

# Inferring Roll Call Estimates of Ideology from Campaign Contributions Using Supervised Machine Learning

Adam Bonica\*

March 12, 2016

**Abstract.** This paper develops a generalized supervised learning methodology for inferring roll call scores for incumbent and non-incumbent candidates from campaign contribution data. Rather than use unsupervised methods to recover the latent dimension that best explains patterns in giving, donation patterns are instead mapped onto a target measure of legislative voting behavior. Supervised learning methods applied to contribution data are shown to significantly outperform alternative measures of ideology in predicting legislative voting behavior. Fundraising prior to entering office provides a highly informative signal about future voting behavior. Impressively, contribution-based forecasts based on fundraising as a non-incumbent predict future voting behavior with the same accuracy as that achieved by in-sample forecasts based on votes cast during a legislator's first two years in Congress. The combined results demonstrate campaign contributions are powerful predictors of roll-call voting behavior and resolve an ongoing debate as to whether contribution records can be used to make accurate within-party comparisons.

---

\*Assistant Professor, 307 Encina West, Stanford University, Stanford CA 94305 (bonica@stanford.edu, <http://web.stanford.edu/~bonica>).

Spatial maps of preferences have become a standard tool for the study of politics in recent decades. As scaling methods are applied to an increasingly diverse set of political actors and types of data, political scientists have come to view DW-NOMINATE and related roll call scaling models as benchmark measures of ideology (Poole and Rosenthal, 2007; Clinton, Jackman, and Rivers, 2004). Part of the appeal of these measures is their ability to summarize the lion's share of congressional voting behavior with a single dimension. Indeed, the predictive power of spatial models of voting have shaped our understanding of Congress as fundamentally one-dimensional. This has in turn aided in testing a variety of theories about representation, accountability, and legislative behavior and has fostered their widespread adoption.<sup>1</sup>

A well known limitation of roll call-based measures of ideology is that they are confined to voting bodies. This precludes estimating scores for non-incumbent candidates prior to taking office, which is arguably where such predictions would be most valuable (Tausanovitch and Warshaw, 2016). Only quite recently has the focus on scaling Congress begun to give way as political scientists have sought to extend ideal point estimation to a wider set of institutions and contexts. In recent years, scaling methods have been applied to a ever more varied types of data, including voter evaluations of candidates (Maestas, Buttice, and Stone, 2014; Hare et al., 2015; Ramey, 2016), legislative speech (Beauchamp, 2012; Lauderdale and Herzog, 2015), social media follower networks (Barberá, 2015; Barberá et al., 2015; Bond and Messing, 2015), and campaign contributions (Bonica, 2013, 2014; Hall, 2015).

As the most widely used measure of ideology, DW-NOMINATE remains a common thread in the literature on ideal point estimation. Benchmarking measures based on comparisons with DW-NOMINATE is a standard practice. Although comparisons with an established measure are useful for establishing face validity, it can encourage scholars to misinterpret roll call estimates as the "true" or definitive measures of ideology. In practice, ideal point estimation is typically performed using unsupervised data reduction techniques.<sup>2</sup> The output of roll call scaling models is most accurately understood as a relative ordering of individuals along a *pre-*

---

<sup>1</sup>According to Google Scholar, Poole and Rosenthal's combined work on NOMINATE has generated nearly 10,000 cites.

<sup>2</sup>Partial exceptions include Gerrish and Blei (2012), Lauderdale and Clark (2014), and Bonica (2016) which use semi-supervised methods to identify the dimensionality of roll calls based on issue weights from topic models.

*dictive* dimension that best explains voting behavior in a given voting body. Although widely understood as measures of ideology, this is an interpretation given by the researcher and not reflective of any defined objective built into the model.

In a recent paper, Tausanovitch and Warshaw (2016) evaluate several alternative measures of ideology recovered from survey data, campaign contributions, and social media data with based on comparisons with DW-NOMINATE. They find that most measures successfully sort legislators by party but are less successful in distinguishing between members of the same-party. This leads the authors question the usefulness of these measures for testing theories of representation and legislative behavior or in predicting how non-incumbent candidates would behave in office. In addition to the obvious implications for researchers, this has important policy implications. One of the main rationales for campaign finance disclosure laid out by the Supreme Court in *Buckley v. Valeo* (424 US 1 [1976]) is that it conveys useful information that would allow “voters to place each candidate in the political spectrum more precisely than is often possible solely on the basis of party labels and campaign speeches.” In a recent study, Ahler, Citrin, and Lenz (Forthcoming) cast doubt on the ability of voters to discern ideological differences between moderate and extreme candidates of the same party, suggesting that the disclosure laws have thus far failed to inform voters along the lines outlined in *Buckley*. Meanwhile, other studies have directly challenged the informational benefits of campaign finance disclosure Primo (2013); Carpenter and Milyo (2012). Finding that even sophisticated statistical methods are unable to leverage the informational value of campaign contributors to generate accurate predictions about how candidates would behave if elected would further undermine an important policy rationale for campaign finance disclosure laws.

