Polarization and State Legislative Elections*

Cassandra Handan-Nader[†]
Andrew C. W. Myers[‡]
Andrew B. Hall[§]

Stanford University

August 6, 2024

Abstract

U.S. state legislatures are critical policymaking institutions that are increasingly polarized, yet data and measurement limitations have prevented researchers from understanding how state legislative elections contribute to this polarization. To address this gap, we construct new measures of candidate ideology based on campaign contributions and roll-call votes, and we use them to offer the first systematic study of the relationship between candidate ideology and electoral outcomes in primary and general elections in state legislatures, 2000-2022. We find that the set of people running for state legislature has polarized substantially in recent decades. More-moderate candidates enjoy a meaningful advantage in contested general elections, but that advantage has declined somewhat in recent years. At the same time, more-extreme candidates are favored in contested primary elections. These new measures and data will allow researchers to build on these basic findings to understand how elections function in lower-information, lower-salience environments like American state legislatures.

Word count: 9938

*For data, the authors thank Adam Bonica, Carl Klarner, Steve Rogers, and Jim Snyder. For comments, the authors thank Avi Acharya, Justin Grimmer, Steve Rogers, Chris Warshaw, and participants of the UC Berkeley American Politics Workshop, the Stanford Political Science Workshop, the Stanford GSB Political Economy Theory Lunch, and the Princeton American Politics Seminar. Some of the computing for this project was performed on the Sherlock cluster. We would like to thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that contributed to these research results. This research was supported in part by funding from the UPS Endowment Fund at Stanford University. For extraordinary research assistance we thank: Elise Kostial, Laniah Lewis, Janet Malzahn, and Anish Sundar.

[†]Assistant Professor of Political Science, New York University.

[‡]Ph.D. Candidate, Department of Political Science, Stanford University.

[§]Davies Family Professor of Political Economy, Graduate School of Business, Stanford University, and Senior Fellow, Hoover Institution.

1 Introduction

To what extent do primary and general elections for state legislatures advantage more-moderate or more-extreme candidates? While research on American candidate ideology and electoral outcomes focuses mainly on the national level (e.g., Ansolabehere, Snyder, and Stewart 2001; Canes-Wrone, Brady, and Cogan 2002; Canes-Wrone and Kistner 2022; Hall 2015), state legislatures provide an ideal laboratory for understanding how the long-studied relationship between candidate ideology and electoral outcomes might extend to lower information, lower salience settings. Such settings are common across the democratic world, and complicate the assumptions behind spatial models of voting that underpin our ideas about the potential advantages of more-moderate candidates (e.g., Downs 1957). State legislatures are themselves highly consequential and increasingly polarized policymaking bodies (Caughey and Warshaw 2022; Rogers 2023; Shor and McCarty 2011), responsible for disbursing nearly two trillion dollars in spending and with authority over many salient policy areas including education, healthcare, and election administration.[1] In addition to being critical components of local government in their own right, state legislatures are also the main source of future House and Senate candidates (e.g., Thomsen 2014), and could therefore be helping to drive national polarization.[2]

Despite the value of these simple empirical questions about state legislative elections, they have been impossible to answer comprehensively because we lack a measure of candidate ideology for state legislatures that both a) corresponds closely to legislative polarization as measured by roll-call votes in state legislatures and b) applies to incumbents and non-incumbents alike. Existing measures available for state legislators either capture state legislative voting behavior but apply only to incumbents (Shor and McCarty 2011), or are meant to represent candidate positions in an ideological space defined primarily by candidates for

[1] `https://www.urban.org/policy-centers/cross-center-initiatives/state-and-local-finance-initiative/state-and-local-backgrounders/state-and-local-expenditures`.

[2] For arguments for why a more-extreme state legislative candidate pool could cause an important part of the rising polarization of Congress, see Hall (2019) and Thomsen (2017).

federal office and subsequent voting behavior in Congress (Bonica 2014, 2018). As a result, previous work specific to state legislative polarization has largely been limited to studying incumbents. Two important recent books show that more-moderate incumbent state legislators modestly outperform more-extreme incumbents in general elections (Caughey and Warshaw 2022; Rogers 2023).[3] But the relationship between ideological positions and electoral performance may be different for challengers and candidates running for open seats who do not have the other advantages that incumbents possess. Moreover, to assess whether the advantages of more-moderate or more-extreme candidates have changed over time, we need to make apples-to-apples comparisons of candidates across years, which is not possible if we have to condition our sample only on the incumbents who have survived in each cycle to face another reelection.

The primary contribution of this paper is to offer new data and measures that overcome these obstacles, allowing researchers to study candidate ideology, polarization, and elections in state legislatures in ways not previously possible. We develop two new measures of state legislative candidate roll-call ideology that apply to incumbents and non-incumbents alike. As a baseline model, we use the approach developed in Hall and Snyder (2015) that imputes candidate NP-Scores—a widely used measure of state legislator's roll-call-based ideology from Shor and McCarty (2011)—using the weighted averages of the NP-Scores of incumbents to which that candidate's donors also donated. To improve predictive acccuracy, we build off of the ideas in Bonica (2018),[4] and use a machine-learning based approach that predicts NP-Scores using campaign donation records. Though they build off prior methodological approaches in the literature, these measures are developed specifically to study state legislative elections in three ways. First, by using a comprehensive source of state contribution records, our measures incorporate more donor information from state candidates' campaigns than measures that must rely on donors that donate to both federal and state

[3]For earlier work on the relationship between incumbent ideology and electoral outcomes in state legislatures, see Birkhead (2015); Hogan (2008); Rogers (2017).

[4]A similar machine learning approach is also used in Bonica and Li (2021).

campaigns to score state candidates (Bonica 2014, 2018). Using these data, our models are able to take advantage of cross-state giving to improve predictive accuracy for smaller states. Second, we train separate machine learning models for each party to improve the measure's ability to distinguish between candidates within partisan primaries. Third, we develop our models to achieve as much parity in accuracy as possible between winners and losers of elections by using only contributions received before a candidate first takes office, helping to allay concerns that campaign-finance-based scalings could partially be a function of having won office previously. Both of the resulting measures correlate highly with NP-Scores, even within party.

In collaboration with Fouirnaies and Hall (2020) and Rogers (2023), we also construct a new dataset on state legislative primary elections, collected and digitized from each state's official records, and extensively cleaned and standardized. We merge this with data on general elections from 2000 through 2022 and combine it with CFscores and our candidate ideology scores to form a dataset containing the estimated ideological positions and primary- and general-election performances of nearly 48,000 candidates for state legislative office. The resulting dataset, including our new measures of candidate ideology, will be made publicly available so that researchers can use them freely for the study of state legislative elections.

With this new data, we first show that the polarization of the whole set of candidates seeking state legislative office has risen dramatically over the past two decades. The growing polarization of state legislators tracks the polarization of the set of candidates running for office quite tightly. We argue that who runs for state legislature may therefore be very important for understanding state legislative polarization, despite the focus of existing research on incumbent positioning, and may therefore be important for explaining polarization at the federal level, too. If the entire pipeline of candidates seeking state legislative office is polarizing, this will increase the polarization of congressional candidates, too.[5]

[5]In fact, Phillips, Snyder, and Hall (2024) shows that the general trend of state legislators becoming more extreme is a more important part of the explanation for the polarization of the congressional candidate pool than changes caused by redistricting or electoral competitiveness.

Next, using a panel design that compares over-time changes in the ideological midpoint between candidates within a given district, we show that contested general elections have favored more-moderate candidates, on average. This result is consistent with canonical spatial models despite the low levels of information in state legislative elections. However, this advantage is relatively modest in magnitude and appears to have declined noticeably in recent years. While there may be many explanations for this decline, it is at least consistent with the growing literature on the nationalization of state legislative elections (Abramowitz and Webster 2016; Hopkins 2018; Rogers 2016, 2023).[6] We bolster these findings with a regression discontinuity design on close primary races, where a more-extreme candidate in the general election is arguably as-if randomly assigned, and find a modest but meaningful vote-share penalty for the more-extreme candidates in the general election.

Next, we take advantage of our large dataset to study important sources of variation in these effects. Spatial models of voting, along with prior empirical work, identify at least three important factors that might increase or decrease the advantage of more-moderate candidates: the degree of competition in the district, which we measure using partisan presidential vote share; the level of legislative professionalization, which makes elections more salient to voters and candidates; and the presence of other, more salient offices at the top of the ballot, which we proxy for with on-cycle vs. off-cycle elections. We find that the advantage to more-moderate candidates is higher in more-competitive districts, yet surprisingly, we find a relatively precisely estimated non-difference between on-cycle and off-cycle elections, and only weak and very imprecise evidence that more professionalized legislatures feature larger advantages. These results suggest that the more important factor in boosting more-moderate candidates and reducing polarization is electoral competition—which has declined in recent decades as more districts have become especially partisan—

[6] On the other hand, despite these patterns of nationalization, meaningful amounts of split-ticket voting still occur in state legislative races (Kuriwaki 2023), especially where information is higher (Moskowitz 2021). It also remains a puzzle why there has been a more substantial advantage to more-moderate candidates in the early 2000s, when partisanship was still important and voter information was presumably still low in state legislative elections.

rather than the timing of elections or the overall level of importance of the state legislature itself.

A related literature, again focused on the national level, posits a push and pull between primary and general elections, with primary elections favoring more-extreme candidates while general-elections favor more-moderate candidates (e.g., Aranson and Ordeshook 1972; Brady, Han, and Pope 2007; Hall 2015). In the final part of the paper, we speak to this literature by offering the first comprehensive estimates of the relationship between candidate ideology and electoral outcomes in contested primary elections. Consistent with this theoretical literature, despite the low levels of information in state legislatures, we find that these elections favor more-extreme candidates, on average, and this advantage has remained large in recent years. These findings build from the analyses presented in Rogers (2023), which shows that incumbents with more-extreme roll-call votes in state legislatures enjoy a modest advantage in primaries.

Taken together, our estimates paint a picture of a changing state legislative system in which more-extreme candidates are increasingly seeking office, face limited competition, are favored in primary elections, and face relatively small and diminishing penalties in the general election. This pattern shows how state legislatures have polarized over the past two decades, and also helps to explain why the set of people running for Congress has also polarized so much over this same time period. There are many possible explanations for why the state legislative electoral system has evolved in this manner, including changes to the media environment, to the structure of districts and elections, to American political culture, and to the policy agenda facing state legislatures. In the conclusion to the paper, we discuss how our findings help to set up future research on these important questions.

2 Campaign-Finance Based Measures of Candidate Roll-Call Ideology

To assess the electoral roots of roll-call based polarization, we first need a measure that closely captures how both winning and losing candidates would cast roll-call votes in state legislatures. However, no existing measures of ideology that extend to candidates for state office are optimized explicitly for capturing roll-call voting behavior in state legislatures. Alternative measures such as CF-Scores (Bonica 2014) and DW-DIME scores (Bonica 2018) map state candidates to an ideological space defined primarily by federal contributions, and therefore must rely on the subset of donors that donate to both state and federal campaigns to estimate the positions of state candidates. While useful for comparing state and federal candidates on the same scale, these measures only incorporate a small number of state candidates' campaign donors by design, and, when optimized specifically to predict roll-call voting in Congress, extend to a very limited number of state candidates (see Appendix A.1, Table A.1). Moreover, these measures incorporate donations from both before and after a candidate wins election, which makes candidates' estimated ideological positions dependent on their past electoral successes or failures. Hence, our goal in this section is to build new measures that are highly predictive of candidates' subsequent roll-call voting behavior in state legislatures within each party, and are tailored specifically to studying the relationship between ideological positions and winning elections.

2.1 Using Campaign Finance Records to Predict NP-Scores

We begin with the key target variable that we want to predict, the ideological mappings for state legislators from Shor and McCarty (2011), called NP-Scores. The most recent version provides scores for 27,629 state legislators between 1993 and 2020.[7] These mappings

[7] To obtain some coverage for 2021-2022 elections, we carry forward the NP-Score for incumbents with scores in the 1993-2020 dataset. The April 2023 release of the data was downloaded via https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/NWSYOS.

are the result of projecting a measure of voting behavior from each state legislature onto a measure of legislator responses to the Project Vote Smart National Political Awareness Test (NPAT). The methodology uses a state-specific OLS regression of a one-dimensional NPAT-based scaling onto a one-dimensional roll-call based scaling using legislators who have both scores available. Therefore, voting disagreement between legislators within a state on the first roll-call dimension is amplified the more it correlates with the NPAT's first dimension.

Though NP-Scores are the most widely used measure of roll-call-based polarization at the state level, they also have some inherent limitations. First, though the NPAT is designed to measure nationally relevant ideological issue positions, voting behavior in legislatures can be driven by party control in addition to individual ideological leanings (Lee 2009; Stiglitz and Weingast 2010). As in other studies of roll-call based polarization (e.g., Poole and Rosenthal 2007), we use "roll-call ideology" as a shorthand for the mixture of ideological and partisan factors that might drive roll-call-based polarization. Second, legislators' positions on certain local issues may not align well with their positions on nationalized issues (Anzia 2021), which could cause the measure to be insensitive to polarization on important local issues in some states. Polarization should therefore be interpreted relative to the national issue space as captured by the NPAT's first dimension, with the understanding that states less polarized on national issues could still be polarized on issues that are not relevant to the two-party axis that characterizes the NPAT's first dimension. Finally, NP-Scores are static by construction throughout a legislator's career, which means they cannot be used to analyze behavioral changes that individual members may make over the course of their careers.

Since these NP-Scores are only available for legislators who won election, we need another set of information to help us predict scores for people who have not, and may not ever, serve in office. While there are many potential data sources one might use for this purpose—such as the text of candidate speeches or behavior on social media—we build on the supervised learning approach of Bonica (2018) to estimate scores for both incumbents and challengers running for state office using donations to their campaigns. We obtain campaign donations for state legislative candidates from the National Institute on Money in Politics (NIMSP), which digitizes and standardizes information from campaign finance reports for all state-office candidates.[8] The data consist of nearly 30 million transactions for state legislative elections between 1989 and 2022, with comprehensive election coverage beginning in the year 2000. We merge this donation information at the legislator-election-year level to the NP-Scores to obtain a unified set of predictors and outcomes for 20,757 state legislators in at least one election, 11,739 of which have enough donations before they won office to enter the training set. Though our research design only makes within-state comparisons (i.e., between two candidates in the same election), our approach to the data was designed to reduce the variance in predictive accuracy between states as much as possible. To impose a unified threshold on data sparsity, we only produce scores for candidates that received money from at least 5 donors who also gave to at least 5 legislators with an NP-score, and to improve predictive accuracy for states with less data, our models take advantage of donors that give to candidates in multiple states to pool information across states.[9] With this approach, we are able to substantially increase the total number of donors that contribute to a candidate's score relative to the federal approach (median of 16 donors vs median of 2 donors).[10]

Though our basic supervised learning approach follows Bonica (2018), we further customize the prediction problem based on the downstream empirical analyses we wish to conduct. First, because our analysis hinges on producing credible ideology estimates for candidates that do not win office, we develop our models only using donations received *before* candidates win state legislative office for the first time. This avoids biasing the predictive models with information following electoral victories when studying performance in elections (Hall and Snyder 2015), and improves the generalizability of the models to losers of elections

[8]See `https://www.followthemoney.org/our-data/about-our-data` for more information on the data. See Appendix A.3 for details on our validation of this dataset.

[9]22% of the donors in our modeling data gave to candidates running in multiple states, and we found that pooling information across states reduces the mean squared prediction error by 38%. Appendix A.2 reports the average proportion of out-of-state donors and contributions per candidate by state.

[10]We show in Appendix A.3 that this is not an artifact of differences in donor identity resolution between the DIME database (Bonica 2023) and NIMSP.

by mimicking the information set that donors have about candidates whom they have not yet observed in state legislative office. As a result, our predicted NP-Scores are static over a legislators' career, matching the construction of NP-Scores, and do not incorporate any donation information from elections in which candidates ran as incumbents. Second, in order to make the assumption that prediction error is equal in expectation between winners and losers more plausible, we assign all legislators in the training data an "out-of-sample" score based on a model that was not trained on their particular donations in any capacity. On top of the cross-validation procedures employed in Bonica (2018), the out-of-sample scores provide an additional guarantee that the predicted scores are not overfit to the training legislators. Third, for our machine learning approach, we train separate models within each party to maximize our ability to distinguish the positions of candidates within partisan primaries.

