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.

		Data Restriction						
	Attribute	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.

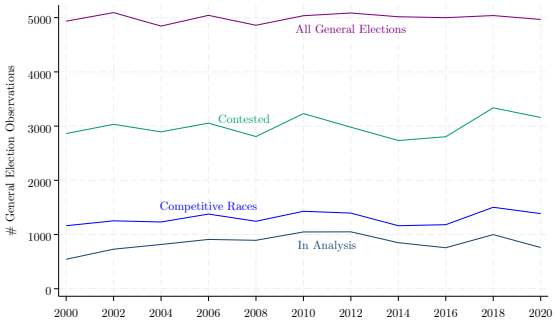
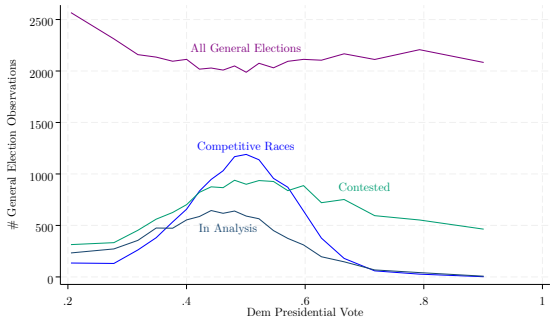


Figure A.2 – ML Score Sample for Midpoint Analysis. Using our ML score, for 20 equal-sample-sized bins of Democratic presidential vote share this figure plots the total number of general elections, contested elections, elections with a margin less than 20%, and the number of observations in our analysis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

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

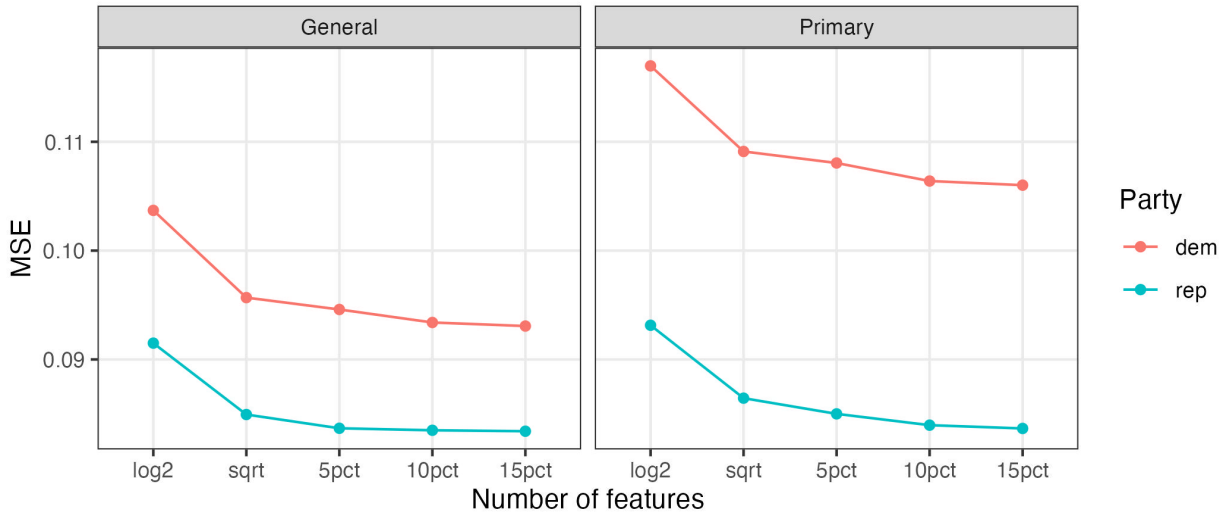
where y_w is the static scaling for candidate w after they take office. With these donor weighted averages, we calculate two types of summary features for candidate i that include no forbidden information from the candidate itself or candidates in the holdout fold. First, we calculate the dollar-weighted 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