This paper introduces a new methodological approach for forecasting legislative voting behavior for candidates who have yet to compile a voting record. Rather than using unsupervised methods to recover the dimension that best explains patterns in the behavior at hand, data on revealed preferences are instead mapped directly onto a target measure of legislative voting behavior—in this case, DW-NOMINATE scores. This is done using supervised machine learning methods similar to those used by many social scientists for text analysis (Grimmer and Stewart, 2013; Laver, Benoit, and Garry, 2003). Supervised machine learning methods excel at this task because they are able to “learn” the mapping between predictor variables and the target variable when the target function is unobserved.

The paper proceeds as follows. It begins by motivating the supervised learning approach

with a discussion that highlights a disconnect the ideal point literature between theory and estimation. This is followed by a brief introduction of supervised learning methods and a presentation of the results. The remaining sections discuss issues raised by the results regarding benchmarking and validation unsupervised models.

## 2 Statement of the Problem

The spatial theory underlying ideal point estimation models is known as two-space theory (Cahoon, Hinich, and Ordeshook, 1976). The theory builds on a concept known as issue constraint first defined by Converse (1964) as “a configuration of ideas and attitudes in which the elements are bound together by some form of constraint or functional interdependence.” Practically speaking, the presence of issue constraint means preferences are correlated across issues. If provided with the knowledge of one or two of an individual’s issue positions, an observer should be able to predict the remaining positions with considerable accuracy.<sup>3</sup> Two-space theory holds that issue constraint implies the existence of a higher-dimensional space that contains positions on all distinct issue-dimensions known as the “action space” and a lower-dimensional mapping of issue preferences onto one or two latent ideological dimensions known as the “basic space.” In practice, we only directly observe positions in the action space, leaving the ideological dimensions to be estimated as latent variables.

Enelow and Hinich (1984) and later Hinich and Munger (1996) extend the two-space model to explain how voters can use ideology as an informational shortcut in deciding between candidates. These models begin with the assumption that voters have preferences over an  $n$ -dimensional issue space. The issue positions of candidates are assumed to be “linked” to an underlying ideological dimension. Given a shared understanding of how issues map on the ideological dimension, voters are able to use ideological cues to infer where candidates locate on issue dimensions. From this perspective, ideology is understood as a mechanism for efficiently summarizing and transmitting information about political preferences.

In recent years, a trend has emerged towards viewing ideal point estimation as directly analogous to a class of latent trait models used in the educational testing literature. Although clear parallels exist with respect to estimation, the analogy quickly wears thin. Educational tests are predicated on the notion that individuals possess latent abilities related to intelligence

---

<sup>3</sup>As explained by Poole (2005), “in contemporary American politics the knowledge that a politician opposes raising the minimum wage makes it virtually certain that she opposes universal health care, opposes affirmative action, and so on. In short, taht she is a *conservative* and almost certainly a Republican.”

or aptitude that generate responses to test questions. What distinguishes the most intelligent individuals is an enhanced cognitive ability that allows them identify the correct answers to a series of carefully designed test questions.

Conceptualizing spatial models of politics in similar terms requires making strong assumptions about the data-generating process. To see why, let  $\mathbf{Y}$  be the  $n$  by  $k$  matrix of issue positions of  $n$  individuals on  $k$  issue dimensions and  $\mathbf{X}$  be the  $n$  by  $s$  matrix of individuals' ideal points on the  $s$  ideological dimensions. The presence of issue constraint implies that all issue positions can be represented as  $\mathbf{X}\beta \implies \mathbf{Y}$ , where  $\beta$  is a projection matrix that maps ideal points onto issue dimensions. This implies the existence of a latent ideological space that is exogenous to the preferences and choices it influences. If  $\mathbf{X}$  generates all the issue positions in  $\mathbf{Y}$ , the relative importance or weighting of issues should have no bearing on the dimensionality of ideology. Neither issue salience nor the frequency upon which issues are voted on should matter to how ideal points project onto issue dimensions, which strictly depends on  $\mathbf{X}\beta$ . This might be referred to as the *holographic interpretation* of ideology in that issue preferences are understood as a higher-dimensional representation of information existing in a low dimensional ideological space.

There is reason to doubt such an interpretation. The crux of the problem is that the sources of constraint remains a "black-box" (Poole, 2005). We observe that issue positions are correlated across individuals but lack a basic understanding of why issues are bundled or how issue dimensions map onto the ideological space. More to the point, the holographic interpretation is at odds with statistical methods used to scale ideology. In practice, scaling models work in reverse, starting with data on revealed preferences on issues that are mapped onto a low dimensional predictive space,  $\mathbf{Y} \implies \mathbf{X}\beta$ . The objective is not necessarily to measure some underlying "true" ability or trait expressed in  $\mathbf{Y}$  but rather to construct a low dimensional representation of the information contained in  $\mathbf{Y}$ . In this respect, these models are more similar to multidimensional scaling and related ordination techniques. The most faithful interpretation of  $\mathbf{X}$  is as whatever dimension best explains variation in  $\mathbf{Y}$ . Consequently, changes to the number or relative importance of issue dimensions contained in  $\mathbf{Y}$  can result in changes to  $\mathbf{X}$ . If we allow issue dimensions to be weighted with respect to salience, their relative importance to policy outcomes, or simply the frequency they are voted on, some issues will matter more in defining  $\mathbf{X}$ . Simply put, an issue that is voted on a hundred times will have greater influence on the dimension recovered from a scaling model than an issue that is only voted on once, or

not at all.