Since we study primary as well as general elections, we also calculate versions of our scores that scale legislators using only *primary* donations from before they won office. These scores are noisier and apply to fewer legislators, but do not change the substantive findings, so we report the main results using these primary-only scores as a robustness check in Appendix A.7 and A.8.

Baseline model: contribution-weighted averages

Our baseline model follows the methodology of Hall and Snyder (2015) and McCarty, Poole, and Rosenthal (2006) by predicting candidates' NP-Scores using the contributed-weighted average NP-Score of the candidates to which their donors contribute.[11] This straightforward method has been employed in previous work on Congressional candidates (Hall and Snyder 2015; Hall 2015; Hall and Thompson 2018; Hall 2019), and builds on an extensive literature that uses campaign contributions to scale candidates without machine learning (e.g., Bonica 2013, 2014, 2018; McCarty and Poole 1998; McCarty, Poole, and Rosenthal 2006; McKay 2008, 2010; Poole and Romer 1985).

[11] A similar method is also used in Caughey and Warshaw (2022).

Drawing on the NP-Score and campaign contributions data, we estimate predicted NP-Scores for candidates in two stages. First, we estimate a preference score for all state legislative donors as the average contribution-weighted NP-Score of the incumbents to which a donor contributes. More formally, let $\boldsymbol{X}$ be an $m \times n$ matrix of campaign contributions, where $\boldsymbol{X}_{ij}$ is the donation amount from donor j to candidate i, and y_i is incumbent i's NP-Score. Then donor j's revealed preference z_j is given by

$$z_j = \frac{\sum_{w \neq i} y_w \; \boldsymbol{X}_{wj}}{\sum_{w \neq i} \boldsymbol{X}_{wj}}, \tag{1}$$

where we leave out candidate i when estimating donor j's preferences to avoid a feedback loop.[12]

The second step is to impute candidate NP-Scores from the preferences of their donors. Specifically, we calculate each candidate's predicted score as

$$\hat{y}_i = \frac{\sum_j z_j \; \boldsymbol{X}_{ij}}{\sum_j \boldsymbol{X}_{ij}}. \tag{2}$$

For the remainder of the paper, we will refer to these contribution weighted-average scores as "HS scores" (for Hall and Snyder).[13]

Machine learning extension: random forest regression

We improve upon the predictive accuracy of our baseline model with a machine learning (ML) approach. It is important to note that, while the ML-based scores can provide the best within-party predictions of NP-scores, they do carry some costs. For one thing, ML-based scores may contain difficult-to-understand biases, since the process trades off bias in exchange for reducing variance in the prediction problem. Depending on the nature of these

[12]Note, however, that our results are substantively identical if we include candidate i when estimating donor j's ideology (i.e., $\boldsymbol{z}_j = \frac{\sum_w y_w \; \boldsymbol{X}_{wj}}{\sum_w \boldsymbol{X}_{wj}}$).

[13]In Appendix Figure A.9, we show that we recover substantively identical Hall-Snyder scores when using an indicator for contributions rather than the actual dollar amount. The results suggest that it is the decision to donate, rather than the donation amount, that primarily drives our ideological scaling.

biases, they could affect our downstream estimates of the electoral advantage of different types of candidates. Compounding this issue, ML-based scores are inherently something of a black box, and we have only a limited ability to examine what determines the scores that different candidates receive. For that reason, we use both the baseline and ML scores in all of our election analyses, along with CFscores.

To produce the ML scores, we learn a party-specific mapping $\hat{f}_p(\cdot)$ between the donations a candidate receives before they ever take office, and their subsequent NP-score that summarizes voting behavior over their entire careers:

$$y_{i,\text{post}} = \hat{f}_p(\boldsymbol{x}_{i,\text{pre}}) + \epsilon_{i,\text{post}}, \tag{3}$$

where $y_{i,\text{post}}$ is the NP-Score for legislator i in party p that summarizes voting behavior post winning office, and $\boldsymbol{x}_{i,\text{pre}}$ is a vector of predictors for legislator i before ever winning office for the first time. We learn $\hat{f}_p(\cdot)$ using a random forest regression (Breiman 2001), an ensemble method that learns a large number of decision trees on bootstrapped samples of the training data by randomly selecting subsets of predictor variables to consider at each split of each tree, and averages predictions across trees to produce a final prediction. We choose the optimal number of predictors to select at each split through ten-fold cross-validation. As in Bonica (2018), we construct donation-based predictors using both standalone donations received from larger donors (represented as dummy variables), and summaries of all donors' preferences using the contribution-weighted average method of the baseline model. These donor summaries are constructed in accordance with the cross-validation scheme to avoid data leakage. Due to limited coverage of candidate demographic information, we include state dummies as the only non-donation based predictor (in addition to, implicitly, the party of the legislator by training separate models by party). Training data legislators are assigned their predicted score from the cross-validation round where they were not used to train the model or build the feature set. Appendix A.2 describes in more detail how we constructed the feature set, reports the results of the cross-validation exercise, and provides a summary of which features were most predictive. We learn two separate mappings for Republicans and Democrats to improve prediction accuracy within party and, hence, our ability to measure extremism within partisan primaries.

For the remainder of the paper, we will refer to these ML-based scores as "ML scores."

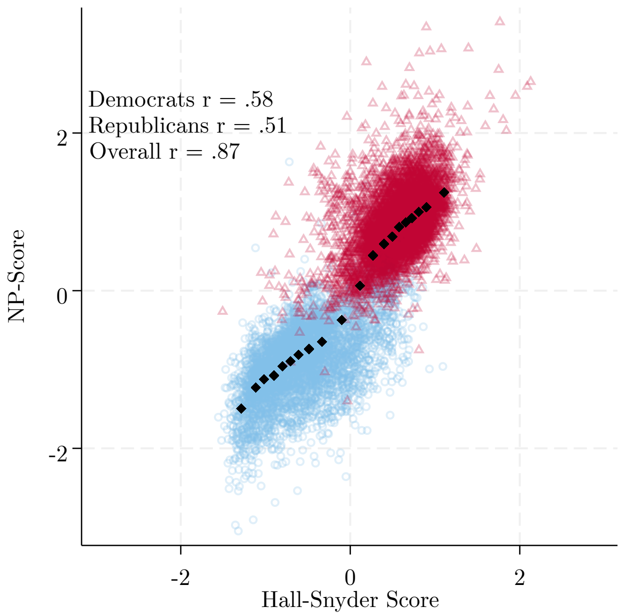
2.2 Validating the Scalings

To validate these new scalings, we conduct two empirical exercises: we compare them to observed NP-Scores for candidates who eventually take office, and we use them to predict a large dataset of roll-call votes cast in all 99 state legislatures from 2010 to 2022. We report the results of the first analyses below and refer the reader to Appendix A.6 for results on the second analysis.

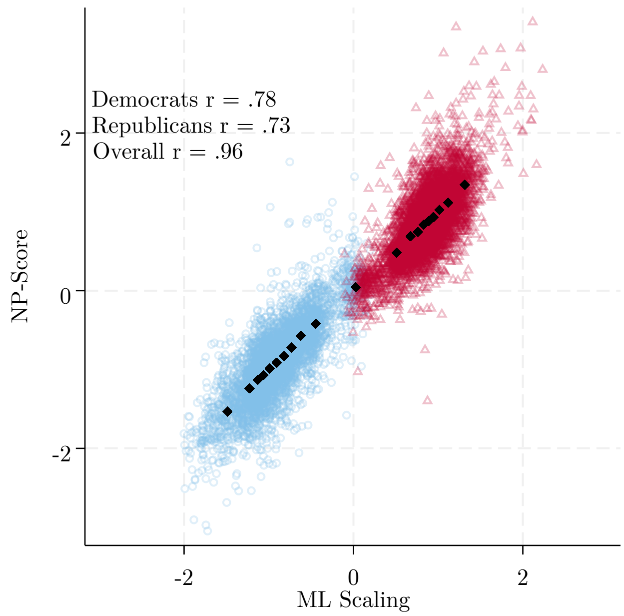
Figure 1 compares legislator NP-Scores (on the horizontal axis of each panel) to their predicted score based on four possible scalings: their Hall-Snyder scores (top left panel); their ML scores (top right panel); their static CFscores (bottom left panel); and their dynamic CFscores (bottom right panel). As the figure shows, while all four scores have relatively high overall correlations (indicating their ability to separate legislators of the two parties), both the Hall-Snyder and ML scores achieve substantially higher within-party correlations. As expected, the ML scores that use machine learning achieve the highest within-party correlations. Because not every candidate can be scored with every model, we report a balance table of candidate characteristics in Appendix A.1. We also report within-state, within-party correlations for the HS and ML scores in Appendix A.4 for completeness, though we note that static differences in predictive accuracy *between* states would not bias our downstream regression results due to the inclusion of state fixed effects. In Appendix A.5, we report that within-state, time-variant trends in prediction error are largely uncorrelated with over-time changes in state-level campaign finance regulations, and thus these types of changes are unlikely to bias our regression results.

Figure 1 – Hall-Snyder and ML Scores Correlate Well With NP-Score Scalings, Even Within Party. Figures plot roll-call-based NP-Scores against the various campaign-finance-based scalings for Democratic (circle) and Republican (triangle) incumbent legislators. Diamonds represent equal-group-size averages.

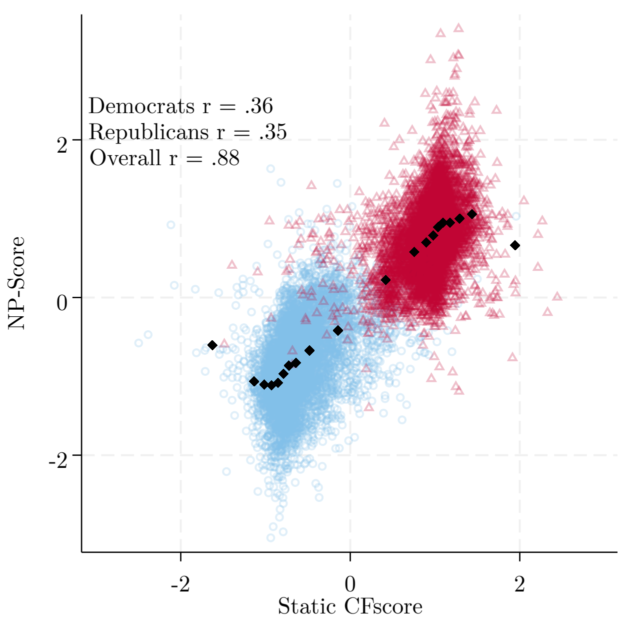
(a) Hall-Snyder Scores



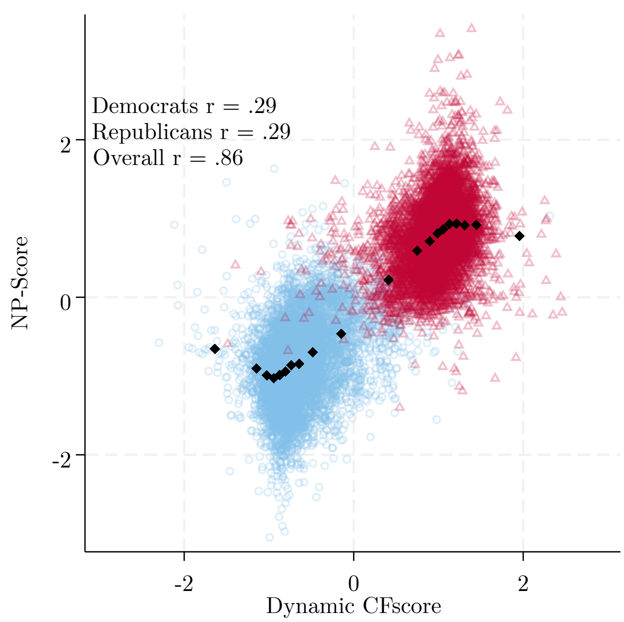
(b) ML Scores



(c) Static CFscores



(d) Dynamic CFscores



Second, leveraging a panel of 72 million raw roll-call votes, we show in Appendix A.6 that our scalings do the best job of replicating the roll-call classification success of the NP-Scores themselves

In sum, across these two exercises, we see that both of our new contribution-based scores correlate well with NP-Scores within party and predict roll-call voting effectively. This makes them useful tools for analyzing the relationship between candidate ideology and electoral performance, which we will turn to now.

3 New Data on State Legislative Elections

In order to provide a comprehensive analysis of candidate ideology and electoral performance in both primary and general elections, we assemble a new dataset of state legislative election results. We begin with the State Legislative Election Returns (SLERs) dataset from Klarner (2023) which covers all general elections in state legislatures, including full coverage of the years of our study, 2000–2022. Though election data is available before the year 2000, we study elections between 2000 and 2022 due to limited coverage of the donation data in the 1990s.

Next, we construct a comprehensive record of primary election outcomes for 2000-2022 in all relevant states. To do this, we started from partial data on 42 states for the period 2000-2014 from Rogers (2023). We added data on primaries in runoff states collected in Fouirnaies and Hall (2020). We then collected the remainder of the data—filling in gaps in the other datasets, adding the remaining states, and extending the data through 2022—from state websites, and cleaned and standardized the resulting combined dataset extensively. Overall, almost exactly 50% of the data we use was collected anew for our study, with the other half coming roughly equally from the two sources referenced above. When applicable, our primary data includes both first-round and runoff primary-election results. The complete primary dataset includes full coverage of all primary elections corresponding to general elections in our sample.

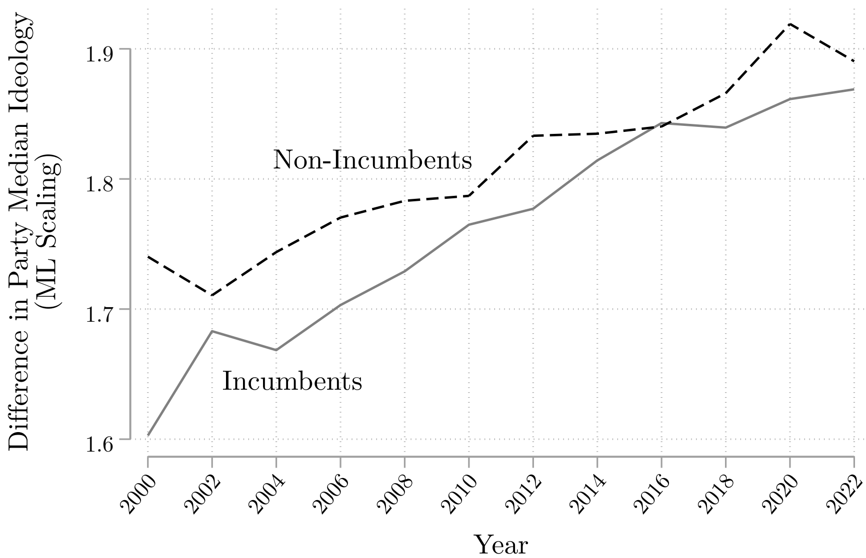
To facilitate meaningful comparisons between candidates, we restrict our analysis data along three margins. First, we focus on Democratic and Republican candidates. Second, we subset our data to include state-chamber-years for which a majority of all available seats are in single-member districts. Finally, we exclude state-chamber-years with non-conventional primary election systems (i.e. top-two and blanket primaries), all special elections, and require each election to send its winner to office for a full term.[14]

We merge the primary and general election data together into a master dataset along with the candidate ideology scores validated in Section 2. Due to the restrictions discussed in Section 2 that ensure that candidates have sufficient donation data to receive predicted scores, we are able to assign ML and HS scores to all candidates within a race in about 10,000 general elections and about 4,000 primary elections, and CFscores to all candidates within a race in about 20,000 general elections and 8,000 primary elections.

What kinds of elections enter our sample? Appendix A.1 reports characteristics of the elections included in the analysis sample compared to all elections ($n = 63,109$ generals and $n = 79,888$ primaries), all contested elections ($n = 37,335$ generals and $n = 18,362$ primaries), and all competitive elections ($n = 16,242$ generals and $n = 3,976$ primaries). Competitive elections are defined as having a winning margin of 20% or less. As Appendix A.1 shows, we are able to capture the majority of competitive races. The races we study are a little less likely to have incumbents than the overall population of competitive races but look very representative of the population in terms of partisanship (measured with Democratic presidential vote share).

[14]The latter two restrictions affect few legislators, reducing our sample by approximately .08%.

