about candidate valences across those states that vote at the same time diminishes. This is intuitive since 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 lower variance of the valence estimates in later states translates into a lower variance of a candidate's vote shares in late states, relative to early states. In particular, this is clear in the limit: If there is (almost) no remaining uncertainty about candidates' valences, then vote shares in late states depend only on μ^s and are otherwise completely deterministic. Any randomness in the valence estimate across late states must increase the variance of the candidates' vote shares. The prediction of Proposition 2 is borne out by the empirical results reported in Section 3.3 above.

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

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

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

Proof. See Appendix. ■

We conjecture that Proposition 3 holds more generally, for any λ , but again this is hard to prove. However, as above, we can also check that Proposition 3 holds around the estimated parameter values by Deltas, Herrera, and Polborn (2016). This result is supported by the estimates in Section 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 is splitting votes with a close ideological neighbor. The U.S. presidential primary system provides a unique opportunity to gauge the presence and size of this vote-splitting effect, because some candidates drop out during the primaries, and the voters that would have voted for a dropped-out candidate need to choose which of the remaining candidates to support. The sequential nature of the primaries also allows us to infer, using the pattern of decline in vote share variability, that voters are becoming better informed about candidate quality by observing the outcomes of earlier election rounds.

References

- Adams, J. and S. Merrill (2008). Candidate and party strategies in two-stage elections beginning with a primary. *American Journal of Political Science* 52(2), 344–359.

- Bartels, L. (1988). *Presidential Primaries and the Dynamics of Public Choice*. Princeton University Press.
- Bartels, L. M. (1985). Expectations and preferences in presidential nominating campaigns. *The American Political Science Review* 79(3), 804–815.
- Bartels, L. M. (1987). Candidate choice and the dynamics of the presidential nominating process. *American Journal of Political Science* 31(1), 1–30.
- Callander, S. (2007). Bandwagons and momentum in sequential voting. *Review of Economic Studies* 74, 653–684.
- Dekel, E. and M. Piccione (2000). Sequential voting procedures in symmetric binary elections. *Journal of Political Economy* 108(1), 34–55.
- Deltas, G., H. Herrera, and M. K. Polborn (2016). Learning and coordination in the presidential primary system. *Review of Economic Studies* 83(4), 1544–1578.
- Grossmann, M. and D. A. Hopkins (2016). *Asymmetric politics: Ideological Republicans and group interest Democrats*. Oxford University Press.
- Kenny, P. J. and T. W. Rice (1994). The psychology of political momentum. *Political Research Quarterly* 47(4), 923–938.
- Klump, T. and M. K. Polborn (2006). Primaries and the New Hampshire effect. *Journal of Public Economics* 90, 1073–1114.
- Knight, B. and P. Hummel (2015). Sequential or simultaneous elections? An empirical welfare analysis. *International Economic Review* 56, 851–887.
- Knight, B. and N. Schiff (2010). Momentum and social learning in presidential primaries. *Journal of Political Economy* 118, 1110–1150.
- Krasa, S. and M. K. Polborn (2010). The binary policy model. *Journal of Economic Theory* 145(2), 661–688.
- Poole, K. and H. Rosenthal (1985). A spatial model for legislative roll call analysis. *American Journal of Political Science* 29(2), 357–384.
- Schwabe, R. (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.

6 Appendix

Derivation of (16) and (17). To calculate the number of voters who rank candidate j highest and candidate j' second, consider first the case that both candidates j and j' have the same position, say,

$a = 0$ (i.e., $j, j' \in J_0$). A voter of type θ ranks j highest and j' second if and only if

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

Consider first the second inequality (i.e., the one that secures that j' is preferred to every candidate except j). For a given $\varepsilon_{j'}$, the second inequality in (22) is satisfied if and only if