**Implications for validation and prediction.** In practice, the output of scaling models is the dimension that best explains variation in the patterns of behavior in the data. In this sense, these models are primarily descriptive in nature as opposed to being designed to measure a target concept. This makes direct comparisons between alternative measures of ideology problematic because neither the mapping function nor the issue weights are observed. As a result, it is difficult to determine whether differences across measures result from measurement error or from systematic differences in how issues are mapped onto the latent dimensions.

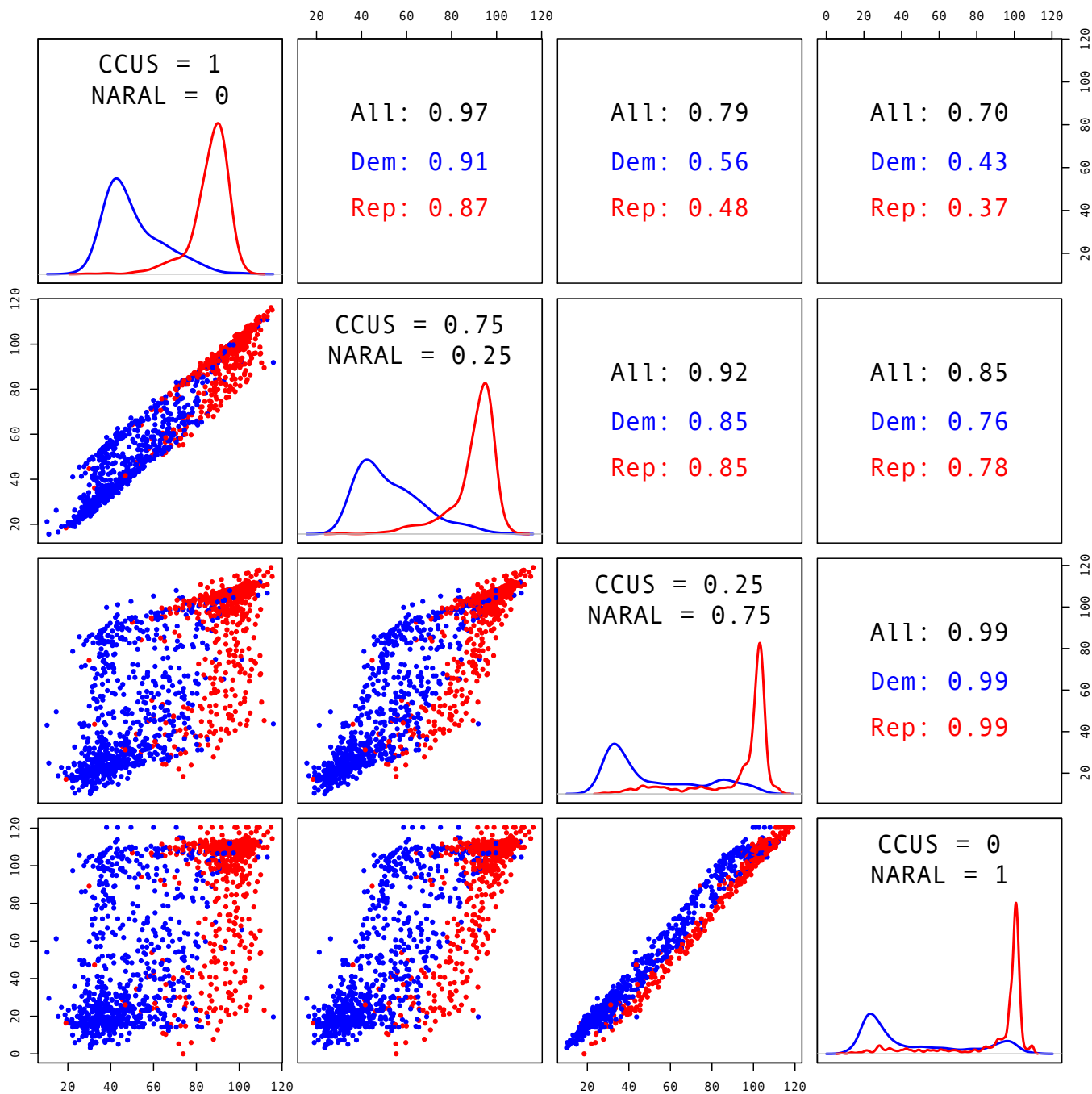
To illustrate, consider a simplified issue space comprised of two issue dimensions. In this example, one issue dimension relates to economic policy and the other relates to social conservatism. Interest group ratings compiled by the US Chamber of Congress (CCUS) and the National Abortion and Reproductive Rights League (NARAL) provide estimates of legislator positions on each issue dimension.<sup>4</sup> Factor analytic techniques can be used to project legislators onto a latent dimension that best explains variation in issue preferences. The somewhat noisier relationship between the CCUS and NARAL scores suggests relatively weak levels of constraint.