Figure 2 – Polarization of Candidates in State Legislatures Over Time, 2000-2022. Plots the absolute difference between each party's median incumbent legislator (blue line) and between each party's median non-incumbent candidate (black line), across all states, by year, as measured using ML scores. Non-incumbent includes both challengers and open-seat candidates.



4 Polarized Candidate Entry in State Legislative Elections

We first use our new data to describe the ideological positions of the people who run for state legislature over time. With relatively low rates of electoral competition, who runs for office becomes especially important in determining the polarization of state legislatures. Figure 2 plots the difference in the median candidate's ideology for each party, using the ML scores, over time. The plot shows separate lines for the entire set of new candidates in each cycle (i.e., all non-incumbent candidates), and for sitting legislators (i.e., incumbents). To keep the plot easily readable, we omit odd-year elections from it.

As the figure shows, we see a steep increase in the polarization of candidates over time; as legislative polarization has increased, so, too, has the polarization of the set of people running for office in the first place. The figure also suggests that, though incumbents are less polarized than non-incumbents throughout the study period, the gap between incumbents and non-incumbents appears to have narrowed since 2010. This suggests that, in addition to a steady increase in polarization among who runs for office over the past two decades, there has also been a shift in electoral selection, from a system that weakly favored more-moderate candidates from among the set of candidates to a system that is indifferent between more-extreme and more-moderate candidates. We will formally document this pattern in the analyses below.

5 General Elections and the Advantage of More-Moderate Candidates

Do contested general elections in state legislatures favor more-moderate candidates, and if so, how much?

This is a classic question in the study of American elections going back to the foundational "median voter theorem" and related ideas about spatial voting explored in Downs (1957), among others. A long debate in political science rages over whether there is, in reality, any advantage to more-moderate candidates as predicted in the spatial model. Behavioral research often argues that voters are not sufficiently informed about candidates and do not have a sufficiently sophisticated view of ideology to favor more-moderate candidates (e.g., Achen and Bartels 2016). Research on federal elections, in contrast, consistently documents an electoral advantage for more-moderate candidates (Ansolabehere, Snyder, and Stewart 2001; Canes-Wrone, Brady, and Cogan 2002; Canes-Wrone and Kistner 2022; Hall 2015), perhaps because there is a critical mass of informed swing voters, or because campaigns, interest groups, parties, the media, and other elite actors are able to help voters coordinate on more-moderate candidates in the absence of widespread individual-level sophistication. State legislatures present an even stronger challenge to the prediction that more-moderate candidates should be advantaged electorally, however; voters are significantly less informed about state legislators than about national-level politicians (Rogers 2023), campaigns are significantly less resourced (e.g., Fouirnaies and Hall 2014), and there is less media coverage of these races (Moskowitz 2021). Given these contrasting predictions, it is valuable to see what the data can tell us.

Existing estimates of the advantage to more-moderate candidates in state legislatures do not fully answer this question, because they are forced to focus on incumbents and cannot account for challenger positioning or study open-seat races. To answer this question using our new measures, we first follow the "midpoint" method of Ansolabehere, Snyder, and Stewart (2001). For each contested election, we compute the distance in ideology between the Democrat and Republican candidates, and we compute the midpoint between their estimated platforms. When this midpoint moves to the right while the distance between the candidates remains constant, it means that the Republican candidate has become more extreme and the Democratic candidate has become more moderate, and vice versa when the midpoint moves to the left while the distance remains constant.

To implement the midpoint method, we estimate regressions of the form

$$Y_{ict} = \beta_1 Midpoint_{ict} + \beta_2 Distance_{ict} + X_{ict} + \gamma_i + \delta_t + \epsilon_{ict}, \tag{4}$$

where Y_{ict} represents the Democratic vote share in district i in chamber c at time t. The vector X_{ict} stands in for an optional vector of control variables, and γ_i and δ_t stand in for district-regime and time fixed effects.

The quantity of interest is β_1, which captures the association between how moderate the Democratic candidate is (when the midpoint between the two candidates shifts right while holding the distance between them equal) and Democratic electoral outcomes. In the original Ansolabehere, Snyder, and Stewart (2001) approach, the unobserved district median voter's preferences are held constant by controlling for presidential vote share in the district. Because presidential vote share is not widely available for all state legislative districts, our

Table 1 – Advantage of More-Moderate Candidates in Contested General Elections, 2000-2022.

	Dem Vote Share			
	(1)	(2)	(3)	(4)
ML Scaling	0.16	0.12	0.13	0.11
	(0.02)	(0.02)	(0.03)	(0.03)
Hall-Snyder Score	0.26	0.19	0.22	0.23
	(0.02)	(0.02)	(0.04)	(0.03)
Static CFscore	0.37	0.25	0.43	0.37
	(0.02)	(0.02)	(0.04)	(0.02)
Dynamic CFscore	0.38	0.25	0.40	0.38
	(0.02)	(0.02)	(0.04)	(0.02)
District-by-Regime FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Controls for Primary Contributions	N	Y	N	N
Only races with below-median contribution gap	N	N	Y	N
Only races with ≥ 10 primary donors per candidate	N	N	N	Y

Each cell in this table reports the coefficient on *Midpoint* from Equation 4 which is scaled to run from 0 (most liberal) to 1 (most conservative) for each scaling. Robust standard errors are clustered by district-regime in parentheses.

preferred specification uses year fixed effects and district fixed effects generated separately for each redistricting period. As a robustness check, for the districts where presidential vote share is available, we report substantively similar results in Appendix A.7.[15]

Table 1 presents the results. Each cell in the table reflects a different estimate of β_1, capturing the relationship between candidate moderation and vote share. The rows show the estimates for different candidate ideology scalings, while the columns are for different regression specifications. The first column is our baseline midpoint specification in Equation 4 using year and district fixed effects (within a districting regime). Because our scalings are not perfect predictors of roll-call voting but rather rely on campaign contributions to estimate candidate positions, columns (2)-(4) subsequently add in different ways to control

[15] We have presidential vote share data for 70% of the state legislative races in our sample.

for possible differences in prediction error between competing candidates due to disparities in the amount of money they raise. In all cells, the midpoint variable is scaled to run from 0 in the race with the leftmost midpoint to 1 in the race with the rightmost midpoint, so that β_1 reflects the predicted change in Democratic vote share for the maximal shift in the midpoint observed in the sample. While we prefer to focus on the first two rows which use our preferred scalings for these analyses—ML and Hall-Snyder scores—we also present estimates for static and dynamic CFscores (rows three and four) in order to offer a point of comparison to previous work. Because CFscores are not designed to predict NP-Scores, and because they pool contributions before and after successful candidates win office, we do not rely on these estimates for our main results.

Looking down the rows for our baseline specification in column 1, we see that we find a consistently positive coefficient, indicating an advantage for more-moderate candidates. The estimates using static CFscores and ML scores are roughly half the size of those using HS scores and dynamic CFscores, but are directionally similar. In column 2, we add controls for the total contributions raised in the primary by each candidate, in logs. We focus only on primary contributions in order to avoid a sort of "post-treatment" bias that might occur where a candidate's ideology both affects their ability to raise money in the general and affects their electoral outcome—such as if a more-moderate candidate is able to raise more money in the general election.[16] These controls are helpful for making sure that our results are not driven by any possible linkage between raising more money and being erroneously scaled as more moderate. As we see, with this control included, all the estimates shrink from their prior magnitudes but remain positive. These estimates are corroborated in columns (3) and (4), which use different ways to control for the same source of confounding due to potential disparities in fundraising between competing candidates.

[16]In Appendix A.7, we re-estimate Table 1 using the ML scores that only use primary donations as a robustness check for possible "post-treatment" bias in the scores themselves, with substantively similar results.

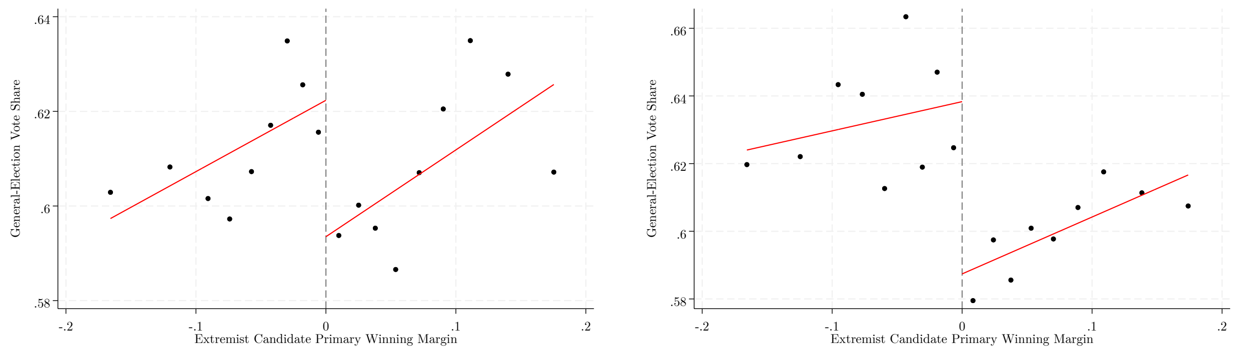
As further robustness checks against confounding due to scaling prediction error, we also report substantively similar results in Appendix A.7 when removing states for which the within-state correlations between the ML and NP-scores are especially low, and when removing states for which Shor and McCarty (2011) report high vote prediction error in the NP-Scores themselves. In Appendix A.9, we show that the midpoint coefficient shrinks in magnitude but remains positive for both Hall-Snyder and ML scores as we increase the number of donors each candidate must have to be included in the analysis.

In general, looking across all of the specifications, we see strong evidence for a positive overall advantage to more-moderate candidates. If we focus on the ML scores estimate in column 2 as our best single estimate, we estimate that shifting from the most-extreme Democratic candidate to the most-moderate predicts a 12 percentage-point increase in vote share. Based on the standard deviation of the midpoint variable in this sample, a one standard-deviation increase in the midpoint variable would predict a 1.56 percentage-point increase in Democratic vote share. This is not nothing, and could certainly matter in a close election, but it does not seem like a very large advantage. For comparison, using the same midpoint approach, Hall (2019) estimates that shifting from the leftmost to the rightmost midpoint corresponds to a 30 percentage-point increase in Democratic vote share, an advantage that is more than double this estimate for state legislatures.

5.1 Regression Discontinuity

The midpoint approach used above has the advantage of using all of our data on contested general elections where we are able to scale both candidates. However, as we discussed, it requires being able to hold fixed the unobserved preferences of the district, which we do either using fixed effects or by controlling for presidential vote. Neither of these is a silver bullet; if districts' political preferences change within redistricting cycles substantially, the fixed effects would fail to capture these trends. A similar issue occurs with presidential vote since it is not observed every year.

Figure 3 – Effect of Extremist Nominee on General Election Vote Share in U.S. State Legislatures, 2000-2022.



(a) ML Scores **(b)** HS Scores

Hall (2015) provides an alternative way to hold fixed the preferences of the district by focusing on close primary elections between a more-extreme and a more-moderate candidate, with the idea that this approximates a natural experiment in which the district "randomly" receives one type of candidate or the other. To the extent this natural experiment is valid—an assumption for which we provide evidence below—then the districts that just barely nominate a more-extreme candidate will be otherwise just like those that nominate a more-moderate candidate, on average, including in their overall political preferences. For each contested primary and general election in which we are able to scale at least two candidates, we compute the estimated ideological distance between the top two vote-getting candidates.[17] To focus on cases where there is a meaningful ideological distance between the more-moderate and the more-extreme candidate, we then restrict the data to cases where the distance between these top two candidates is at or above the median distance across all cases. We follow standard approaches to estimate the "jump" at the discontinuity that occurs when the more-extreme of the two candidates just barely switches from losing the primary to winning it.

Figure 3 shows the results graphically. As can be seen in both panels, when the more-extreme candidate goes from just barely losing the primary (left side of each plot) to just

[17]Ideally, we would construct the ideological distance measure using only primary donations, since general donations are post-treatment in this setting. However, due to the sparsity of primary donations for non-incumbents, the primary donation measure is too noisy to cleanly estimate the RD, so we rely on the measure that uses both primary and general donations.

barely winning (right side of each plot), the party's general-election vote share drops noticeably. The size of this drop is meaningful but not huge.

We estimate the size of this drop formally using standard approaches including the optimal bandwidth approach of Calonico, Cattaneo, and Titiunik (2014). Table 2 presents estimates for four different specifications and all four possible scaling approaches. In the first column, we focus on data in a 10 percentage-point window around 50/50 and use only a linear specification of the running variable. In the second and third columns, we include all of the data and use either a third-order or fifth-order polynomial specification of the running variable. Finally, in the fourth column, we use the automated procedure from Calonico, Cattaneo, and Titiunik (2014).

Looking across the first row, we see that the estimates using the ML scores range from a 2 percentage-point penalty to a 5 percentage-point penalty. These estimates grow modestly with Hall-Snyder scores and CFscores but remain relatively stable across specifications. Looking across the estimates, we find strong evidence for a modest penalty to extremist nominees. While a 2-5 percentage-point penalty in vote share is enough to tip close elections, it is small enough to not matter in many cases, too. In comparison, Hall (2019) uses the same RD setup with Hall-Snyder Scores and estimates an 8 percentage-point effect on vote share.

As is standard with RD analyses, in Appendix A.10 we show that there is no evidence for sorting or for an imbalance that would contribute to these negative estimates. Also in Appendix A.10, we investigate how our estimated effect varies as we change the minimal ideological distance between candidates that is required for a race to enter our sample (i.e., the cutoff). As the figure shows, we find that the estimated penalty to nominating an extremist increases noticeably as we increase the cutoff and thereby focus on more "intensive" treatments in which the more-extreme candidate is farther away from the more-moderate candidate.

Table 2 – Effect of Extremist Nominee on General Election Vote Share, U.S. State Legislatures 2000-2022.

	Party Vote Share			
	(1)	(2)	(3)	(4)
ML Scaling	-0.04	-0.04	-0.05	-0.02
	(0.02)	(0.01)	(0.02)	(0.02)
Hall-Snyder Score	-0.05	-0.05	-0.06	-0.06
	(0.02)	(0.01)	(0.01)	(0.02)
Static CFscore	-0.03	-0.03	-0.02	-0.03
	(0.01)	(0.01)	(0.01)	(0.01)
Dynamic CFscore	-0.02	-0.02	-0.01	-0.02
	(0.01)	(0.01)	(0.01)	(0.01)
Polynomial	1	3	5	CCT
Bandwidth	.10	-	-	-

Note: Each cell in this table reports the coefficient on *Extremist Primary Win.* Robust standard errors are reported in parentheses.

6 Variation in the Advantage of More-Moderate Candidates

So far, we have established that there is a modest, positive advantage to more-moderate candidates in contested general elections in state legislatures. This estimate pools across roughly 20 years of data and across 47 states. We can learn more about the roots of this modest advantage now by exploring where and when the advantage is larger and smaller.

6.1 Declining General-Election Advantage to Moderates Over Time

First, we explore whether the advantage to more-moderate candidates in contested general elections has changed in recent years. As state legislatures have polarized and elections have nationalized, we might suppose that the advantage to more-moderate candidates has gone down in state legislatures. This would be consistent with the argument advanced in Rogers (2023, 2016) that voters in state legislative elections are highly partisan, focus on national races at the top of the ballot, and rarely know much or anything about their state legislative candidates. On the other hand, decades of research at the federal level shows how campaigns, interest groups, parties, the media, and other elite actors can structure elections such that more-moderate candidates are favored even if most voters are unaware of candidate positions. As such, it is not clear whether the advantage to more-moderate candidates has actually declined or not in state legislatures; we need to examine the data directly.

As we discussed in the Introduction, making over-time comparisons requires having access to measures of candidate ideology for not only incumbents, but also for challengers and open-seat entrants, so that we do not confuse over-time changes in conditioning on incumbency with changes to the unconditional advantage to more-moderate candidates. Our new measures and data allow us perform this over-time comparison for the first time.

In Figure 4, we estimate Equation 4 separately for each year, for both the HS scores and the ML scores. This is only possible when we use presidential vote to control for district preferences, since our main approach uses district fixed effects that require multiple years to be pooled in order to work.