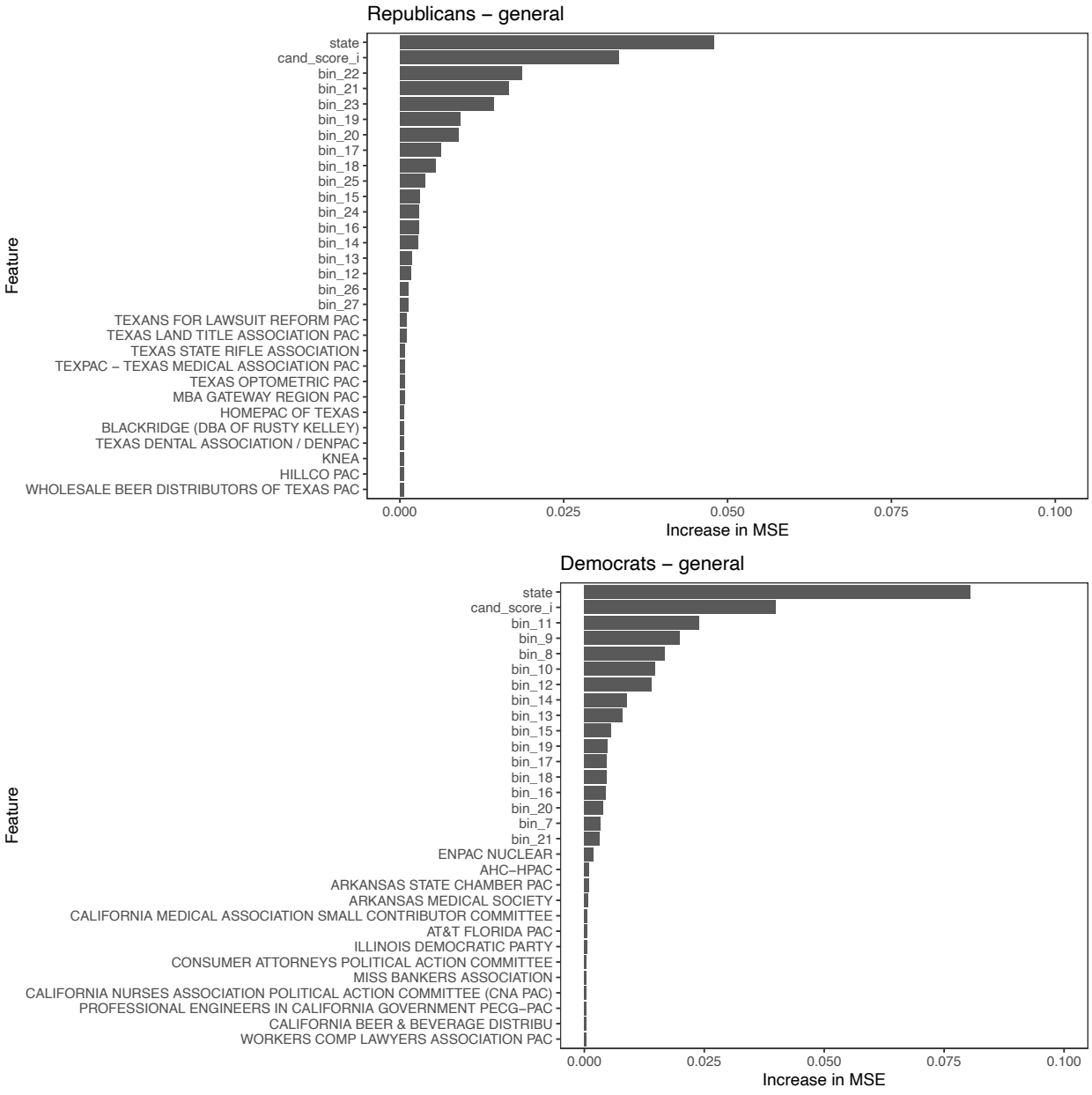
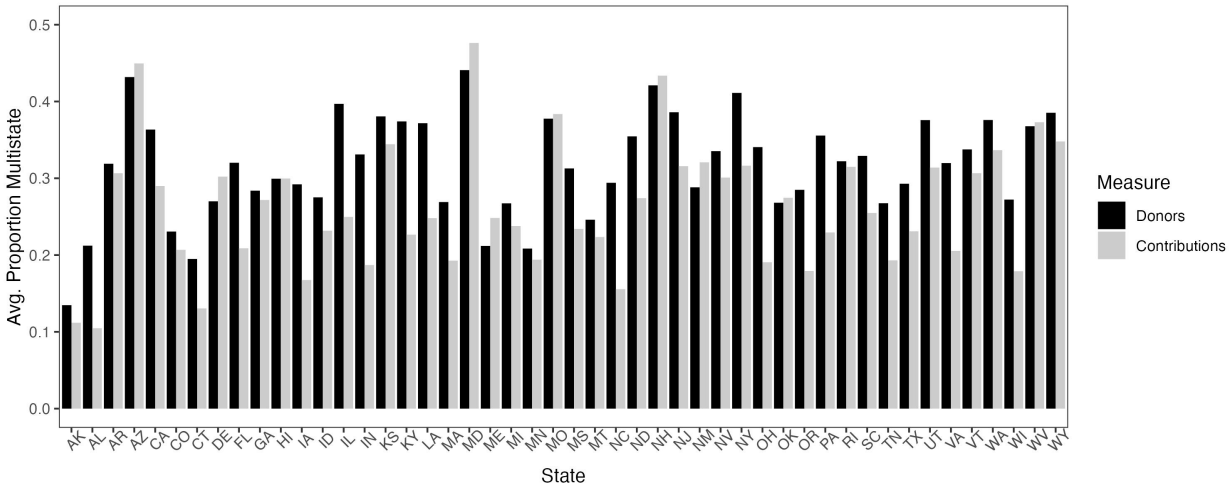