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

Since the ε_k 's are distributed independently $N(0, 1)$, the probability that a voter of type $\theta = 0$ ranks candidate j' higher than any other candidate (except j) is

$$\prod_{J_0^s \setminus \{j, j'\}} \Phi(\hat{v}_{j'}^s - \hat{v}_k^s + \varepsilon_{j'}) \cdot \prod_{J_1^s} \Phi(\lambda + \varepsilon_{j'} + \hat{v}_{j'}^s - \hat{v}_k^s). \quad (24)$$

Turning to the first inequality in (22), it must also be true that $\varepsilon_j \geq \varepsilon_{j'} + \hat{v}_{j'}^s - \hat{v}_j^s$, which, for given $\varepsilon_{j'}$, has probability $\left[1 - \Phi(\varepsilon_{j'} + \hat{v}_{j'}^s - \hat{v}_j^s)\right] = \Phi(\hat{v}_j^s - \hat{v}_{j'}^s - \varepsilon_{j'})$, where the equality uses the identity $1 - \Phi(x) = \Phi(-x)$ for the cdf of the normal distribution.

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

$$\int_{-\infty}^{\infty} \Phi(\hat{v}_j^s - \hat{v}_{j'}^s - \varepsilon_{j'}) \prod_{J_0^s \setminus \{j, j'\}} \Phi(\hat{v}_{j'}^s - \hat{v}_k^s + \varepsilon_{j'}) \cdot \prod_{J_1^s} \Phi(\lambda + \varepsilon_{j'} + \hat{v}_{j'}^s - \hat{v}_k^s) \cdot \phi(\varepsilon_{j'}) d\varepsilon_{j'}. \quad (25)$$

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

$$\int_{-\infty}^{\infty} \Phi(\varepsilon_{j'} + \hat{v}_j^s - \hat{v}_{j'}^s) \prod_{J_0^s \setminus \{j, j'\}} \Phi(\hat{v}_{j'}^s - \hat{v}_k^s + \varepsilon_{j'}) \cdot \prod_{J_1^s} \Phi(-\lambda + \varepsilon_{j'} + \hat{v}_{j'}^s - \hat{v}_k^s) \cdot \phi(\varepsilon_{j'}) d\varepsilon_{j'}. \quad (26)$$

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

$$\begin{aligned} & (1 - \mu^s) \int_{-\infty}^{\infty} \Phi(\hat{v}_j^s - \hat{v}_{j'}^s - \varepsilon_{j'}) \prod_{J_0^s \setminus \{j, j'\}} \Phi(\hat{v}_{j'}^s - \hat{v}_k^s + \varepsilon_{j'}) \cdot \prod_{J_1^s} \Phi(\varepsilon_{j'} + \hat{v}_{j'}^s - \hat{v}_k^s + \lambda) \cdot \phi(\varepsilon_{j'}) d\varepsilon_{j'} + \\ & \mu^s \int_{-\infty}^{\infty} \Phi(\hat{v}_j^s - \hat{v}_{j'}^s - \varepsilon_{j'}) \prod_{J_0^s \setminus \{j, j'\}} \Phi(\hat{v}_{j'}^s - \hat{v}_k^s + \varepsilon_{j'}) \cdot \prod_{J_1^s} \Phi(\varepsilon_{j'} + \hat{v}_{j'}^s - \hat{v}_k^s - \lambda) \cdot \phi(\varepsilon_{j'}) d\varepsilon_{j'} \end{aligned} \quad (27)$$

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

$$\begin{aligned} & (1 - \mu^s) \int_{-\infty}^{\infty} \Phi(\hat{v}_j^s - \hat{v}_{j'}^s + \lambda - \varepsilon_{j'}) \prod_{J_0^s \setminus \{j, j'\}} \Phi(\hat{v}_{j'}^s - \hat{v}_k^s + \varepsilon_{j'} - \lambda) \cdot \prod_{J_1^s} \Phi(\varepsilon_{j'} + \hat{v}_{j'}^s - \hat{v}_k^s) \cdot \phi(\varepsilon_{j'}) d\varepsilon_{j'} + \\ & \mu^s \int_{-\infty}^{\infty} \Phi(\hat{v}_j^s - \hat{v}_{j'}^s - \lambda - \varepsilon_{j'}) \prod_{J_0^s \setminus \{j, j'\}} \Phi(\hat{v}_{j'}^s - \hat{v}_k^s + \varepsilon_{j'} + \lambda) \cdot \prod_{J_1^s} \Phi(\varepsilon_{j'} + \hat{v}_{j'}^s - \hat{v}_k^s) \cdot \phi(\varepsilon_{j'}) d\varepsilon_{j'}. \end{aligned} \quad (28)$$