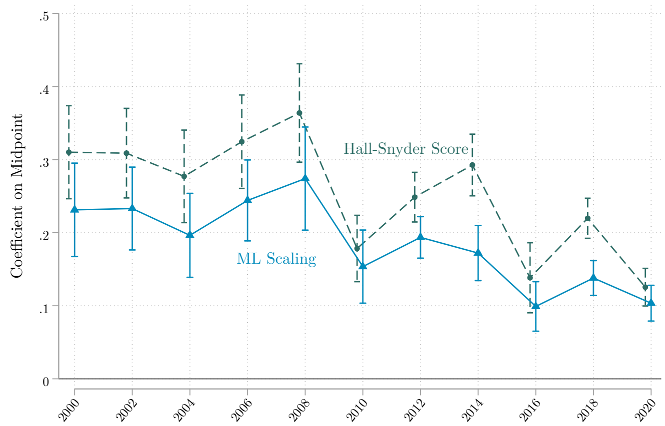
The figure reveals a relatively steady decline in the size of the coefficient on midpoint over time, indicating that the advantage to more-moderate candidates in contested general elections is shrinking. Although exactly how and when this decline has occurred varies across the two scores, the trends are very similar, and in both cases, the estimate in 2020 is the smallest of the whole time period, and indicates a quite modest advantage.[18]

6.2 Regression Results

Table 3 presents formal regressions results for the above heterogeneity tests as well as some additional ones, using the ML scores. In each column, we re-estimate Equation 4, our basic midpoint approach, and we interact the midpoint variable with a key moderator variable. Consistent with prior work on state legislatures, we focus on moderators that speak to the

[18]The year 2022 is excluded from this figure because we lack data on presidential election returns at the legislative district level for elections after the 2020 decennial redistricting process.

Figure 4 – General-Election Advantage to Moderates Over Time. This figure repots the coefficient on candidates' *Midpoint* estimated separately by year along with the upper and lower 95% confidence intervals (vertical bars). This figure uses district presidential vote share to hold the median voter constant.



nationalization of state elections (e.g., Rogers 2023, 2016), the professionalization of state legislatures (e.g., Birkhead 2015; Rogers 2017), and the degree of electoral competition (e.g., Rogers 2017), as these factors have been hypothesized to contribute to polarization at the state level.

Column 1 of Table 3 simply provides a formal test related to Figure 4. Specifically, we interact the midpoint variable with an indicator for whether the election takes place in 2012 or later. We chose 2012 because it is the first cycle that occurs after the 2010 redistricting cycle, which aligns the way we cut the data with the district-by-regime fixed effects. In the second row we see that this interaction is negative and statistically significant, indicating a decline in the advantage post 2010.

In column 2, we explore whether the advantage to more-moderate candidates varies across the types of elections: presidential elections (captured in the main effect on Midpoint in the first row of the table); off-cycle elections (the interaction in the third row of the table); and odd-year elections (the interaction in the fourth row). The results indicate that there is no interaction for off-cycle races, meaning that we estimate that more-moderate candidates have similar advantages whether or not they share the ballot with a presidential race. Odd

Table 3 – Variation in Midpoint Coefficient.

	ML Score				
	(1)	(2)	(3)	(4)	(5)
Midpoint	0.174	0.122	0.122	0.160	0.078
	(0.031)	(0.021)	(0.036)	(0.032)	(0.025)
Midpoint · Year $\geq$ 2012	-0.091				
	(0.038)				
Midpoint · Off Cycle		0.015			
		(0.009)			
Midpoint · Odd Year		-0.452			
		(0.139)			
Midpoint · Prof. (Squire dynamic average)			0.028		
			(0.096)		
Midpoint · Prof. (Squire staff)				-0.267	
				(0.099)	
Midpoint · Prof. (Squire salary)				0.155	
				(0.099)	
Midpoint · Prof. (Squire session length)				-0.054	
				(0.064)	
Midpoint · Competitive					0.042
					(0.019)
N	6,727	6,727	6,669	6,663	6,727
Controls for Primary Contributions	Y	Y	Y	Y	Y
District FEs	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y

Note: Robust standard errors clustered by district in parentheses. *Midpoint* is defined according to Equation 4 and is scaled to run from 0 (most liberal) to 1 (most conservative) for each scaling. *Professionalization* scaled to run from 0 (least professionalized state) to 1 (most professionalized state). *Competitive* districts are districts where neither party received greater than 70% of the two-party presidential vote share. *Off-Cycle* elections occur in non-presidential election years. *Distance* is included in all models but not reported in the table. Data on professionalization for Louisiana and West Virginia were not available in columns 3 and 4.

years potentially provide a more-interesting test because in the few states that hold odd-year elections for state legislature, there are often no national races at all on the ballots. Here we find a large negative interaction that is imprecisely estimated, as there are few cases that hold elections in odd years.

In column 3, we interact the Midpoint variable with Squire's (2017) measure of state legislative professionalization, as implemented in Birkhead (2015). Professionalization is scaled from 0 (least professionalized) to 1 (most professionalized). In contrast to the finding in Birkhead (2015), we find if anything a positive though imprecise interaction coefficient indicating that more professionalized state legislatures exhibit somewhat larger advantages for more-moderate candidates. In column 4, we further explore this interaction by decomposing the professionalization measure into its key constituent parts, following Rogers (2017). Consistent with the findings in Rogers (2017), we find a large negative interaction with staff size, indicating that state legislatures with larger staffs exhibit, on average, lower advantage for more-moderate candidates. Coefficients for salary and session length are smaller and noisier, roughly consistent with Rogers (2017) as well. Together, these estimates indicate that there is no simple relationship between legislative professionalization and the advantage to more-moderate candidates, possibly because, as both Birkhead (2015) and Rogers (2017) point out, professionalization bundles together both factors that make elections more salient with factors that may allow legislators to gain voters through other, non-ideological dimensions. Sussing out the precise mechanisms between legislative professionalization, candidate ideology, and polarization is beyond the scope of this study, but the new measures and data we provide should prove useful to future research in this direction.

Finally, in column 5, we show that the advantage to more-moderate candidates is stronger in competitive districts. Competitive districts are districts where neither major party averaged greater than 70% of the two-party presidential vote share across the districting regime. As the table shows, the advantage is estimated to be larger in these districts. This is consistent both with the logic of the spatial model, as well as with classic findings for the federal level (Ansolabehere, Snyder, and Stewart 2001) and previous work looking at state legislative incumbents (Rogers 2017). This is also important because it suggests that, to the extent there are fewer competitive districts in state legislatures than there used to be, this decline in competition could lead to a decline in the advantage to more-moderate candidates.

7 Primary Elections and the Advantage of More-Extreme Candidates

Having explored the links between candidate ideology and electoral outcomes in general elections, we now turn to estimating the advantage for more-extreme candidates in contested primary elections. A long literature at the federal level documents the conflict between primary and general electorates, with primary electorates thought to prefer more-extreme candidates, while general electorates are thought to prefer more-moderate candidates (e.g., Aranson and Ordeshook 1972; Brady, Han, and Pope 2007; Hall 2015). The mechanisms underlying the advantage of more-extreme primary candidates are not well understood, but may include the activation of ideologically extreme interest groups whose influence is heightened in primaries because they are lower salience (e.g., Bawn et al. 2015), as well as the differential participation of more-extreme voters in primaries (Hill and Tausanovitch 2016).

Whether this same dynamic is at play in state legislatures is unclear, given the general lack of information and competition in state legislative primaries. If the voters turning out in primaries tend to be more extreme, state legislative primaries might favor more-extreme candidates like at the federal level; on the other hand, if these voters are turning out to vote on the top-of-ballot primaries and are generally uninformed about their potential state legislative nominees, then we should not expect an advantage for more-extreme candidates. Further, there is no evidence on whether these potential effects have increased, decreased, or remained the same over time.

The one existing analysis of state legislative primary elections and candidate ideology only has access to a measure of incumbent ideology. While Rogers (2023) finds that more-extreme incumbents do somewhat better electorally in primary elections, it is unclear whether the same pattern persists in open-seat primaries, which are essential for sending new incumbents to office.

To measure how extreme primary candidates are in relation to one another, we follow Hall and Snyder (2015) and define

$$Relative\ Centrism_{ipt} = |Cand\ Ideology_{ipt} - Most\ Extreme\ Ideology_{pt}|, \tag{5}$$

where $Cand\ Ideology_{ipt}$ reflects the ideology score of candidate i running in the contested primary for party p in year t. The variable $Most\ Extreme\ Ideology_{pt}$ represents the most-extreme candidate running in primary pt, i.e., the candidate with the maximum scaling, in a Republican primary, and the candidate with the minimum scaling, in a Democratic primary. The basic idea here is to give each candidate in a contested primary a score that indicates how much more moderate she is than the most-extreme candidate in the race. This measure is better than using the simple absolute value of the scaling to measure extremism because it deals with cases where Republicans have scalings less than zero or Democrats have scalings greater than zero. This occurs with non-trivial frequency because 0 is an arbitrary value in the candidate scalings.

Armed with this measure, we then estimate regressions of the form

$$Y_{ipdt} = \beta_p Relative\ Centrism_{ik} + X_{ipdt} + \epsilon_{ipdt}, \tag{6}$$

where Y_{jpct} reflects the vote share for candidate i in the primary for party p in district d in year t.[19] The vector X stands in for a vector of covariates, which we include to address

[19]Note that the regression discontinuity design from section 5.1 does not apply to primary elections because we lack an as-if random assignment mechanism for running in a primary election. We also cannot study subsequent primary election outcomes because doing so would condition on a post-treatment variable

the fact that the relative centrism score in Equation 5 may still be sensitive to differences in the predictive accuracy of the scalings between primary districts and election years. For example, the location of the most extreme candidate within a race could vary more widely in races with comparatively less donation data, producing systematically larger swings in candidate extremism for some districts relative to others. To address this potential source of confounding, we employ two different baseline model specifications in Table 5. Column one performs a difference-in-differences in which we compare within-primary-district variation in candidate extremism over time, conditional on the number of candidates in the primary (as vote share decreases mechanically with an increasing number of primary candidates). Column two instead includes fixed effects for the specific primary election (that is, for each state-district-party-year), which makes comparisons only amongst candidates in a given race. In this latter specification we do not need to include fixed effects for the number of candidates, since it is fixed within each race. This is arguably the strongest specification, since it does not require making any cross-district comparisons, but it may be statistically noisier. Columns three and four supplement these baseline specifications with controls for each candidate's primary contributions as a further robustness check against confounding due to fundraising disparities between competing candidates, as in column 2 of Table 1.

Table 4 presents the results for contested primary elections, excluding districts where the opposing party received greater than 70% of the two-party presidential vote share.[20,21] Each cell in the table reflects the coefficient on *Relative Centrism* for a different scaling and specification of the regression. Looking at the first two rows that use our preferred scalings, we see a consistent, large, negative coefficient—indicating that more-moderate candidates

(whether the candidate seeks reelection again). This post-treatment bias could be severe since, in our sample, moderates are nearly 10 percentage points more likely to run for reelection following a close primary election than extremist candidates.

[20]In Appendix A.11 we document that all four scalings tend to underestimate extremism for candidates in very uncompetitive presidential districts, particularly for Democratic candidates, likely due to access-seeking behavior on the part of donors. The ML scores do the best at ameliorating this relationship, but for robustness we exclude very uncompetitive districts from the analysis. Our substantive findings are unchanged when adding the remaining 30% of races.

[21]In our sample, 39% of primary elections are contested, while 60% of general election races are contested.

Table 4 – Advantage of More-Extreme Candidates in Contested Primary Elections, 2000-2022.

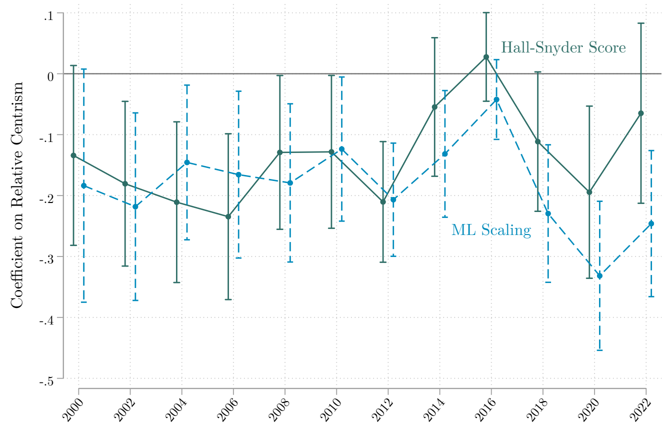
	Primary Vote Share			
	(1)	(2)	(3)	(4)
ML Scaling	-0.17	-0.17	-0.17	-0.17
	(0.02)	(0.02)	(0.02)	(0.02)
ML Scaling (Primary donations only)	-0.15	-0.15	-0.16	-0.17
	(0.03)	(0.02)	(0.03)	(0.02)
Hall-Snyder Score	-0.12	-0.12	-0.13	-0.14
	(0.02)	(0.02)	(0.02)	(0.02)
Static CFscore	0.23	0.28	0.19	0.21
	(0.04)	(0.03)	(0.04)	(0.03)
Dynamic CFscore	0.19	0.24	0.15	0.15
	(0.04)	(0.03)	(0.04)	(0.03)
District-by-Party FE	Y	N	Y	N
Party-by-Year FE	Y	N	Y	N
Number of Candidates FE	Y	N	Y	N
Race FE	N	Y	N	Y
Controls for Primary Contributions	N	N	Y	Y

Each cell in this table reports the coefficient on *Relative Centrism* from Equation 5 which is scaled to run from 0 (most extreme) to 1 (most moderate) for each scaling. The sample is restricted to contested primary elections and excludes races in districts where the opposing party received greater than 70% of the two-party presidential vote share. Robust standard errors in parentheses.

do worse, on average, in contested primaries. These results are corroborated by the static and dynamic CFscores.

Consider our preferred estimate, which uses race fixed effects, the ML scores, and no additional controls (since using primary contributions as a control is arguably post-treatment in this case). Here we estimate that going from the most moderate candidate to the most extreme candidate in a primary predicts a 17 percentage-point decrease in primary vote share. This maps to a 1 percentage-point decrease in predicted vote share for every one standard-deviation shift in relative centrism, which seems like a small but potentially important effect. In Appendix A.8, we report substantively similar results for our preferred specification when removing states with high prediction error, as in earlier robustness checks for the midpoint

Figure 5 – More-Moderate Candidates Disadvantaged in Primaries Over Time



models. In Appendix A.9, we show that the advantage to extremists shrinks in magnitude but remains statistically different from zero for both HS and ML scores as we increase the minimum number of unique donors that both candidates in a race must receive a contribution from to be included in the analysis. Finally, in Appendix Table A.10 we show that our results are of similar magnitudes after restricting the analysis to races with below-median contribution gaps between candidates and races with at least 20 donors per candidate. In sum, our results corroborate at the state level what has been found at the federal level – more extreme candidates appear to fare better in contested primary elections.

7.1 Advantage to Extremists in Primaries Persists Over Time

Finally, we can also again explore whether these relationships are changing over time. In Figure 5, we plot our estimates for the coefficient on *Relative Centrism* by year, using Equation 5 as our specification. We see a brief decline in the disadvantage for more-moderate candidates between 2012 and 2016, but in recent years, the disadvantage has returned to being statistically indistinguishable from pre-2012 levels, though the point estimate is lower.[22]

[22] In a previous version of the paper, we found that the advantage to more-extreme primary candidates increased in the 2010s relative to earlier years. With our updated data and model specifications, we no longer find that this is the case.

8 Conclusion

Understanding how state legislatures have polarized is important both because the state legislatures are themselves highly important policymaking bodies, and because they are the main pathway for candidates to Congress. In this paper, we have offered the first systematic analyses of the links between candidate ideology, electoral competition, and legislative polarization in state legislatures that cover all three stages of the process: candidate entry, primary elections, and general elections. Using new data and new measures of candidate ideology based on campaign contributions, we have established a number of empirical patterns relevant for future work on elections and polarization.

Our study is meant to be only the first key step in what must be a broader effort to understand why state legislative elections work the way that they do. Why are the people running for state legislature themselves so much more polarized than they used to be? Why are more-extreme candidates advantaged in these primaries, and why has their disadvantage in general elections decreased? At the same time, how do state legislative elections sustain *any* advantage for more-moderate candidates, when information on state legislative candidates has always been quite low? These are key questions for future research, and should be aided by the new measures and data that we have assembled to understand state legislative elections.

References

Abramowitz, Alan I., and Steven Webster. 2016. "The Rise of Negative Partisanship and the Nationalization of US Elections in the 21st Century." *Electoral Studies* 41: 12–22.

Achen, Christopher H., and Larry M. Bartels. 2016. *Democracy for Realists: Why Elections Do Not Produce Responsive Government.* Princeton University Press.