Figure 1 compares ideal points projected on the latent dimension recovered using weighted factor analysis under four hypothetical weighting profiles. The two corner scenarios assume that a single issue receives 100 percent of the weight. In the other two scenarios, one issue dimension receives 75 percent of the weight while the other receives 25 percent. Comparing ideal points across scenarios illustrates just how sensitive scaling models can be to how issues are weighted. Depending on the issue weights, the distributions of ideal points on the latent dimension can look very different.

One application where the weighting of issue dimensions comes into play is in bridging across voting bodies. A common identification strategy uses legislators who served in one legislature before entering another as bridge observations. Linear projections are used to re-scale ideal points recovered from voting in state legislatures to the same actors' ideal points recovered from voting in Congress (Shor, Berry, and McCarty, 2010; Windett, Harden, and Hall, 2015). This approach rests on the assumption that the dimension that best explains roll call

---

<sup>4</sup>The adjusted interest group ratings are provided by Groseclose, Levitt, and Snyder (1999) and cover Congress members who served between 1979 and 2008. The scores are averaged across periods so that each legislator is assigned a single score.



**Figure 1:** Pairwise comparisons of interest group ratings under different weighting assumptions

*Note:* The points for legislators are color coded with respect to party. The upper-right panels report the Pearson correlation coefficients between measures overall and within party. The diagonal panels list the weights assigned to each issue dimension and plot the ideal point distributions by party.

voting in a given state legislature is identical to the dimension that best explains roll call voting in Congress and that, after rescaling, differences in ideal points recovered from each voting body are simply a matter of measurement error. If the issue weightings in state legislatures differ from those in Congress, the shared dimensionality assumption will likely be violated.

It is doubtful that the shared dimensionality assumption would hold in most cases. Voting within a legislature is a narrow and somewhat peculiar task. Further complicating matters, the set of questions that legislators are asked to consider is largely endogenous to the voting institution. Both the set of bills that are penned into existence and the subset of those which ultimately make it to the floor are the products of a highly strategic and closely managed agenda setting process (see for example Cox and McCubbins (2006)). Moreover, many issues that are central to state policy, such as education policy, are less of a focus for Congress. On the flip-side, issues related to defense, foreign policy, and trade are almost strictly the domain of Congress.

This problem complicated even further when bridging across measures derived from different types of preference data. In any given Congress, it is rare to see more than a dozen roll call votes on issues directly relating to socially-charged issues such as abortion and same-sex marriage. In contrast, these same issues feature prominently in campaign rhetoric and are a frequent subject of ballot initiatives. PACs and ballot committees that focus on social issues consistently draw large numbers of donors. The likely consequence of this is that positions on social issues will receive more weight when scaling contributions and less weight when scaling Congressional roll calls.

One way researchers have attempted to get around the comparability problem is to use National Political Awareness Test (NPAT) candidate surveys as an intermediary (Shor and McCarty, 2011). First, state legislators and members of Congress are jointly scaled using their NPAT responses. Congress and state legislatures are then each scaled separately using roll call data and projected onto the NPAT common space via an error-in-variables regression model. While this greatly increases the number of available bridge observations and addresses some of the issues related to assumptions about the consistency of behavior when bridge actors move from one chamber to another, the identification strategy still rests on the assumption that scaling models applied to the various legislatures all recover positions along the same latent ideological dimension.

In what follows, I propose a general methodology for mapping revealed preference data



generated in one context onto a target latent dimension recovered from data generated in a different context.

### **3 Supervised Learning Algorithms for Predicting Congressional Voting**

This section outlines the methodology for inferring DW-NOMINATE scores for candidates based on non-roll call sources of data. The idea underlying supervised machine learning is that given a target data set where outcomes are either observed or have been systematically assigned by human coders, an algorithm can "learn" to predict outcomes by recognizing patterns in a corresponding feature set (i.e. matrix of predictor variables).

Two main tasks are involved in using supervised learning models for this purpose. The first is to identify a common source of data that is shared by incumbent and non-incumbents. Nearly all candidates engage in fundraising, making contribution data ideal for this purpose. This positions the modeling strategy developed here to generalize well beyond Congress to the general population of candidates and political elites across the nation. The second task is to determine which supervised learning algorithms are best suited for the data. In this case, the target variable (DW-NOMINATE) is measured along a continuous dimension, which suggests a regression based modeling approach.

Machine learning methods have become an increasingly popular tool in recent years for social scientists dealing with data sets with many hundreds or thousands of variables (Hainmueller and Hazlett, 2014; Grimmer and Stewart, 2013; Cantú and Saiegh, 2011). By far, the most common application for these models has been text analysis. In a typical scenario, a researcher might begin with a sample of a few hundred hand-coded documents sorted into a predefined set of topics. The hand-coded documents are used to train a supervised machine learning model. The trained model can then be used to infer the topics for remaining documents. This provides an efficient means of topic coding large corpuses of text. In an alternative arrangement, a model might be trained to classify legislators by party or ideological groupings based on a corpus of legislative text, where each document associated with a legislator (Yu, Kaufmann, and Diermeier, 2008; Diermeier et al., 2012).