Analogous conditions to (27) and (28) can be derived for candidate j being located at $a = 1$.

Equations (16) and (17) are special cases of (27) and (28). ■

Proof of Proposition 1. Existence follows by construction: Since the vector W^r is generated using the realized vector of estimated valences $(\hat{v}_j^r)_{j=1,\dots,k}$, a solution to (20) exists. Furthermore, it is clear that any vector of the form $(0, x_2, x_3, \dots, x_k) + (c, c, \dots, c)$ also satisfies (20). It remains to be shown that there cannot be a solution of the form $(0, y_2, y_3, \dots, y_k)$ with $(0, y_2, y_3, \dots, y_k) \neq (0, x_2, x_3, \dots, x_k)$. Assume to the contrary, and let \bar{k} be the candidate for whom $y_j - x_j$ is maximal. If $y_{\bar{k}} - x_{\bar{k}} > 0$, then substituting in the corresponding equation of (20) shows that candidate \bar{k} receives a strictly higher vote share than $W_{\bar{k}}^r$, a contradiction. Similarly, let \underline{k} be the candidate for whom $y_j - x_j$ is minimal. If $y_{\underline{k}} - x_{\underline{k}} < 0$, then substituting in the corresponding equation of (20) shows that candidate \underline{k} receives a strictly smaller vote share than $W_{\underline{k}}^r$, a contradiction. But then, it must be true that $y_j = x_j$ for all $j = 2, \dots, k$. ■

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

$$E\left(\frac{\sigma_v^2}{\sigma_v^2 + \sigma_\eta^2} Z_j\right) = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_\eta^2} v_j \quad (29)$$

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

$$\left(\frac{\sigma_v^2}{\sigma_v^2 + \sigma_\eta^2}\right)^2 \text{Var}(Z_j) = \left(\frac{\sigma_v^2}{\sigma_v^2 + \sigma_\eta^2}\right)^2 \sigma_\eta^2 \quad (30)$$

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

$$\hat{v}_{j0,late} = \frac{\frac{\sigma_\eta^2}{R}}{\sigma_v^2 + \frac{\sigma_\eta^2}{R}} \cdot 0 + \frac{\sigma_v^2}{\sigma_v^2 + \frac{\sigma_\eta^2}{R}} \frac{\sum Z_j^r}{R} \quad (31)$$

with a variance of (using (19))

$$\frac{\sigma_v^2 \frac{\sigma_\eta^2}{R}}{\sigma_v^2 + \frac{\sigma_\eta^2}{R}}. \quad (32)$$

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

$$\hat{v}_j^s = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \frac{\sigma_v^2 \sigma_\eta^2}{R}} \cdot \hat{v}_{j0,late} + \frac{\frac{\sigma_v^2 \sigma_\eta^2}{R}}{\sigma_\eta^2 + \frac{\sigma_v^2 \sigma_\eta^2}{R}} \cdot Z_j^s \quad (33)$$