Ansolabehere, Stephen, James M. Snyder, Jr., and Charles Stewart, III. 2001. "Candidate Positioning in US House Elections." *American Journal of Political Science* 45(1): 136–159.

Anzia, Sarah F. 2021. "Party and Ideology in American Local Government: An Appraisal." *Annual Review of Political Science* 24: 133–150.

Aranson, Peter H., and Peter C. Ordeshook. 1972. "Spatial Strategy for Sequential Elections." In *Probability Models of Collective Decision Making*, ed. Richard G. Niemi, and Herbert F. Weisberg. Columbus, OH: Charles E. Merrill.

Bawn, Kathleen, Knox Brown, Angela Ocampo, Shawn Patterson, John Ray, and John Zaller. 2015. "Social Choice and Coordination Problems in Open House Primaries." Presented at the American Political Science Association Conference, San Francisco, CA, 2016.

Birkhead, Nathaniel A. 2015. "The Role of Ideology in State Legislative Elections." *Legislative Studies Quarterly* 40(1): 55–82.

Bonica, Adam. 2013. "Ideology and Interests in the Political Marketplace." *American Journal of Political Science* 57(2): 294–311.

Bonica, Adam. 2014. "Mapping the Ideological Marketplace." *American Journal of Political Science* 58(2): 367–386.

Bonica, Adam. 2018. "Inferring Roll-Call Scores from Campaign Contributions Using Supervised Machine Learning." *American Journal of Political Science* 62(4): 830–848.

Bonica, Adam. 2023. "Database on Ideology, Money in Politics, and Elections.".

Bonica, Adam, and Zhao Li. 2021. "Inferring Candidates' Issue-Specific Positions from Itemized Campaign Contributions Using Supervised Machine Learning." `https://www.dropbox.com/s/6uie90whk4n7xdl/Bonica%20and%20Li%20%282021%29.pdf?dl=0`.

Brady, David W., Hahrie Han, and Jeremy C. Pope. 2007. "Primary Elections and Candidate Ideology: Out of Step with the Primary Electorate?" *Legislative Studies Quarterly* 32(1): 79–105.

Breiman, Leo. 2001. "Random forests." *Machine Learning* 45: 5–32.

Calonico, Sebastian, Mattias D. Cattaneo, and Rocío Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82(6): 2295–2326.

Canes-Wrone, Brandice, and Michael R. Kistner. 2022. "Out of Step and Still in Congress? Electoral Consequences of Incumbent and Challenger Positioning across Time." *Quarterly Journal of Political Science* 17(3): 389–420.

Canes-Wrone, Brandice, David W. Brady, and John F. Cogan. 2002. "Out of Step, Out of Office: Electoral Accountability and House Members' Voting." *American Political Science Review* 96(1): 127–140.

Caughey, Devin, and Christopher Warshaw. 2022. *Dynamic Democracy: Public Opinion, Elections, and Policymaking in the American States.* University of Chicago Press.

Downs, Anthony. 1957. *An Economic Theory of Democracy.* Harper and Row.

Eggers, Andrew, Anthony Fowler, Jens Hainmueller, Andrew B. Hall, and James M. Snyder, Jr. 2015. "On the Validity of the Regression Discontinuity Design for Estimating Electoral Effects: Evidence From Over 40,000 Close Races." *American Journal of Political Science* 59(1): 259–274.

Fouirnaies, Alexander, and Andrew B. Hall. 2014. "The Financial Incumbency Advantage: Causes and Consequences." *Journal of Politics* 76(3): 711–724.

Fouirnaies, Alexander, and Andrew B. Hall. 2020. "How Divisive Primaries Hurt Parties: Evidence from Near-Runoffs in US Legislatures." *The Journal of Politics* 82(1): 43–56.

Fouirnaies, Alexander, and Andrew B. Hall. 2022. "How Do Electoral Incentives Affect Legislator Behavior? Evidence from U.S. State Legislatures." *American Political Science Review* 116(2): 662–676.

Hall, Andrew B. 2015. "What Happens When Extremists Win Primaries?" *American Political Science Review* 109(1): 18–42.

Hall, Andrew B. 2019. *Who Wants to Run? How the Devaluing of Political Office Drives Polarization.* University of Chicago Press.

Hall, Andrew B., and Daniel M. Thompson. 2018. "Who Punishes Extremist Nominees? Candidate Ideology and Turning Out the Base in U.S. Elections." *American Political Science Review* 112(3): 509–524.

Hall, Andrew B., and James M. Snyder, Jr. 2015. "Candidate Ideology and Electoral Success." https://www.andrewbenjaminhall.com/Hall_Snyder_Ideology.pdf.

Hill, Seth J., and Chris Tausanovitch. 2016. "Southern Realignment, Party Sorting, and the Polarization of American Primary Electorates, 1958-2012." Working Paper.

Hogan, Robert E. 2008. "Policy Responsiveness and Incumbent Reelection in State Legislatures." *American Journal of Political Science* 52(4): 858–873.

Hopkins, Daniel J. 2018. *The Increasingly United States: How and Why American Political Behavior Nationalized*. University of Chicago Press.

Klarner, Carl. 2023. *State Legislative Election Returns, 1967-2022*.

Kuriwaki, Shiro. 2023. "Ticket Splitting in a Nationalized Era." https://osf.io/preprints/socarxiv/bvgz3/.

Lee, Frances E. 2009. *Beyond Ideology: Politics, Principles, and Partisanship in the US Senate*. University of Chicago Press.

McCarty, Nolan M., and Keith T. Poole. 1998. "An Empirical Spatial Model of Congressional Campaigns." *Political Analysis* 7: 1–30.

McCarty, Nolan M., Keith T. Poole, and Howard Rosenthal. 2006. *Polarized America: The Dance of Ideology and Unequal Riches*. MIT Press Cambridge, MA.

McKay, Amy. 2008. "A Simple Way of Estimating Interest Group Ideology." *Public Choice* 136(1): 69–86.

McKay, Amy. 2010. "The Effects of Interest Groups' Ideology on Their PAC and Lobbying Expenditures." *Business and Politics* 12(2): 1–21.

Moskowitz, Daniel J. 2021. "Local News, Information, and the Nationalization of US Elections." *American Political Science Review* 115(1): 114–129.

Phillips, Connor Halloran, James M. Snyder, Jr., and Andrew B. Hall. 2024. "Who Runs for Congress? A Study of State Legislators and Congressional Polarization." *Quarterly Journal of Political Science* 19(1): 1–25.

Poole, Keith T., and Thomas Romer. 1985. "Patterns of Political Action Committee Contributions to the 1980 Campaigns for the United States House of Representatives." *Public Choice* 47(1): 63–111.

Poole, Kieth T, and Howard Rosenthal. 2007. *Ideology and Congress*. New Brunswick, NJ: Transaction.

Rogers, Steven. 2016. "National Forces in State Legislative Elections." *The Annals of the American Academy of Political and Social Science* 667(1): 207–225.

Rogers, Steven. 2017. "Electoral Accountability for State Legislative Roll Calls and Ideological Representation." *American Political Science Review* 111(3): 555–571.

Rogers, Steven. 2023. *Accountability in American Legislatures*. University of Chicago Press.

Shor, Boris, and Nolan McCarty. 2011. "The Ideological Mapping of American Legislatures." *American Political Science Review* 105(3): 530–551.

Squire, Peverill. 2017. "A Squire Index Update." *State Politics & Policy Quarterly* 17(4): 361–371.

Stiglitz, Edward H, and Barry R Weingast. 2010. "Agenda Control in Congress: Evidence from Cutpoint Estimates and Ideal Point Uncertainty." *Legislative Studies Quarterly* 35(2): 157–185.

Thomsen, Danielle M. 2014. "Ideological Moderates Won't Run: How Party Fit Matters for Partisan Polarization in Congress." *The Journal of Politics* 76(7): 786–797.

Thomsen, Danielle M. 2017. *Opting Out of Congress: Partisan Polarization and the Decline of Moderate Candidates*. Cambridge University Press.

Online Appendix

Intended for online publication only.

Contents

A.1 Scaling Coverage Diagnostics and Robustness

In this section, we report balance statistics for the candidates and elections included in our regression sample compared to those that are excluded because they could not be assigned an ideology score. As a point of comparison, we also report coverage statistics for DW-DIME (Bonica 2018), even though we do not use it in the paper due to the small number of state legislative candidates with available scores.

Table A.1 – Scaling Coverage Balance Table. Table reports the count (rows 1-2), median count (row 3), and share (rows 4-10) of observations with non-missing scalings broken down by candidate *Attribute*. *Full Dataset* refers to the population values in the complete election returns dataset.

		Scaling					
	Attribute	Full Dataset	ML Scaling	HS Score	Static CFscore	Dynamic CFscore	DW-DIME
1	Total Candidate-Years	129,058	62,768	63,092	120,494	119,881	3,164
2	Total Distinct Candidates	67,965	26,506	26,546	63,268	63,171	1,187
4	Incumbent	0.373	0.460	0.461	0.414	0.415	0.621
5	Democrat	0.506	0.491	0.490	0.499	0.499	0.495
6	Lower Chamber	0.770	0.791	0.790	0.759	0.759	0.558
7	Vote Share General	0.622	0.671	0.671	0.643	0.643	0.732
8	Win General	0.484	0.661	0.660	0.538	0.540	0.849
9	Vote Share Primary	0.416	0.529	0.528	0.452	0.454	0.561
10	Win Primary	0.785	0.907	0.907	0.821	0.822	0.944

Table A.2 – Midpoint Coverage Balance Table. This table reports the number of general election *races* stratified by various race attributes and data restrictions.

		Data Restriction						
	Attribute	All Races	Contested Races	Competitive Races	Races with HMH Midpoint	Races with HS Midpoint	Races with StaticCF Midpoint	Races with DW-DIME Midpoint
1	N Races	63,109	37,335	16,242	10,202	10,287	28,721	14
2	Average Win Margin	0.56	0.26	0.10	0.16	0.16	0.23	0.15
3	Share Incumbents	0.83	0.72	0.68	0.60	0.60	0.70	0.57
4	Average General Elec. Contribs. (1000s)	169	169	255	304	303	178	498
5	Average Dem. Pres. Vote Share	0.51	0.50	0.48	0.46	0.46	0.48	0.51
6	Average Year	2011	2011	2011	2011	2011	2011	2005
7	Share Western States	0.17	0.20	0.19	0.24	0.24	0.21	0.57
8	Share Midwestern States	0.25	0.30	0.30	0.37	0.37	0.28	0.29
9	Share Southern States	0.34	0.26	0.25	0.28	0.28	0.28	0.07
10	Share Eastern States	0.24	0.25	0.26	0.10	0.10	0.22	0.07

Shares for state geography may not sum to one due to rounding. Races with Midpoint must feature competition between one scalable candidate for each party.

Table A.3 – Primary Extremism Coverage Balance Table. This table reports the number of primary election *races* stratified by various race attributes and data restrictions.

	Attribute	Data Restriction						
		All Races	Contested Races	Competitive Races	Races with HMH Relative Centrism	Races with HS Relative Centrism	Races with StaticCF Relelative Centrism	Races with DW-DIME Relelative Centrism
1	N Races	79,888	18,362	3,976	4,062	4,113	13,781	7
2	Average Win Margin	0.28	0.28	0.05	0.16	0.16	0.24	0.13
3	Share Incumbent	0.61	0.37	0.18	0.18	0.18	0.33	0.00
4	Average Primary Elec. Contribs. (1000s)	72	120	141	230	229	137	1,356
5	Average Dem. Pres. Vote Share	0.50	0.50	0.49	0.50	0.50	0.50	0.59
6	Average Year	2011	2012	2011	2011	2011	2011	2005
7	Share Western States	0.17	0.20	0.22	0.24	0.24	0.20	0.43
8	Share Midwestern States	0.26	0.27	0.27	0.29	0.29	0.27	0.29
9	Share Southern States	0.32	0.39	0.39	0.41	0.41	0.42	0.29
10	Share Eastern States	0.24	0.14	0.13	0.06	0.07	0.11	0.00

Note: Shares for state geography may not sum to one due to rounding. Races with Rel. Centrism must feature at least two scalable-candidates. Following Table 5, the sample is restricted to contested primary elections and excludes races in districts where the opposing party received greater than 70% of the two-party presidential vote share

Figure A.1 – ML Score Sample for Midpoint Analysis. Using our ML score, this figure plots the total number of general elections, contested elections, elections with a margin less than 20%, and the number of observations in our analysis sample for every even-numbered year.

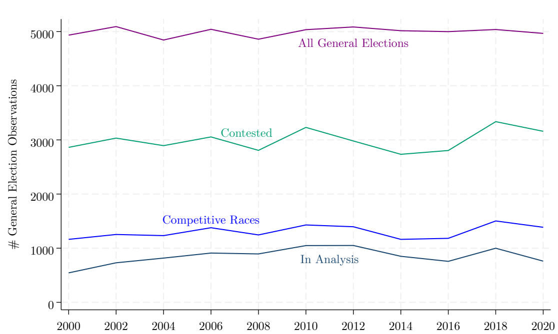
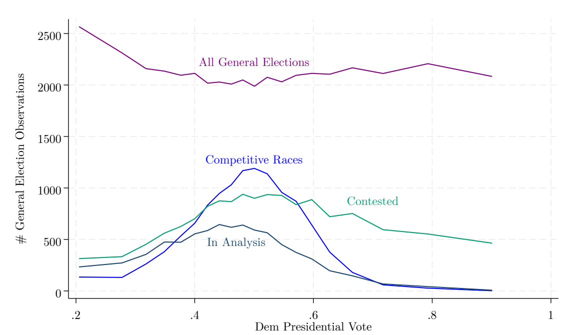


Figure A.2 – ML Score Sample for Midpoint Analysis. Using our ML score, for 20 equal-sample-sized bins of Democratic presidential vote share this figure plots the total number of general elections, contested elections, elections with a margin less than 20%, and the number of observations in our analysis sample.



A.2 Feature Engineering and Cross-Validation for ML Scores

In this section, we provide details on the construction of the feature set for the random forest, the design and results of the cross validation procedure to choose the optimal number of predictors considered at each split in the trees, and the most predictive features from the final model. We also report statistics on multistate donors and their role in enhancing the models' predictive accuracy.

To construct the feature set, we start by summing the total contribution amounts from each donor to each candidate in each election cycle. When candidates run in multiple states or multiple parties across different election cycles, we treat them as separate candidates. We reduce these contributions down to a contribution matrix $\boldsymbol{X}$ where $\boldsymbol{X}_{ij}$ represents the *average* amount that donor j gave to candidate i over all available election cycles. We use averages to reduce the scale differences between candidates that run in different numbers of election cycles.

Using $\boldsymbol{X}$, we create two types of donation summary features. The summary features were calculated for candidates in the training set in accordance with the ten-fold cross-validation scheme as follows. Let $\mathcal{F}$ be the set of indices for candidates in the holdout fold at any step of the cross-validation procedure. For each donor, we calculate the dollar-weighted average scaling for each donor j to candidate i as:

$$z_j^{(i)} = \frac{\sum_{w \neq i, w \notin \mathcal{F}} y_w \; \boldsymbol{X}_{wj}}{\sum_{w \neq i, w \notin \mathcal{F}} \boldsymbol{X}_{wj}},$$

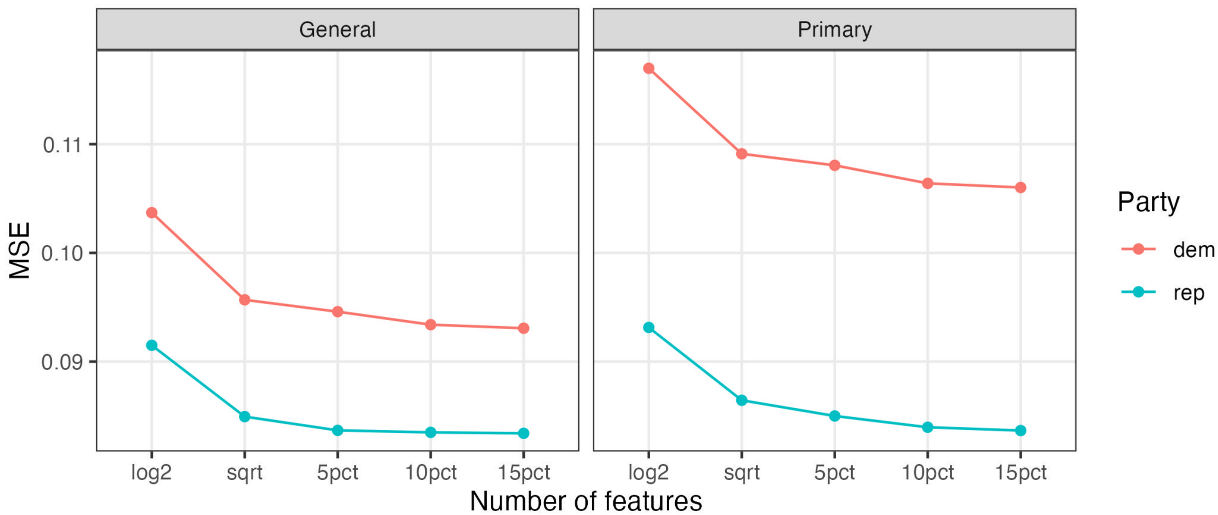
where y_w is the static scaling for candidate w after they take office. With these donor weighted averages, we calculate two types of summary features for candidate i that include no forbidden information from the candidate itself or candidates in the holdout fold. First, we calculate the dollar-weighted average scaling for candidate i using the donor scalings $z_j^{(i)}$ as in Equation 2, where the weights are the proportion of donations candidate i received from donor j. Second, we bin the $z_j^{(i)}$'s into bins between -4 and 4 of width 0.2, and calculate the proportion of donations to candidate i that fall into each bin. Legislators in the training set receive the score from the cross-validation step where they were in the holdout fold.