Figure A.5 – Average Proportion of Multistate Donors and Donations per Candidate, by State.



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.

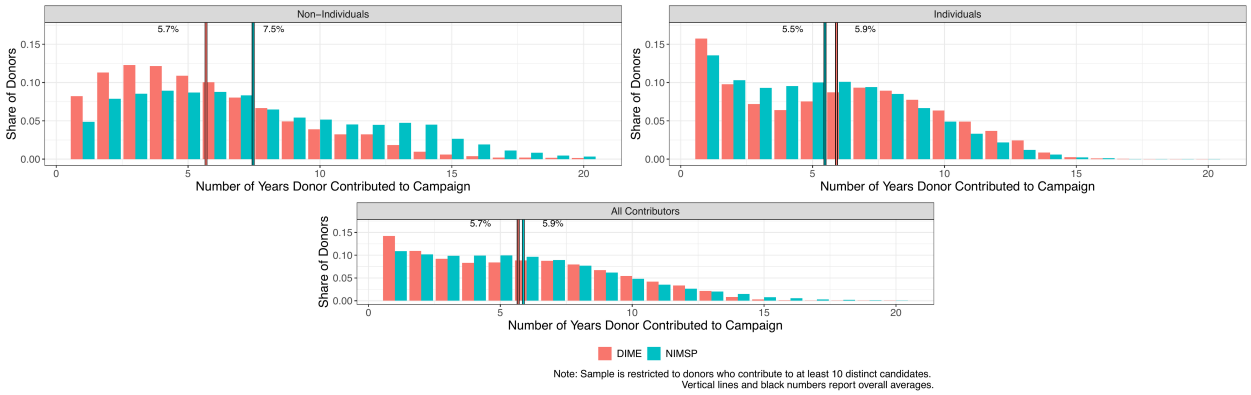
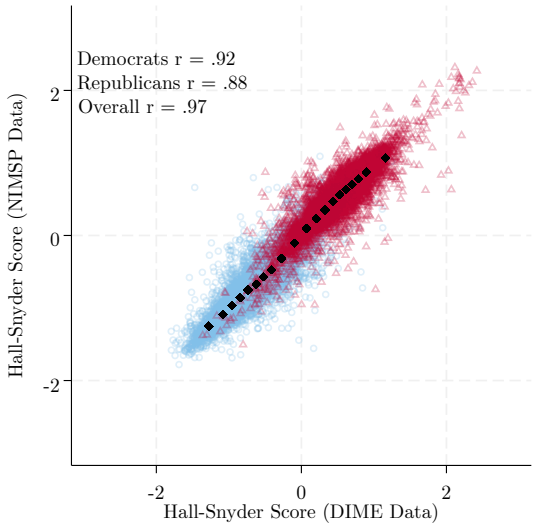