The supervised machine learning task undertaken here can be thought of in a similar vein. The candidate-contributor matrix takes on a nearly identical structure to that of a document-term matrix, where the contribution profiles associated with candidates can be thought of as

documents and contributors as words. Given a training set of candidates that have been assigned DW-NOMINATE scores, the model will attempt to discern the ideological content of contributors, just as models applied to legislative text attempt to discern the ideological content of words.

In this framework, the set of candidates with DW-NOMINATE scores are used to train the model. Insofar as information relevant for predicting roll call behavior is present in the contribution matrix, it becomes a matter of training a model to learn from the observed patterns of giving. To state the problem more formally, suppose there are  $N_{train}$  candidates for whom DW-NOMINATE scores are observed ( $i = 1, \dots, N_{train}$ ) and another  $N_{test}$  candidates for whom DW-NOMINATE scores are not available. Let  $\mathbf{Y}_{train}$  be an  $N_{train}$ -length vector of observed DW-NOMINATE scores and let  $\mathbf{W}_{train}$  be an  $N_{train} \times m$  matrix of contribution amounts. The remaining  $N_{i \in test}$  candidates represents values to be predicted. The model assumes there is some unobserved target function,  $f(\cdot)$ , that best describes the relationship between  $\mathbf{Y}_{train}$  and  $\mathbf{W}_{train}$ ,

$$\mathbf{Y}_{train} = f(\mathbf{W}_{train}). \quad (1)$$

The supervised learning algorithm attempts to learn this relationship by estimating a function,  $\hat{f}(\cdot)$ , that approximates  $f(\cdot)$ .  $\hat{f}(\cdot)$  is then used to infer values of  $\mathbf{Y}_{test}$  from  $\mathbf{W}_{test}$ ,

$$\hat{\mathbf{Y}}_{test} = \hat{f}(\mathbf{W}_{test}). \quad (2)$$

Although several regression-based supervised machine learning methods would be applicable here, support vector regression (Drucker et al., 1997; Smola and Schölkopf, 2004) and random forests (Breiman, 2001) are particularly well-suited for the task at hand.

**Support vector regression.** Support vector regression is a generalization of support vector machines (SVM) to real-valued functions. The objective of support vector regression is to find a function  $\hat{f}(\cdot)$  that minimizes the number predicted values with residuals larger than  $\epsilon$ . This differs from standard regression models in that the loss function tolerates deviations where  $|\hat{y} - y| \leq \epsilon$ , with only deviations  $|\hat{y} - y| > \epsilon$  being penalized. This is known as an epsilon-insensitive loss function,

$$|\hat{\xi}_i|_\epsilon = \begin{cases} 0 & \text{if } |\hat{y} - y| \leq \epsilon \\ |\hat{y} - y| - \epsilon & \text{if } |\hat{y} - y| > \epsilon \end{cases} \quad (3)$$

where the value of  $\epsilon$  either set a priori or, as is more commonly the case, treated as a tuning parameter during computation. To estimate a linear regression,

$$f(x) = \sum_{i=1}^{N_{train}} (\alpha_i^* - \alpha_i) k(x_i, x) + b \quad (4)$$

where  $k(\cdot)$  is the kernel function and  $b$  is the “bias” term. A linear kernel  $k(x_i, x_j) = x_i^\top x_j$  is used because it suits contribution data well. The SVM algorithm solves the constrained optimization problem,

$$\begin{aligned} \arg \max_{\alpha^*} W(\alpha^*) &= \sum_{i=1}^{N_{train}} (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) k(x_i, x_j), \\ \text{subject to} \\ \sum_{i=1}^{N_{train}} (\alpha_i^* - \alpha_i) &= 0, \\ \alpha^* &\in \left[ 0, \frac{C}{M} \right], \\ \sum_{i=1}^{N_{train}} (\alpha_i^* - \alpha_i) &< C \cdot v. \end{aligned} \quad (5)$$

**Random forests.** Random forests are an ensemble approach to supervised learning that operates by constructing many random decisions trees from the input data and aggregating over the output to generate predictions. The main advantages of random forests are efficiency with large datasets, resistance to overfitting, and built-in estimates of variable importance, which aids in feature analysis. (See Breiman (2001) for an overview.)

## 4 Model Training

**Constructing the training set.** The analysis here focuses on candidates running for federal office during the 1980-2014 election cycles. The common-space DW-NOMINATE scores, which provide estimates from a joint scaling of the House and Senate for the 1-113th Congresses, are used as the target variable. Unlike chamber-specific scalings of the House or Senate that model dynamic legislator ideal points, the common-space scores are static. The data on campaign contributions is from the Database on Ideology, Money in Politics, and Elections (DIME) Bonica (2016). The DIME data covers a period from 1980-2014 and contains records for 72,065 candidates from state and federal elections (1,718 of whom have DW-NOMINATE scores). In addition, indicator variables for three basic candidate traits—party, home state, and gender—are included in the feature matrix.