We also include dummy variables for state, and dummy variables for larger individual donors. To improve coverage within states while reducing the computational complexity of the model, we include individual donors as dummy variables if they gave to at least

25% of the candidates within at least one state for the model that includes general election donations, and 15% of the candidates for the model that only includes primary donations.

Figure A.3 reports the results of the cross-validation for both the primary and general election models. We experimented with choosing $log2(n), sqrt(n), 0.05 \cdot n, 0.10 \cdot n, 0.15 \cdot n$ predictors at each split, where n is the total number of features. Figure A.4 shows that the most predictive features were by far the donation summary features and the state dummy variables.

Figure A.3 – Cross-validation results for choosing number of predictors



As we note in the methods section, the state contribution data is more sparse than the federal contribution data, so borrowing information from donors across states is an important way that our model is able to make more accurate predictions for states with less available data. We directly assessed this possibility by experimenting with training separate models by state, and found that the overall mean squared prediction error decreases by 38% when we allow the ML model to pool information across states.

The reason for this is that out-of-state giving is a common enough phenomenon among larger state donors to help the model make better predictions by pooling information across states. Out of the donors and candidates that meet our modeling data restrictions (i.e., donors who gave to at least 5 candidates with an NP score and candidates that received donations from at least 5 of these donors), 22% of donors gave to campaigns in at least two different states. These "multistate" donors contributed to campaigns in 5 different states on average, and represent 15% of the contributions in the modeling data. Figure A.5 plots the average proportion of multistate donors and contributions per candidate by state in our modeling data.

Figure A.4 – Feature importance for general election model

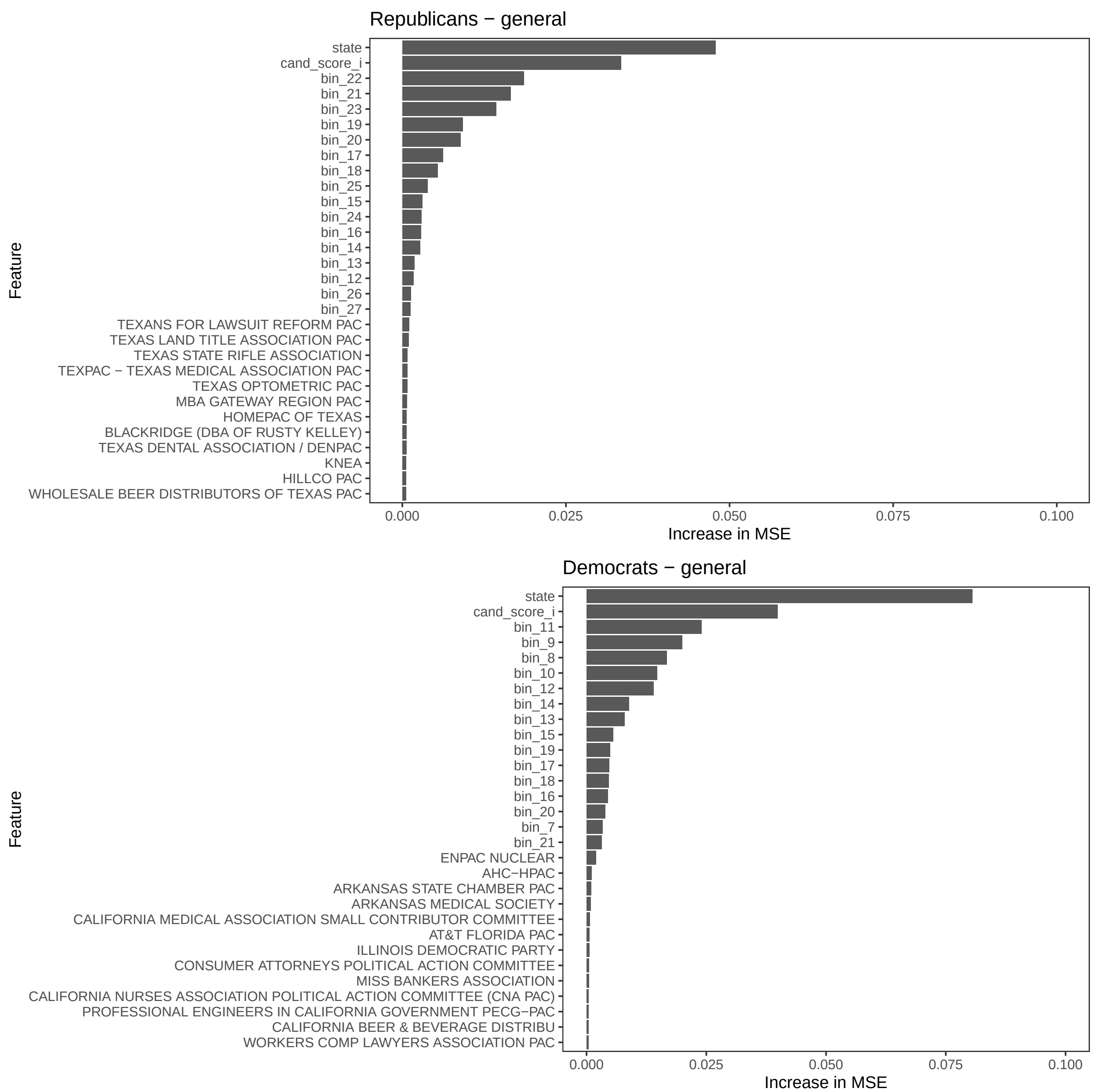
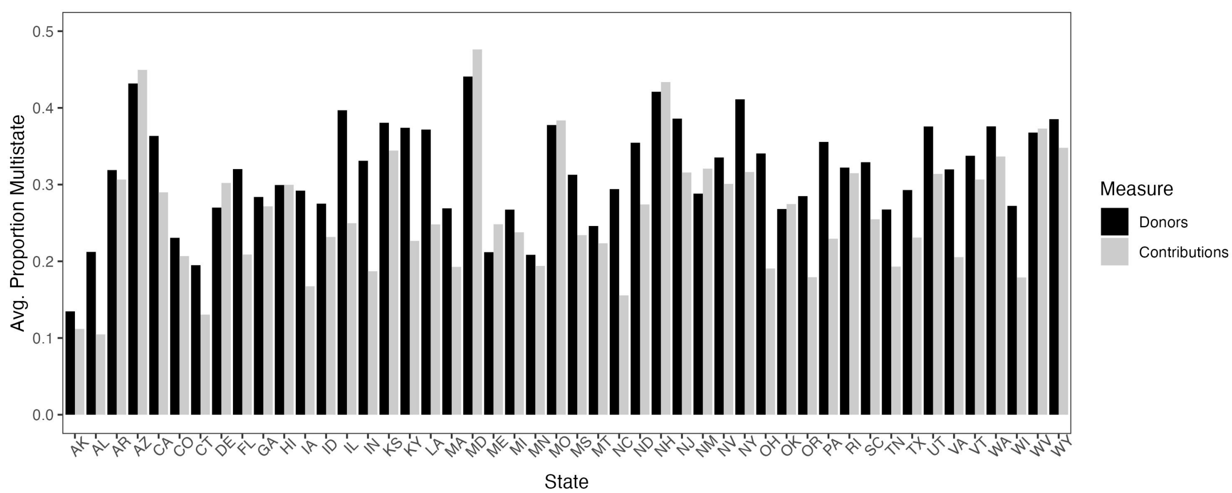


Figure A.5 – Average Proportion of Multistate Donors and Donations per Candidate, by State.



A.3 Contribution Data Validation

To evaluate the quality of the National Institute on Money in Politics's (NIMSP) donor identity resolution software, we benchmark the set of NIMSP donor IDs against the donor IDs reported in the Database of Ideology, Money in Politics, and Elections (DIME) (Bonica, 2023). DIME is widely considered to employ literature-standard entity resolution processes. Conveniently, the coverage of state legislative campaign finance data in DIME is nearly identical to that of the NIMSP data, but donor IDs in DIME are constructed independently of the IDs reported in NIMSP.

To compare the donor IDs produced by NIMSP and DIME, we identify a set of donors that provide the most information in our scaling process: donors that contribute to at least 10 distinct candidates. (We have confirmed that the following results are very similar for cutoffs between 5 and 50). Then for every donor in each data source, we calculate the number of different election cycles in which that donor ID is observed making at least one contribution. Finally, we aggregate these results across donors within the same data source. The resulting quantity—the number of election cycles in which the average donor contributes—captures the extent to which the NIMSP and DIME identity resolution softwares match individual donors across time. This is important because our scalings rely heavily on donors that "bridge" candidates across election cycles and jurisdictions.

We conduct this exercise separately for non-individual and individual donors and all donors. The results of this exercise are reported in the Figure A.6. The horizontal axis reports the average number of election cycles in which a donor contributes to at least one candidate and the vertical axis reports the share of donors. Results are plotted separately for NIMSP (blue) and DIME (red) and the vertical lines and black numbers report averages within data sources. Overall, we find that the distribution of donor persistence across time is highly similar between NIMSP and DIME. In aggregate, donors in NIMSP contribute in 5.9 election cycles while the same value in DIME is 5.7. Individual donors contribute in 5.5 election cycles in NIMSP and 5.9 in DIME, on average. And non-individual donors contribute in 7.5 election cycles in DIME and 5.7 in NIMSP, on average. We conclude that the donor IDs in NIMSP and DIME are highly stable.

As a further robustness check, we reconstruct our baseline Hall-Snyder scores using the DIME data rather than NIMSP data. Figure A.7 plots Hall-Snyder scores calculated using the NIMSP data (i.e., scalings reported in the main paper) on the vertical axis and Hall-Snyder scores calculated using DIME data on the horizontal axis. As is apparent, the two scalings are highly correlated (r=.97 overall, .92 for Democrats, .88 for Republicans).

Figure A.6 – Distribution of Donor Persistence in NIMSP and DIME Data.

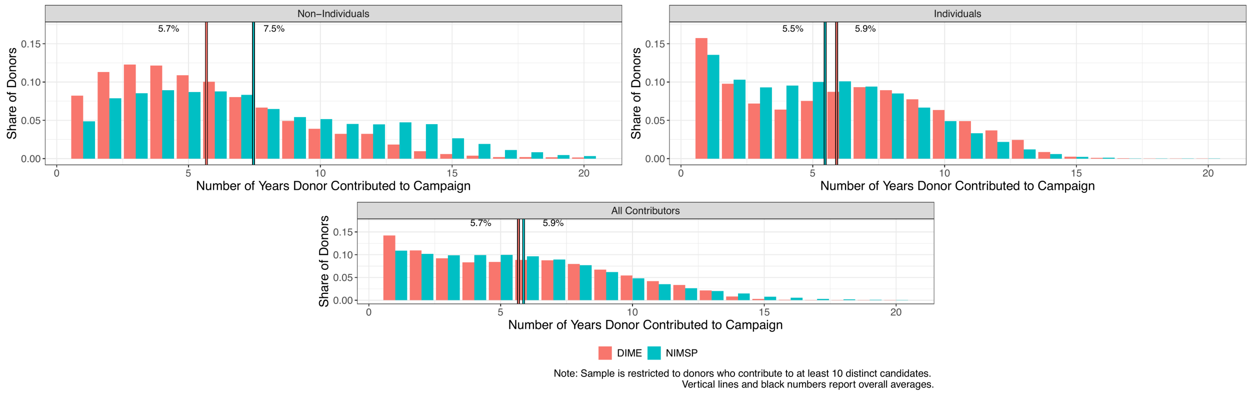
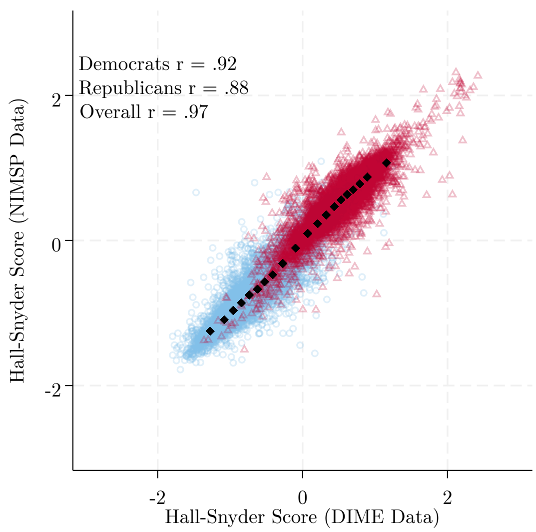


Figure A.7 – Hall-Snyder Scores Generated NIMSP and DIME Data Correlate Highly. This figure shows the correlation between Hall-Snyder scores generated using NIMSP data (vertical axis) and DIME data (horizontal axis).



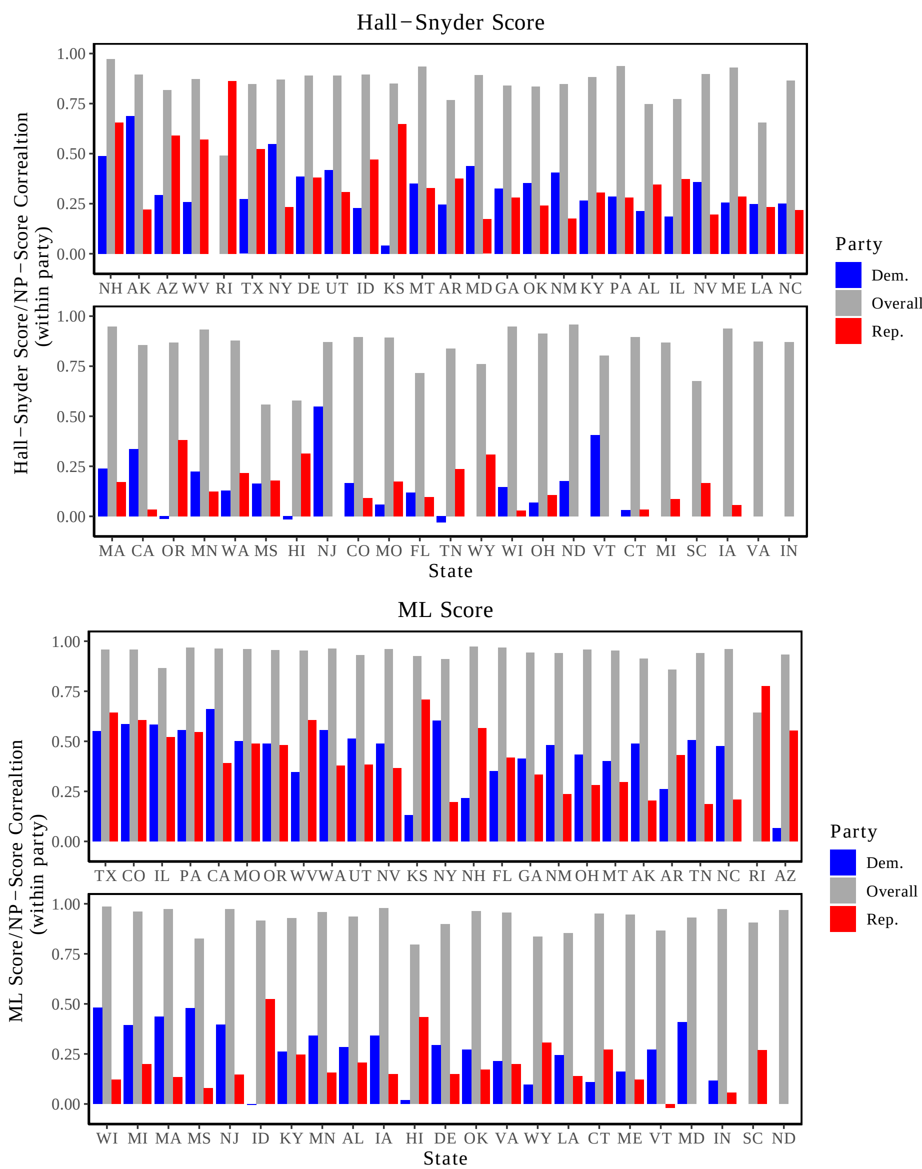
Finally, we re-run our main analyses using the DIME-based Hall-Snyder scores. Overall, the results are highly similar between NIMSP- and DIME-based Hall-Snyder scores. We conclude that our results are both replicable and robust to alternative identity resolution softwares

A.4 Within-State Scaling Correlations

In this section, we examine the within-party correlations of both the Hall-Snyder and ML scores with NP-Scores across states.