Figure A.7 – Hall-Snyder Scores Generated NIMSP and DIME Data Correlate Highly. This figure shows the correlation between Hall-Snyder scores generated using NIMSP data (vertical axis) and DIME data (horizontal axis).



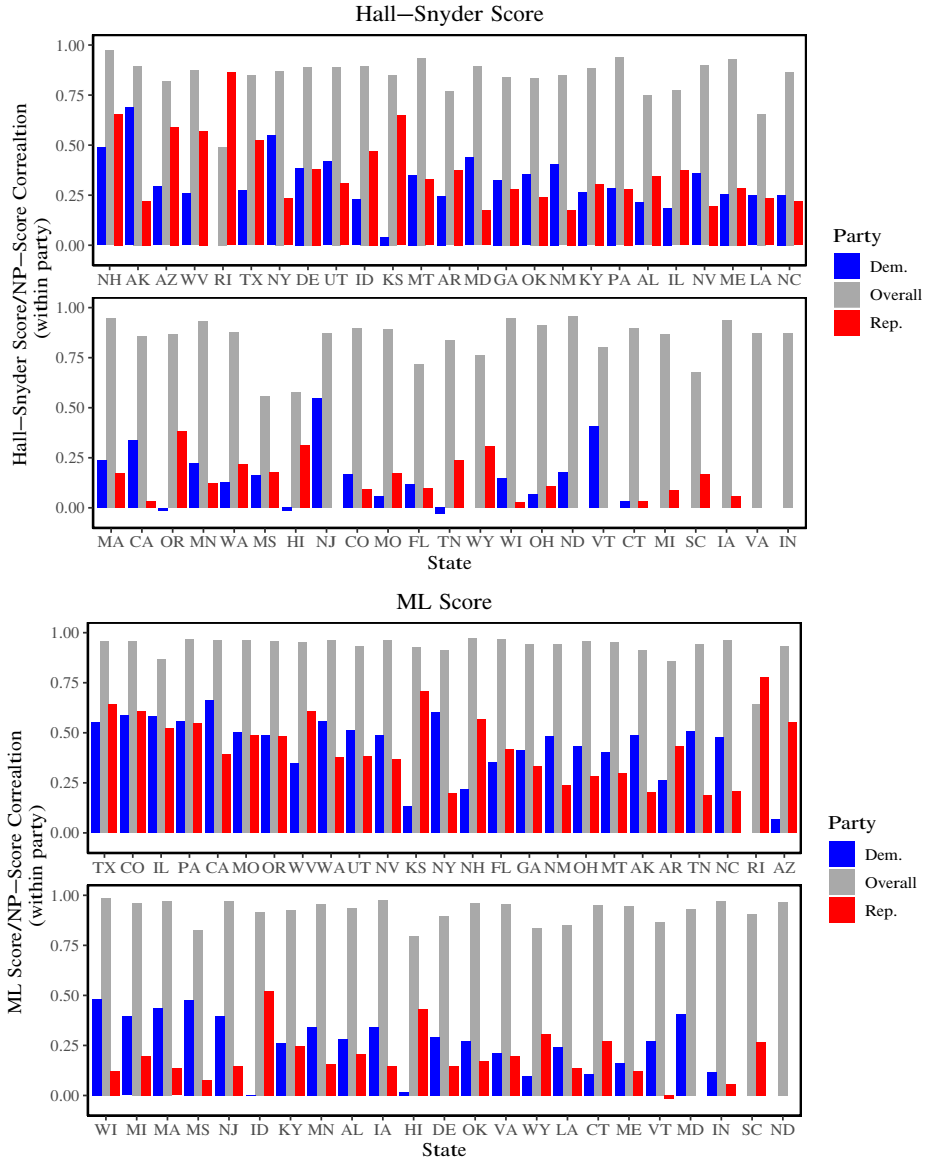
Finally, we re-run our main analyses using the DIME-based Hall-Snyder scores. Overall, the results are highly similar between NIMSP- and DIME-based Hall-Snyder scores. We conclude that our results are both replicable and robust to alternative identity resolution softwares

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:

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

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

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.