**Feature selection.** Given the large size of the potential feature set, donors that did not meet the threshold of giving to at least 15 distinct candidates included in the training set (e.g. that have DW-NOMINATE scores) were thinned from the feature set. This reduces the number of features to 63,992. Recursive feature elimination techniques, which rely on iterative methods to narrow the feature set, were also used in building the model.

While feature selection allows for improved handling of the sparsity in the contribution matrix, it does risk excluding potentially useful information from the millions of less active donors. In order to avoid discarding information from donors who do not meet the threshold for inclusion, I employ feature extraction methods. Specifically, I construct an  $n \times m$  matrix that summarizes the percentage of funds a candidate raised from donors that fall within  $m = 10$  ideological quantiles. This is done by calculating contributor coordinates from the dollar-weighted ideological average of contributions based on the DW-NOMINATE scores of the recipients and then binning the coordinates into deciles. The contributor coordinates are calculated in a manner consistent with the cross-validation scheme by removing rows for candidates in the held-out set for each round. With the coarsened contributor scores in hand, I then calculate the proportions of contribution dollars raised by each candidate from each decile of donors. The resulting  $n$  by 10 matrix of decile shares is then included in the feature set.<sup>5</sup>

**Model fitting.** The random forest regression was trained using the *caret* package in *R* (Kuhn, 2008). The support vector regression model was trained using the Liblinear library Fan et al. (2008). Repeated  $k$ -fold cross-validation is used in training ( $k = 10$ ). This is done by partitioning the sample into  $k$  groups and repeatedly fitting the model each time with one of the  $k$ -sets held out-of-sample. This process is repeated five times on different partitions of the data and results are averaged over rounds.

One thing to note is that the DW-NOMINATE scores are treated as known quantities despite being measured with error. This makes assessing model fit slightly less straightforward as it is unclear the extent to which cross-validation error reflects measurement error in the target variable. The presence of measurement error is relatively common for supervised machine learning exercises, especially those that rely on human coding to generate a training the set. Although measurement error in the target variable can lead to overfitting, regularized kernel-regression methods and random forests are less prone overfitting in the presence of low levels of measurement error.

---

<sup>5</sup>Alternatively, using common-space CFscores for feature extraction performs nearly as well.

## 5 Results

This section reports results to assess the predictive performance of the support vector regression model. For purposes of comparison, fit statistics are reported for common-space CFscores, another set of contribution-based scores estimated using a structural model applied to federal PAC contributions (IRT CFscores), Turbo-ADA interest group ratings compiled by Americans for Democratic Action and normalized by Groseclose, Levitt, and Snyder (1999), NPAT scores based on candidate surveys from the 1996 elections (Ansolabehere, Snyder, and Stewart, 2001), Shor and McCarty (2011) state legislator ideal points based on roll call voting in state legislatures, and two alternative roll call measures developed by Bailey (2013) and Nokken and Poole (2004).

Lastly, I report results from a supervised version of the CFscore model that is estimated in a manner akin to the Wordscores algorithm (Laver, Benoit, and Garry, 2003), where candidates with DW-NOMINATE scores act as the reference documents. To estimate the scores, donors are assigned ideal points based on the dollar-weighted average DW-NOMINATE score of their recipients. The process is then reversed and scores for candidates are calculated based on the dollar-weighted average of their contributors. Similar to the other supervised models, and 10-fold cross-validation is used to assess model performance. The scores reported below are predicted out-of-sample so that a legislator's DW-NOMINATE score does not factor into the estimates for their contributors.<sup>6</sup>

Several of the alternative roll call measures rely on the same underlying data as DW-NOMINATE to scale legislators but make different modeling assumptions. The Bailey scores are estimated use a scaling model similar to that of DW-NOMINATE but incorporate additional data on position-taking by non-legislative actors to bolster identification. The Nokken-Poole scores are a period-specific measure derived from DW-NOMINATE scores. Using the set of roll call parameter estimates recovered from DW-NOMINATE to fix the issue space, the technique estimates congress-specific ideal points for legislators based on voting during each two-year period. As such, these scores represent in-sample estimates of DW-NOMINATE based on subsets of a legislator's voting history. The Nokken-Poole estimates appear twice in the results. First with the observations spanning the course of legislators' careers then as a fixed score based on a legislator's first term in Congress.<sup>7</sup> The first term Nokken-Poole DW-NOMINATE scores

---

<sup>6</sup>This scaling model is similar to the one used by Hall (2015).

<sup>7</sup>Only first term scores for legislators that served in more than one Congress are included.