Figure A.8 plots within-state correlations between Hall-Snyder and ML scores with NP-Scores. Correlations are high in many states, but there are some states and parties where the correlations are quite low, which is to be expected given the large number of different states and contexts in the data.

Figure A.8 – State-Level Within Party Correlations Between Scalings and NP-Scores. Hall-Snyder Scores correlate highly with NP-Scores within party and state.



A.5 Scaling Robustness to Changes in Campaign Finance Regulations

In this section we explore the sensitivity of the NP-score predictions to changes in contribution limits and other campaign finance regulations during the period we study. Because our specifications always include controls for between-state and between-year differences in prediction error, the most important source of confounding to investigate is within-state prediction error trends related to these changes net of the global time trend in prediction error. To probe the possibility that the prediction error may be sensitive to these within-state regulatory changes, we have included an analysis of whether the prediction error in all the scalings we employ is related to changes in campaign finance regulations, including contribution limits, disclosure rules, and public financing (obtained from the Campaign Finance Institute's state law database, `http://www.cfinst.org/law/stateLinks.aspx`). To address different concerns about bias, we use two definitions of prediction error:

$$\text{Squared Prediction Error} = (\hat{y} - y)^2$$

$$\text{Signed Prediction Error} = \begin{cases} \text{sign}(y) \cdot (\hat{y} - y) & \text{if sign}(y) = \text{sign}(\hat{y}) \\ |\hat{y}| - |y| & \text{if sign}(y) \neq \text{sign}(\hat{y}) \end{cases} \quad \text{where sign}(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0. \end{cases}$$

The squared prediction error metric simply captures the magnitude of the errors in any direction, while the signed prediction error captures whether the model predicts the candidate as too "moderate" (i.e., model errs in the direction of too much shrinkage towards zero, always represented as negative errors) or too "extreme" (i.e., model errs in the direction of too much inflation away from zero, always represented as positive errors). There are two cases in the signed error function to properly sign errors for the edge cases where the predicted and actual NP Score do not share the same sign ($< 3\%$ of cases for the ML models).

The squared prediction error results presented in Table A.4 suggest that most regulatory changes are not statistically significantly related to changes in prediction error across the 5 scalings we employ, and in the cases where they are statistically significant, the coefficients are small relative to the overall mean squared prediction error. Similarly, the signed prediction error coefficients in Table A.5 are often not statistically significant and are small relative to the overall root mean squared prediction error (negative coefficients indicate more shrinkage errors made after the policy change, positive coefficients indicate more inflation errors). The results also underscore the advantage of using 5 different predictions in our analyses, since no regulatory change is associated with the prediction error in the same way across the 5

Table A.4 – Squared Prediction Error.

	Squared Prediction Error				
	HS	ML (All)	ML (Primary)	Static CF	Dynamic CF
Contribution Limits (1000s)					
Individual	-0.002	-0.007	-0.015	0.000	0.003
	(0.005)	(0.003)	(0.006)	(0.007)	(0.008)
PAC	-0.002	-0.001	0.003	-0.008	-0.008
	(0.008)	(0.004)	(0.007)	(0.005)	(0.004)
Corp	0.003	0.000	-0.002	0.005	0.006
	(0.004)	(0.001)	(0.002)	(0.003)	(0.003)
Labor	0.003	0.005	0.017	-0.002	-0.001
	(0.004)	(0.003)	(0.005)	(0.005)	(0.004)
Other Candidate	0.005	0.003	-0.004	0.000	-0.006
	(0.011)	(0.003)	(0.006)	(0.005)	(0.006)
Public Funding Provided	-0.013	0.004	0.007	0.000	0.038
	(0.019)	(0.008)	(0.009)	(0.014)	(0.012)
Donor Disclosure Minimum Amt. (10s)	-0.002	-0.002	-0.004	0.001	0.003
	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)
Electronic Disclosure Mandatory	-0.013	-0.001	-0.004	-0.024	-0.020
	(0.019)	(0.008)	(0.009)	(0.016)	(0.022)
N	17505	17423	9676	28313	28188
Mean Squared Prediction Error	0.264	0.081	0.087	0.251	0.284
State FEs	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y

Note: Std. errors clustered by state.

predictions. For example, the HS, ML Primary, and static CF scores make more inflation errors after states switch to allowing public funding, but the main ML and dynamic CF scores are insensitive to this policy change. This gives us confidence that our results across the 5 predictions are unlikely to suffer from the same source of bias related to regulatory changes.

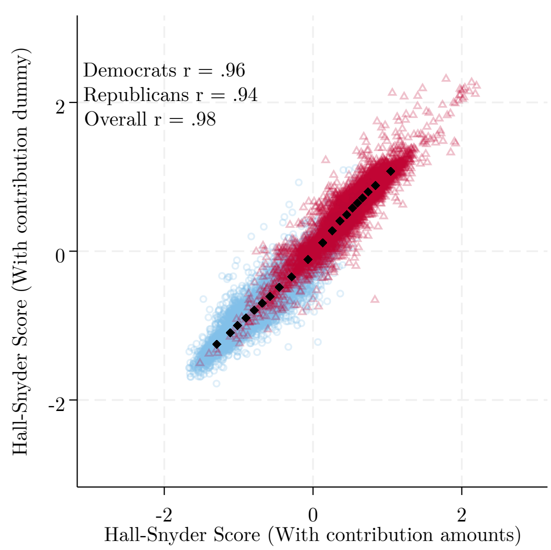
As a further check on sensitivity to contribution limits, we also show that the dollar amount of contributions does not appear to be driving our predictive performance. Figure A.9 shows the correlation between Hall-Snyder scores computed using contribution amounts (horizontal axis) and an alternative version computed using an indicator for contributions (vertical axis). The former scaling is the same Hall-Snyder scaling employed in the main paper, while the latter scaling leverages only the decision to donate and not the actual contribution amount. As is apparent, the within-party and overall correlations between these scalings are quite high (r=.96 for Democrats, r=.94 for Republicans, r=.98 overall). The results suggest that it is the decision to donate, rather than the donation amount, that primarily drives our ideological scaling, matching the conclusions of Bonica (2014, 2018).

Table A.5 – Signed Prediction Error.

	Signed Prediction Error				
	HS	ML (All)	ML (Primary)	Static CF	Dynamic CF
Contribution Limits (1000s)					
Individual	-0.017	-0.003	-0.010	-0.001	0.003
	(0.006)	(0.005)	(0.012)	(0.009)	(0.011)
PAC	-0.007	-0.001	0.003	-0.001	-0.006
	(0.010)	(0.004)	(0.014)	(0.006)	(0.007)
Corp	0.000	-0.001	-0.004	0.000	0.000
	(0.004)	(0.001)	(0.003)	(0.004)	(0.004)
Labor	0.013	0.005	0.014	-0.004	-0.005
	(0.006)	(0.004)	(0.012)	(0.008)	(0.009)
Other Candidate	-0.003	-0.007	-0.014	0.013	0.014
	(0.013)	(0.005)	(0.013)	(0.010)	(0.012)
Public Funding Provided	0.032	0.003	0.034	0.048	-0.011
	(0.022)	(0.008)	(0.014)	(0.028)	(0.035)
Donor Disclosure Minimum Amt. (10s)	-0.001	-0.001	-0.002	-0.003	0.000
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
Electronic Disclosure Mandatory	0.001	-0.001	-0.028	-0.010	-0.007
	(0.019)	(0.009)	(0.016)	(0.018)	(0.021)
N	17505	17423	9676	28313	28188
Root Mean Squared Prediction Error	0.514	0.284	0.296	0.501	0.533
State FEs	Y	Y	Y	Y	Y
Year FEs	Y	Y	Y	Y	Y

Note: Std. errors clustered by state.

Figure A.9 – Hall-Snyder Scores Generated Contribution Amounts and Contribution Indicators Correlate Highly. This figure shows the correlation between Hall-Snyder scores generated using contribution amounts (vertical axis) and contribution indicators (horizontal axis).



A.6 Roll Call Classification Exercise

Another way to validate the new scalings is to use them to predict the outcome of specific roll-call votes. To do so, we follow Bonica (2014, 2018) and calculate the percentage of state legislative roll-call votes that can be correctly classified using an optimal cutting-point procedure described in Poole and Rosenthal (2007).[23] For this exercise, we construct a panel containing the near-universe of roll-call votes cast in all 99 state legislative chambers for the years 2010-2022, and a subset of states for the years 2000-2009. Overall, this panel includes 72 million roll-call votes.

Table A.6 – Number of State Legislative Roll Call Votes, 2000-2022.

Year	Overall	House	Senate	Year	Overall	House	Senate
2000	525,030	502,200	22,830	2012	3,901,469	2,860,733	1,040,736
2001	1,335,741	1,313,014	22,727	2013	4,901,037	3,647,518	1,253,519
2002	647,393	628,493	18,900	2014	3,726,559	2,726,239	1,000,320
2003	1,469,279	1,448,997	20,282	2015	5,448,711	4,052,937	1,395,774
2004	905,406	880,660	24,746	2016	4,058,217	2,962,530	1,095,687
2005	1,423,359	1,396,849	26,510	2017	5,914,265	4,297,685	1,616,580
2006	893,547	867,604	25,943	2018	4,622,352	3,315,950	1,306,402
2007	1,296,335	1,275,055	21,280	2019	6,164,053	4,456,106	1,707,947
2008	908,425	884,248	24,177	2020	3,619,255	2,527,984	1,091,271
2009	1,834,702	1,534,968	299,734	2021	6,224,710	4,552,591	1,672,119
2010	2,212,753	1,570,450	642,303	2022	4,748,004	3,444,087	1,303,917
2011	4,710,315	3,489,983	1,220,332				

This state legislative roll call data was assembled from two sources. First, data for the near-universe of roll call votes cast in all 99 state legislative chambers for the years 2010-2022 was collected by the authors from www.Legiscan.com. This data consists of 60.8 million individual votes. We supplement this data with 11.2 million roll call votes for the years 2000-2009 from Fouirnaies and Hall (2022) for a varying panel of 21 states.[24] Combined, our roll call dataset encompasses 72 million distinct votes. Following Bonica (2014, 2018) and Poole and Rosenthal (2007), we remove lopsided roll calls with margins greater than 97.5% and omit abstentions and missed votes. Table A.6 reports the total number roll-call votes in our dataset by chamber and year.

[23] Specifically, for every roll-call in our dataset, we find the maximally-classifying point in one-dimensional space that predicts "Yea" votes on one side and "Nay" votes on the other. We then report the percentage of all votes cast that are correctly predicted.

[24] We include the unbalanced panel of states from 2000-2009 in our main analyses to evaluate the predictive capacity of our Hall-Snyder scores over an extended time frame. Our results in Table A.7 are very similar if we instead focus on the years 2010-2022 for which we have a balanced panel.

Table A.7 – Percent of State Legislative Roll Call Votes Classified Correctly, 2000-2022.

Scaling	Overall	House	Senate
NP-Score	0.914	0.913	0.920
	(0.755)	(0.752)	(0.767)
ML Scaling	0.900	0.899	0.906
	(0.706)	(0.705)	(0.712)
Hall-Snyder Score	0.890	0.889	0.899
	(0.678)	(0.676)	(0.690)
Static CFscore	0.883	0.881	0.888
	(0.663)	(0.661)	(0.676)
Party	0.856	0.858	0.847
	(0.586)	(0.593)	(0.549)

Note: Aggregate Proportional Reduction in Error reported in parantheses.[25] Table is ordered by overall classification rate.

Using this data, for each roll call and scaling, we calculate the optimal cutting point between "yea" and "nay" votes (Poole and Rosenthal 2007). Leveraging these cutpoints, we impute predicted roll call votes and compare the result to the true votes cast.

Table A.7 reports the classification rates and aggregate proportional reduction in error (APRE) for our new scores and, for comparison, NP-Scores, dynamic and static CFscores, and the naive indicator for party.[26] The table orders the scalings by overall prediction rate. As can be seen, the order is as expected: the ML scores do the best job of replicating the classification success of the NP-Scores themselves, the Hall-Snyder scores do almost as good a job, the CFscores do slightly worse, and all four outperform the naive Party model.

[25] $\text{APRE}_i = \frac{\sum_{j=1}^{J}\{\text{minority vote}_j - \text{classification errors}_{ij}\}}{\sum_{j=1}^{J}\text{minority votes}_j}$ for scaling i and roll call j.

[26] We exclude DW-DIME Scores from this analysis because their coverage is insufficient to accurately calculate representative cutting-points.

A.7 Midpoint Estimate Robustness Checks

Estimates with Presidential Vote Share

In this table, we re-estimate the midpoint regressions using presidential vote share to control for district preferences instead of district fixed effects. As the table shows, we find generally larger estimates of the advantage in this specification, but with significantly less data.

Table A.8 – Advantage of More-Moderate Candidates in Contested General Elections, 2000-2022.

	Dem Vote Share			
	(1)	(2)	(3)	(4)
ML Scaling	0.17	0.12	0.14	0.15
	(0.01)	(0.01)	(0.01)	(0.01)
ML Scaling (Primary donations Only)	0.09	0.05	0.06	0.09
	(0.01)	(0.01)	(0.02)	(0.01)
Hall-Snyder Score	0.24	0.15	0.18	0.20
	(0.01)	(0.01)	(0.01)	(0.01)
Static CFscore	0.60	0.30	0.53	0.59
	(0.02)	(0.02)	(0.03)	(0.02)
Dynamic CFscore	0.38	0.25	0.40	0.38
	(0.02)	(0.02)	(0.04)	(0.02)
District-by-Regime FE	N	N	N	N
Year FE	Y	Y	Y	Y
Controls for Primary Contributions	N	Y	N	N
Controls for Dem. Pres. Vote Share	Y	Y	Y	Y
Only races with below-median contribution gap	N	N	Y	N
Only races with ≥ 10 primary donors per candidate	N	N	N	Y

Each cell in this table reports the coefficient on *Midpoint* from Equation 4 which is scaled to run from 0 (most liberal) to 1 (most conservative) for each scaling. Robust standard errors are clustered by district-regime in parentheses.

Estimates without Low-Correlation and High-Error States

In this table, we re-estimate the midpoint regressions after excluding states with below-median average within-party correlations between ML scores and NP-Scores (first row) and above-median average NP-Score prediction error as reported in Shor and McCarty (2011) (second row). The estimates reported in this table are substantively identical to those estimated using the full sample in Table 1.

Table A.9 – Advantage of More-Moderate Candidates in Contested General Elections, 2000-2022.

	Dem Vote Share			
	(1)	(2)	(3)	(4)
ML Scaling (Excludes low-correlation states)	0.15	0.11	0.14	0.09
	(0.03)	(0.03)	(0.04)	(0.03)
ML Scaling (Excludes high-error states)	0.23	0.16	0.14	0.16
	(0.04)	(0.04)	(0.05)	(0.06)
District-by-Regime FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Controls for Primary Contributions	N	Y	N	N
Only races with below-median contribution gap	N	N	Y	N
Only races with ≥ 10 primary donors per candidate	N	N	N	Y

Each cell in this table reports the coefficient on *Midpoint* from Equation 4 which is scaled to run from 0 (most liberal) to 1 (most conservative) for each scaling. Robust standard errors are clustered by district-regime in parentheses. Sample restrictions are reported in parentheses in the first column.