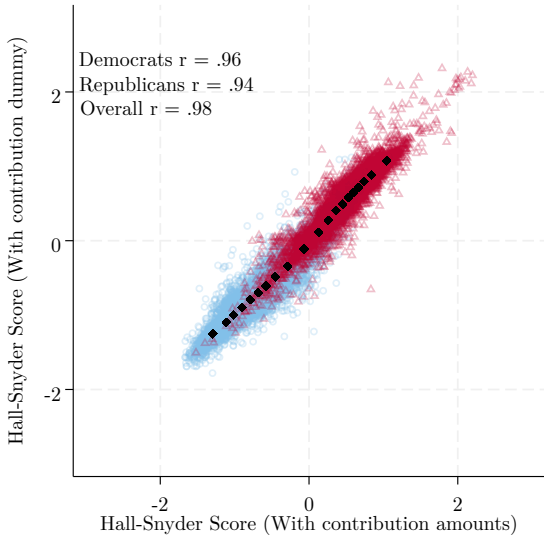
As a further check on sensitivity to contribution limits, we also show that the dollar amount of contributions does not appear to be driving our predictive performance. Figure A.9 shows the correlation between Hall-Snyder scores computed using contribution amounts (horizontal axis) and an alternative 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

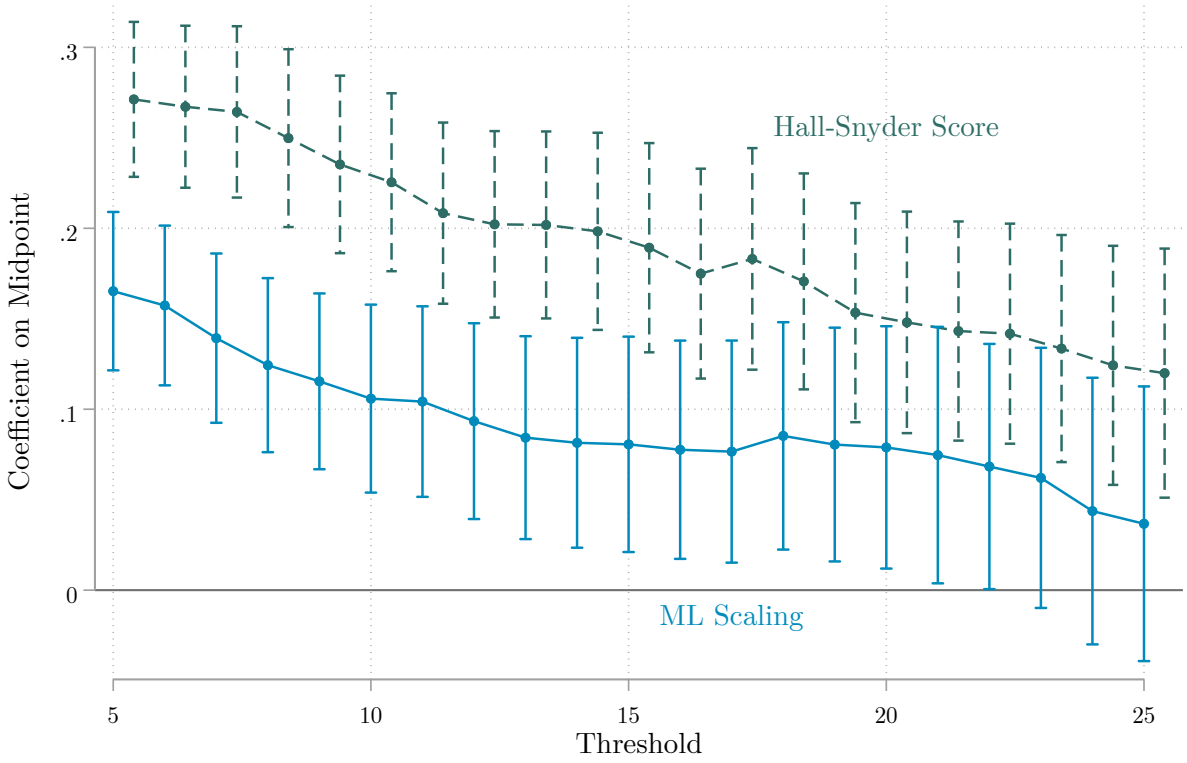
Each cell in this table reports the coefficient on *Relative Centrism* from Equation 5 which is scaled to run from 0 (most extreme) to 1 (most moderate) for each scaling. The sample is restricted to contested primary elections and excludes races in districts where the opposing party received greater 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.