	<u>All Cands</u>			<u>Dem Cands</u>			<u>Rep Cands</u>		
	R	RMSE	N	R	RMSE	N	R	RMSE	N
<b><u>Cross-validated</u></b>									
Random Forest	0.97	0.10	1718	0.81	0.10	874	0.82	0.10	838
Support Vector Regression	0.96	0.11	1718	0.78	0.11	874	0.77	0.11	838
Supervised CFscores	0.94	0.14	1718	0.67	0.15	874	0.75	0.13	838
<b><u>In-Sample</u></b>									
Random Forest	0.99	0.04	1718	0.98	0.04	874	0.98	0.04	838
Support Vector Regression	0.99	0.05	1718	0.96	0.05	874	0.97	0.04	838
<b><u>Roll Call Measures</u></b>									
Nokken-Poole (Dynamic)	0.97	0.09	9200	0.91	0.07	4954	0.83	0.11	4227
Nokken-Poole (First Term)	0.96	0.10	1488	0.87	0.09	763	0.84	0.12	721
Bailey Scores (Dynamic)	0.92	0.16	12724	0.78	0.15	6523	0.64	0.16	6161
Bailey Scores (Mean)	0.89	0.18	1662	0.80	0.17	844	0.66	0.18	814
Turbo-ADA	0.90	0.17	1444	0.69	0.18	762	0.59	0.15	681
Shor-McCarty	0.93	0.17	226	0.58	0.17	105	0.51	0.16	121
<b><u>Alternative Measures</u></b>									
Common-space CFscores	0.91	0.17	1718	0.52	0.19	874	0.68	0.14	838
IRT CFscores	0.89	0.18	1275	0.63	0.19	668	0.53	0.17	605
NPAT (1996)	0.92	0.16	257	0.77	0.16	115	0.63	0.16	142

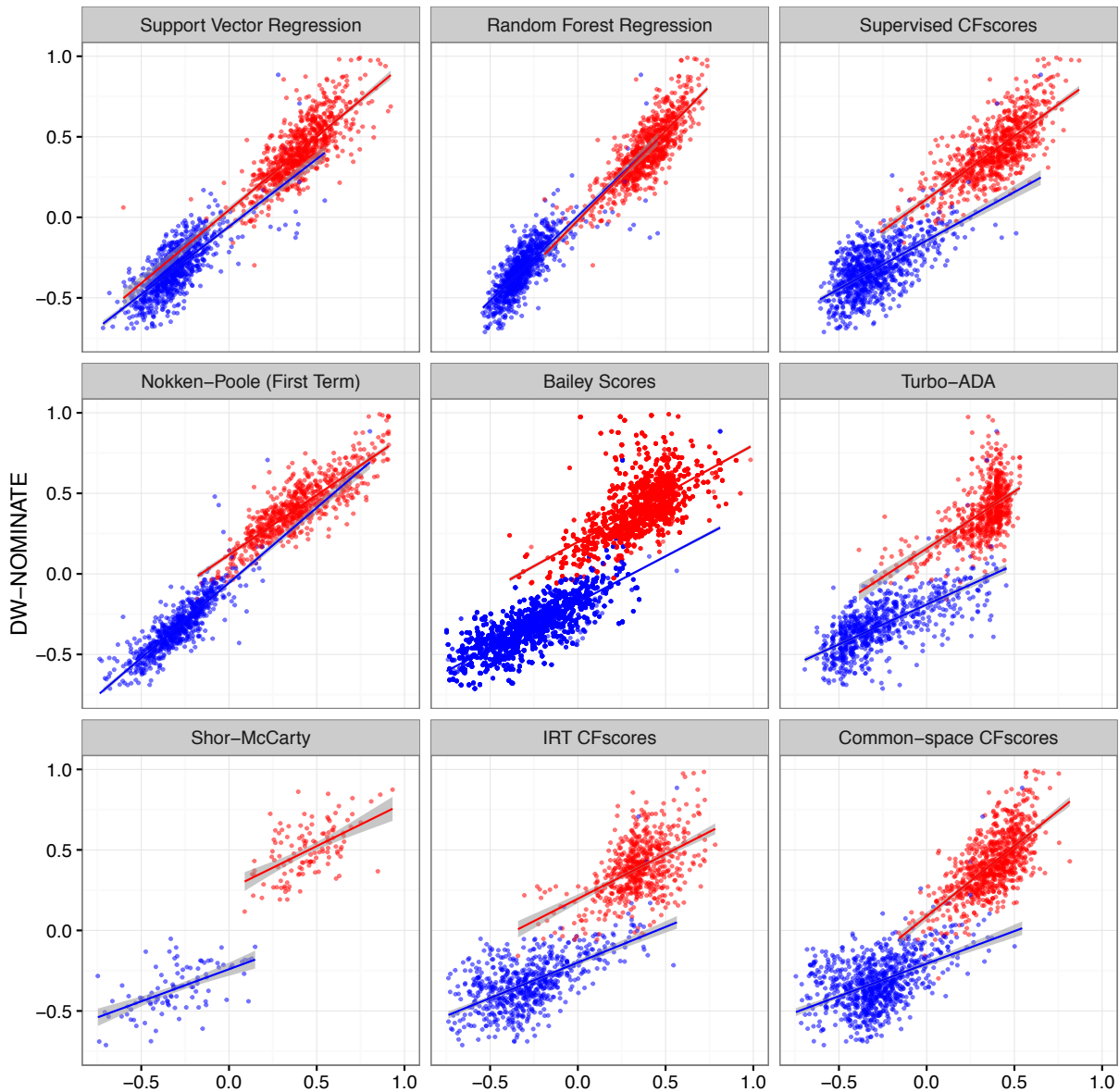
**Table 1:** Predicting DW-NOMINATE Scores: Fit statistics for alternative measures of ideology.

are a particularly informative benchmark for assessing predictive accuracy. It tells us how well voting patterns observed during the first two year in Congress predicts voting behavior over the course of a legislative career.