A.8 Primary Extremism Estimate Robustness Checks

Primary Extremism Estimates After Restricting to Races with Above-Median Contribution Amounts

In this table, we re-estimate the primary extremism regressions after restricting the sample to races with above-median contribution amounts. Our conclusions remain unchanged.

Table A.10 – Advantage of More-Extreme Candidates in Contested Primary Elections, 2000-2022.

	Primary Vote Share			
	(1)	(2)	(3)	(4)
ML Scaling	-0.20	-0.20	-0.08	-0.08
	(0.03)	(0.02)	(0.03)	(0.02)
ML Scaling (Primary donations only)	-0.18	-0.18	-0.11	-0.10
	(0.04)	(0.03)	(0.04)	(0.03)
Hall-Snyder Score	-0.15	-0.16	-0.11	-0.10
	(0.03)	(0.02)	(0.04)	(0.03)
Static CFscore	0.31	0.31	0.30	0.33
	(0.06)	(0.05)	(0.05)	(0.05)
Dynamic CFscore	0.29	0.29	0.27	0.31
	(0.06)	(0.05)	(0.05)	(0.05)
District-by-Party FE	Y	N	Y	N
Party-by-Year FE	Y	N	Y	N
Number of Candidates FE	Y	N	Y	N
Race FE	N	Y	N	Y
Controls for Primary Contributions	N	N	N	N
Only races with below-median contribution gap	Y	Y	N	N
Only races with ≥ 20 donors per candidate	N	N	Y	Y

Each cell in this table reports the coefficient on *Relative Centrism* from Equation 5 which is scaled to run from 0 (most extreme) to 1 (most moderate) for each scaling. The sample is restricted to contested primary elections and excludes races in districts where the opposing party received greater than 70% of the two-party presidential vote share. Robust standard errors in parentheses.

Estimates without Low-Correlation and High-Error States

In this table, we re-estimate the primary extremism regressions after excluding states with below-median average within-party correlations between ML scores and NP-Scores (first row) and above-median average NP-Score prediction error as reported in Shor and McCarty (2011) (second row). The estimates reported in this table for our preferred specification (column 2) are very similar to those estimated using the full sample in Table 4.

Table A.11 – Advantage of More-Extreme Candidates in Contested Primary Elections, 2000-2022.

	Primary Vote Share							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ML Scaling (Excludes low-correlation states)	-0.23	-0.23	-0.24	-0.23	-0.25	-0.25	-0.11	-0.11
	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.03)
ML Scaling (Excludes high-error states)	-0.14	-0.15	-0.13	-0.14	-0.14	-0.14	-0.04	-0.04
	(0.04)	(0.04)	(0.04)	(0.03)	(0.07)	(0.05)	(0.07)	(0.05)
District-by-Party FE	Y	N	Y	N	Y	N	Y	N
Party-by-Year FE	Y	N	Y	N	Y	N	Y	N
Number of Candidates FE	Y	N	Y	N	Y	N	Y	N
Race FE	N	Y	N	Y	N	Y	N	Y
Controls for Primary Contributions	N	N	Y	Y	N	N	N	N
Only races with below-median contribution gap	N	N	N	N	Y	Y	N	N
Only races with ≥ 20 donors per candidate	N	N	N	N	N	N	Y	Y

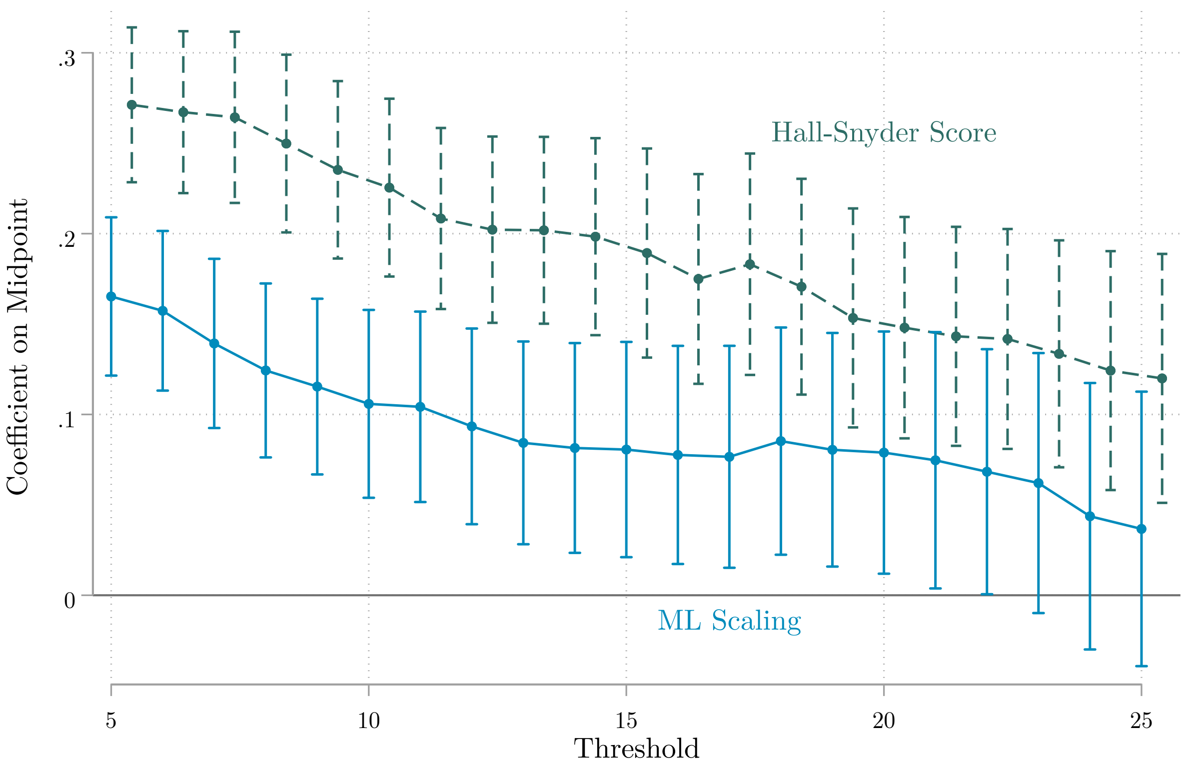
Each cell in this table reports the coefficient on *Relative Centrism* from Equation 5 which is scaled to run from 0 (most extreme) to 1 (most moderate) for each scaling. The sample is restricted to contested primary elections and excludes races in districts where the opposing party received greater than 70% of the two-party presidential vote share. Robust standard errors in parentheses.

A.9 Results Across Scaling Thresholds

In this section, we explore how our main midpoint and primary extremism results vary when we change the threshold required to include a candidate in the regression.

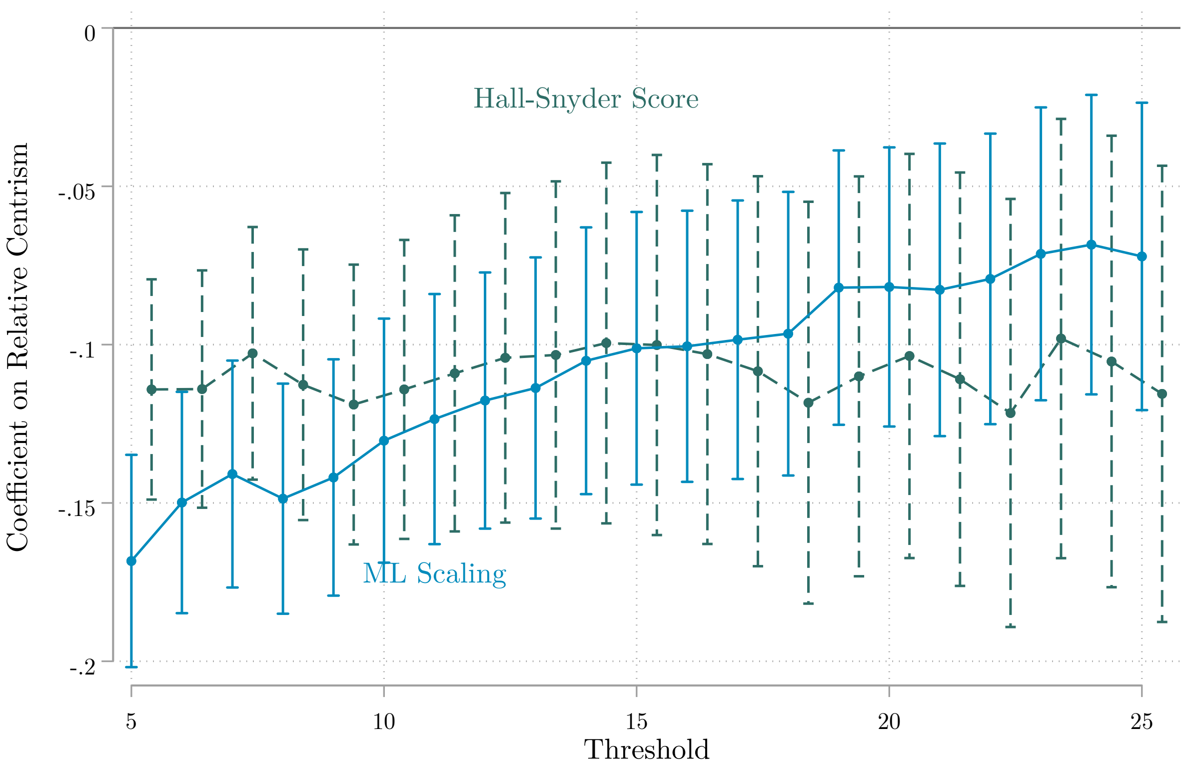
Midpoint Estimates

Figure A.10 – Robustness of General Election Analysis to Scaling Thresholds. This figure reports the coefficient on *Midpoint* across donor thresholds. The donor threshold is the minimum number of unique donors that both candidates in a race must receive a contribution from to be included in the analysis. Vertical bars report 95% confidence intervals.



Primary Extremism Estimates

Figure A.11 – Robustness of Primary Election Analysis to Scaling Thresholds. This figure reports the coefficient on *Relative Centrism* across donor thresholds. The donor threshold is the minimum number of unique donors that all candidates in a primary race must receive a contribution from to be included in the analysis. Vertical lines report 95% confidence intervals from robust standard errors.



A.10 Regression Discontinuity Details and Robustness Checks

In this section, we expand on the RD results presented in the paper. The key assumption for the RD to be a valid estimate is that there is no sorting at the discontinuity: that is, in virtually tied elections, it should not be the case that either the more-moderate or more-extreme candidate systematically end up winning. As discussed and validated in Eggers et al. (2015), this is plausible since it is exceedingly unlikely that primary candidates are able to manipulate the results of these elections. Nevertheless, we can also directly test this assumption—and look for chance imbalances in our sample—by estimating the same RD "effect" where the outcome is the vote share of the nominee's party in the *previous* election cycle. We carry out these tests in Table A.12 and find no evidence for sorting or for an imbalance that would contribute to our negative estimates.

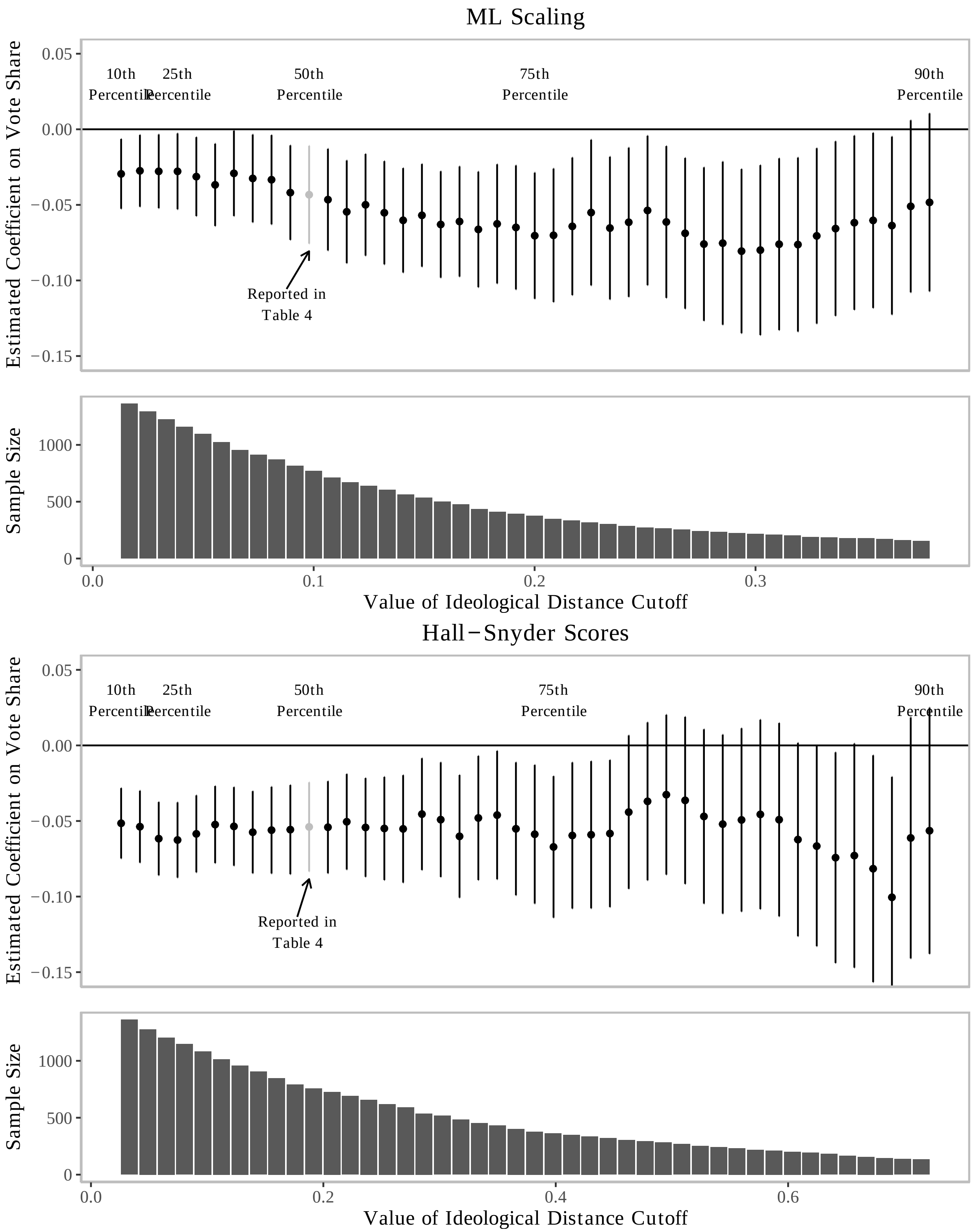
Table A.12 – Effect of Extremist Nominee on Lagged General Election Vote Share, U.S. State Legislatures 2000-2022.

	Party Vote Share			
	(1)	(2)	(3)	(4)
ML Scaling	0.00	0.03	0.00	0.02
	(0.06)	(0.05)	(0.05)	(0.08)
Hall-Snyder Score	0.03	0.02	0.02	0.04
	(0.06)	(0.05)	(0.05)	(0.07)
Polynomial	1	3	5	CCT
Bandwidth	.10	-	-	-

Note: Each cell in this table reports the coefficient on *Extremist Primary Win.* Robust standard errors are reported in parentheses.

Second, in Figure A.12 we also evaluate how the RD estimate changes as we change the cutoff in terms of ideological distance between candidates used to determine which races enter the sample. In each panel, the figure plots the RD estimate across cutoff size, from the 10th to the 90th percentile. At the left of the plot, nearly all cases are being included in the data, including those where the two candidates are quite similar ideologically so that the "treatment" of nominating the more-extreme one is weak. Towards the right of the plot, we are strengthening the treatment by only including cases where the more-extreme candidate is substantially more extreme than the more-moderate candidate. As the figures show, with both measures, we find that the penalty grows as we strengthen the treatment.

Figure A.12 – Effect of Extremist Nominee on General Election Vote Share Across Possible Cutoffs.



A.11 Scaling Error and MSE Correlations

In this section, we document the correlation between measurement error and district competitiveness for our four scaling measures. Because donors are access seeking as well as ideological, candidates in very uncompetitive districts are likely to be scaled as too moderate relative to their true NP-Score. The ML scores do the best at ameliorating this relationship out of the four scores.

Figure A.13 – Dem. Presidential Vote Share (General Election)