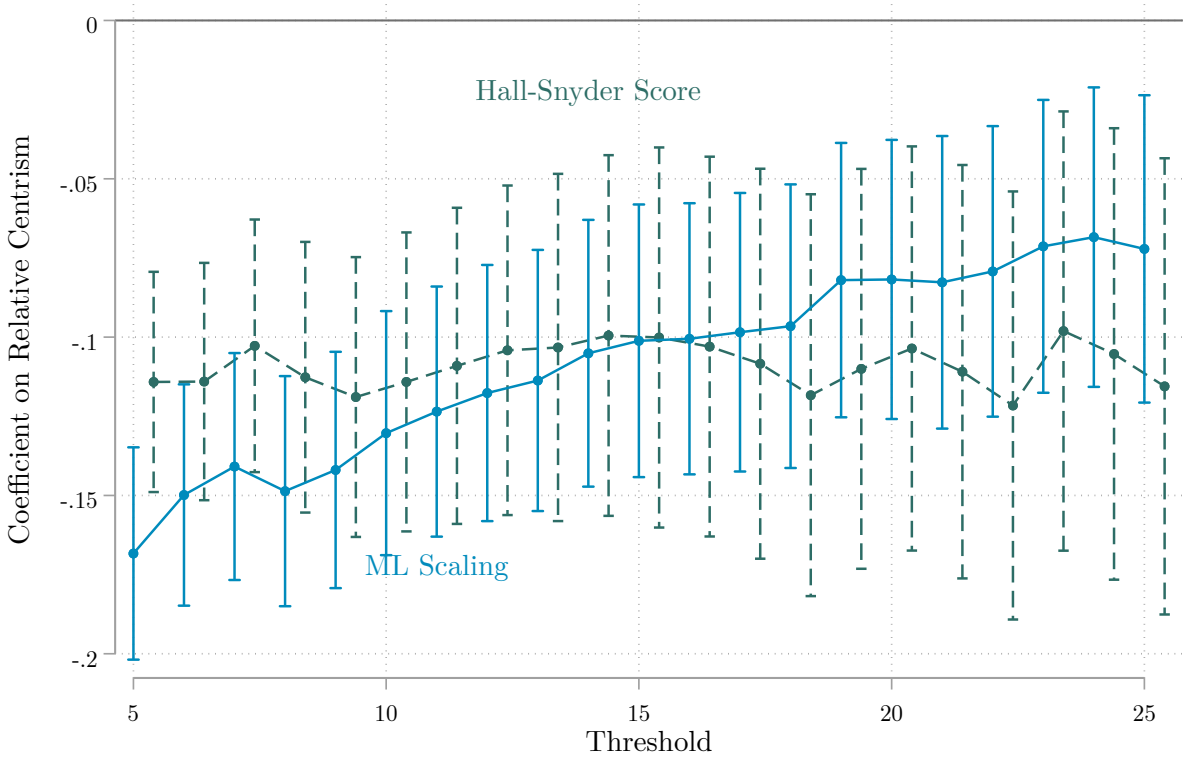
Midpoint Estimates

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.