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

the mean valence estimate in late states) is therefore

$$\left(\frac{\frac{\sigma_v^2 \frac{\sigma_\eta^2}{R}}{\sigma_v^2 + \frac{\sigma_\eta^2}{R}}}{\sigma_\eta^2 + \frac{\sigma_v^2 \frac{\sigma_\eta^2}{R}}{\sigma_v^2 + \frac{\sigma_\eta^2}{R}}} \right)^2 \cdot \text{Var}(Z_j^s) = \left(\frac{1}{1 + \frac{\sigma_\eta^2 (\sigma_v^2 + \frac{\sigma_\eta^2}{R})}{\sigma_v^2 \frac{\sigma_\eta^2}{R}}} \right)^2 \cdot \text{Var}(Z_j^s) = \left(\frac{1}{1 + \frac{\sigma_\eta^2}{\sigma_v^2} + R} \right)^2 \cdot \text{Var}(Z_j^s), \quad (34)$$

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

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

$$Z_A = \frac{\frac{dVS_A}{d(1-\mu)}}{VS_A} = \frac{VS_{A,0} - VS_{A,1}}{(1 - \mu)VS_{A,0} + \mu VS_{A,1}} \quad (35)$$

and, similarly,

$$Z_B = \frac{\frac{dVS_B}{d(1-\mu)}}{VS_B} = \frac{VS_{B,0} - VS_{B,1}}{(1 - \mu)VS_{B,0} + \mu VS_{B,1}} \quad (36)$$

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

$$\frac{VS_{A,1}}{VS_{B,1}} > \frac{VS_{A,0}}{VS_{B,0}}. \quad (37)$$

A type 0 voter prefers candidate A to B if and only if $v_A + \varepsilon_A \geq v_B + \varepsilon_B$. If λ is large, so that there are almost no cross-over voters (i.e., type 0 voters who vote for C, or type 1 voters who vote for A or B), then the ratio on the right-hand side of (37) is $\Phi\left(\frac{v_A - v_B}{\sqrt{2}}\right) / \Phi\left(\frac{v_B - v_A}{\sqrt{2}}\right)$.

Consider now the term on the left-hand side of (37). The probability that a type 1 voter ranks both A and B higher than C is exceedingly small for λ large and neglected in the following. A type 1 voter prefers A to C if $v_A + \varepsilon_A - \lambda \geq v_C + \varepsilon_C$, and thus, $VS_{A,1} = \Phi\left(\frac{v_A - v_C - \lambda}{\sqrt{2}}\right)$. Similarly, $VS_{B,1} = \Phi\left(\frac{v_B - v_C - \lambda}{\sqrt{2}}\right)$. We therefore have

$$\frac{VS_{A,1}}{VS_{B,1}} = \frac{\int_{-\infty}^{\frac{v_A - v_C - \lambda}{\sqrt{2}}} \frac{1}{\sqrt{2\pi}} \exp(-t^2/2) dt}{\int_{-\infty}^{\frac{v_B - v_C - \lambda}{\sqrt{2}}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt} = \frac{\int_{-\infty}^{\frac{v_A - v_C - \lambda}{\sqrt{2}}} \frac{1}{\sqrt{2\pi}} \exp(-t^2/2) dt}{\int_{-\infty}^{\frac{v_A - v_C - \lambda}{\sqrt{2}}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\left(t - \frac{v_A - v_B}{\sqrt{2}}\right)^2}{2}\right) dt} \quad (38)$$

Compare the integrands on the right-hand side. Note that $\left(t - \frac{v_A - v_B}{\sqrt{2}}\right)^2 - t^2 = \frac{v_A - v_B}{2}(v_A - v_B - 2\sqrt{2}t)$ is decreasing in t , and is positive for all $t \leq \frac{v_A - v_C - \lambda}{\sqrt{2}}$, provided that λ is sufficiently large (clearly, $\lambda \geq v_A - v_C$ is a sufficient condition for this). Thus, substituting the upper limit of the integral for t , the integrand in the denominator is at most $\exp\left(-\frac{v_A - v_B}{2}(v_A - v_B - 2\sqrt{2}\frac{v_A - v_C - \lambda}{\sqrt{2}})\right)$ times the integrand in the numerator, and thus the same relation holds for the values of the two integrals. Since this factor goes to zero as λ grows, (38) goes to infinity, which proves that (37) holds for λ sufficiently large. This proves the claim. ■