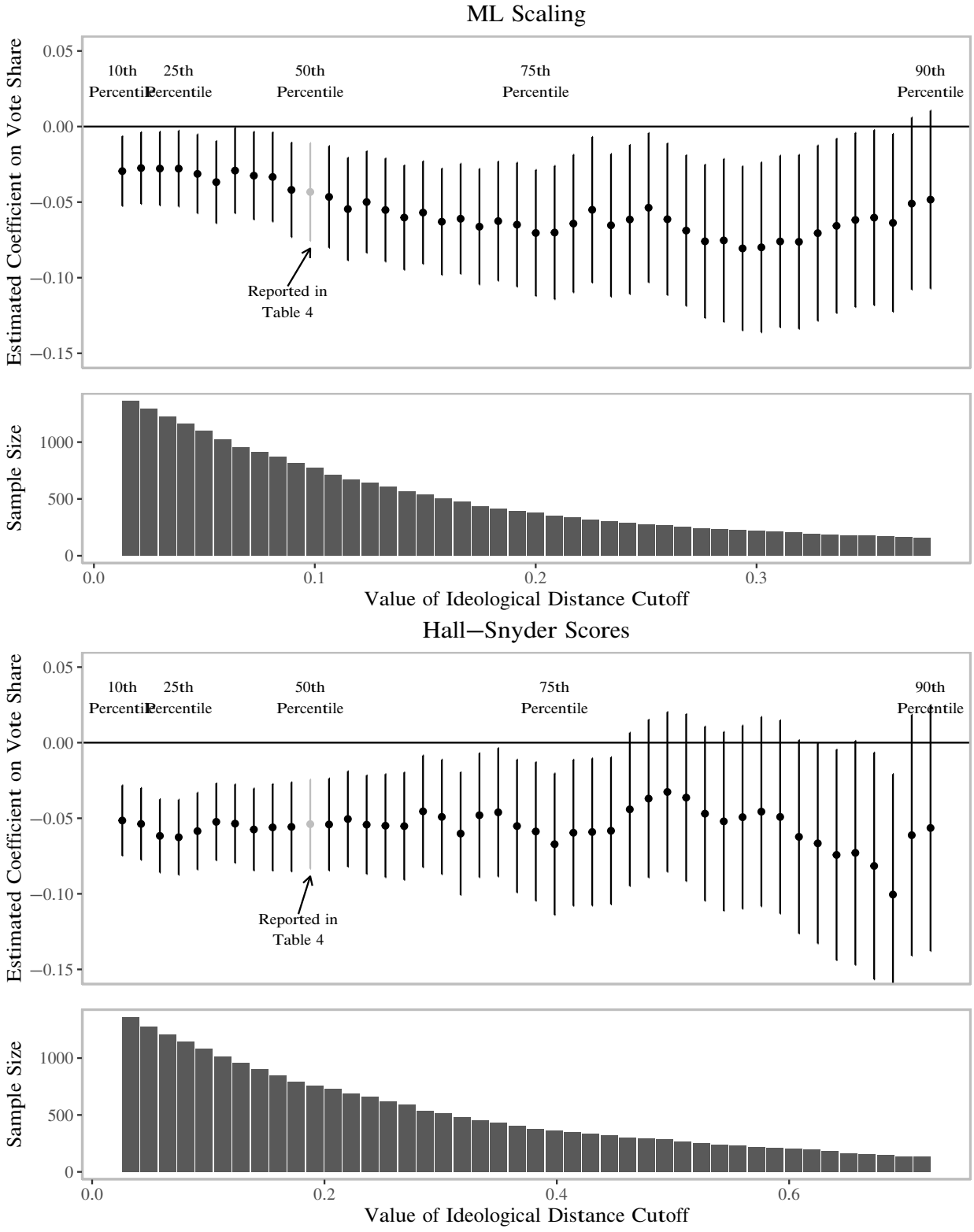
Table A.12 – Effect of Extremist Nominee on Lagged General Election Vote Share, U.S. State Legislatures 2000-2022.

	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)