Table 1 reports comparisons with DW-NOMINATE for the supervised methods and several alternative measures of ideology. Note that model fit is defined here in terms of similarity with DW-NOMINATE scores. For the supervised models, cross-validated and in-sample fit statistics are reported separately. (For the remainder of the paper, the cross-validated estimates are used throughout.) For all other measures, the fit statistics are based on comparisons after being projected onto the DW-NOMINATE scores.

The supervised models perform well in explaining DW-NOMINATE scores, overall and within party, with the random forest regression model doing best overall. The supervised learning models significantly outperforms both common-space CFscores and the PAC-based IRT CFscores, both of which are based on campaign contributions. The predictive accuracy of the supervised models even exceeds that of measures derived from congressional roll call votes. They outperform the Turbo-ADA and Bailey scores by sizable margins.<sup>8</sup> Of the included roll calls measures, the Shor-McCarty scores are the only measures based on non-congressional vote data. They also exhibit the weakest within party correlations, speaking to the challenges inherent in bridging across institutions even when we observe bridge actors engaging in the

<sup>8</sup>Note that model performance is narrowly defined here in terms of similarity with DW-NOMINATE. The lower classification rates associated with the Bailey scores reflects a deliberate departure from the modeling assumptions of DW-NOMINATE.



**Figure 2:** Comparing measures of legislator ideology against DW-NOMINATE scores. *Note: The scales for non-supervised methods have been rescaled for purposes of comparison. Linear trend lines are fit separately for each party.*

same type of behavior in both settings.

Perhaps most telling is that the supervised models are on par with the Nokken-Poole first term estimates in terms of predictive accuracy. This demonstrates that *it is possible to infer a legislator's DW-NOMINATE score from her contribution records just as accurately as we can from observing how she votes during her first two years in Congress.*

Figure 2 presents the relationships between measures as a series of scatter plots. The shaded trend lines show the linear fit by party. As compared with DW-NOMINATE, all of

the independent measures exhibit increased levels of partisan overlap.<sup>9</sup> This suggests that DW-NOMINATE may tend to overstate the extent to which the parties in Congress have polarized. In contrast, both supervised measures appear to successfully capture the gap between parties present in DW-NOMINATE, which helps to explain their higher overall correlations.

**Classifying roll call votes.** Another way to compare predictive accuracy across ideal point measures is to calculate the percentage of votes that can be correctly predicted with a linear classifier (Poole, 2000; Poole and Rosenthal, 2007).<sup>10</sup> Table 2 reports the percentage of votes correctly classified and the aggregate proportional reduction in error (APRE) for roll call voting in the House and Senate for the 96-113th Congresses. Only measures for which scores are available for the majority of the period are included. The table also includes the classification rate associated with a partisan model that assumes each legislator always votes with the majority of her party. This provides a baseline for evaluating how well a given measure improves classification over partisan affiliation. At the other extreme, the classification rate associated with the first dimension of DW-NOMINATE provides an effective upper limit for how well a single dimension can successfully predict vote choices. Legislators who switched parties during this period are excluded from the analysis. (DW-NOMINATE assigns separate ideal points based on votes casts before and after a legislator switched parties, but most of the other measures do not.) Following Poole and Rosenthal (2007), lopsided votes with winning margins greater than 97.5 percent are excluded.

The table orders measures with respect to their success in classifying roll call outcomes, from best to worse. It shows the random forest model to be second only to DW-NOMINATE itself, even outperforming other roll call measures that are estimated in-sample. Notably, the random forest model outperforms the first term Nokken-Poole scores in predicting roll call behavior. The difference in classification rate between DW-NOMINATE and the random forest model is about half a percentage point.

Figure 3 tracks correct classification (joint with the House and Senate) for the partisan model, DW-NOMINATE, and the random forest model across time. The model fit associated with the random forest model relative to DW-NOMINATE has remained more or less stable over the period. Also of note is that while the partisan model provides a natural baseline,

---

<sup>9</sup>One possible explanation for this pattern is the high percentage of procedural votes taken on the floor which are often voted on along party lines (Roberts and Smith, 2003).

<sup>10</sup>For each roll call, the cutting-line procedure draws a maximally classifying line through the ideological map that predicts that those voting "yea" are on one side of the line and those voting "nay" are on the other.



	House	Senate
DW-NOMINATE	0.9 (0.703)	0.887 (0.662)
Random Forest	0.894 (0.687)	0.881 (0.644)
Nokken-Poole (First Term)	0.892 (0.68)	0.879 (0.638)
Support Vector Regression	0.891 (0.677)	0.875 (0.626)
Supervised CFscores	0.884 (0.657)	0.877 (0.633)
Common-space CFscores	0.883 (0.653)	0.874 (0.623)
Bailey Scores (Mean)	0.879 (0.641)	0.852 (0.558)
Turbo-ADA	0.873 (0.621)	0.857 (0.575)
Party	0.87 (0.616)	0.844 (0.536)

**Table 2:** Percentage of Votes Correctly Classified (96th - 113th Congresses)

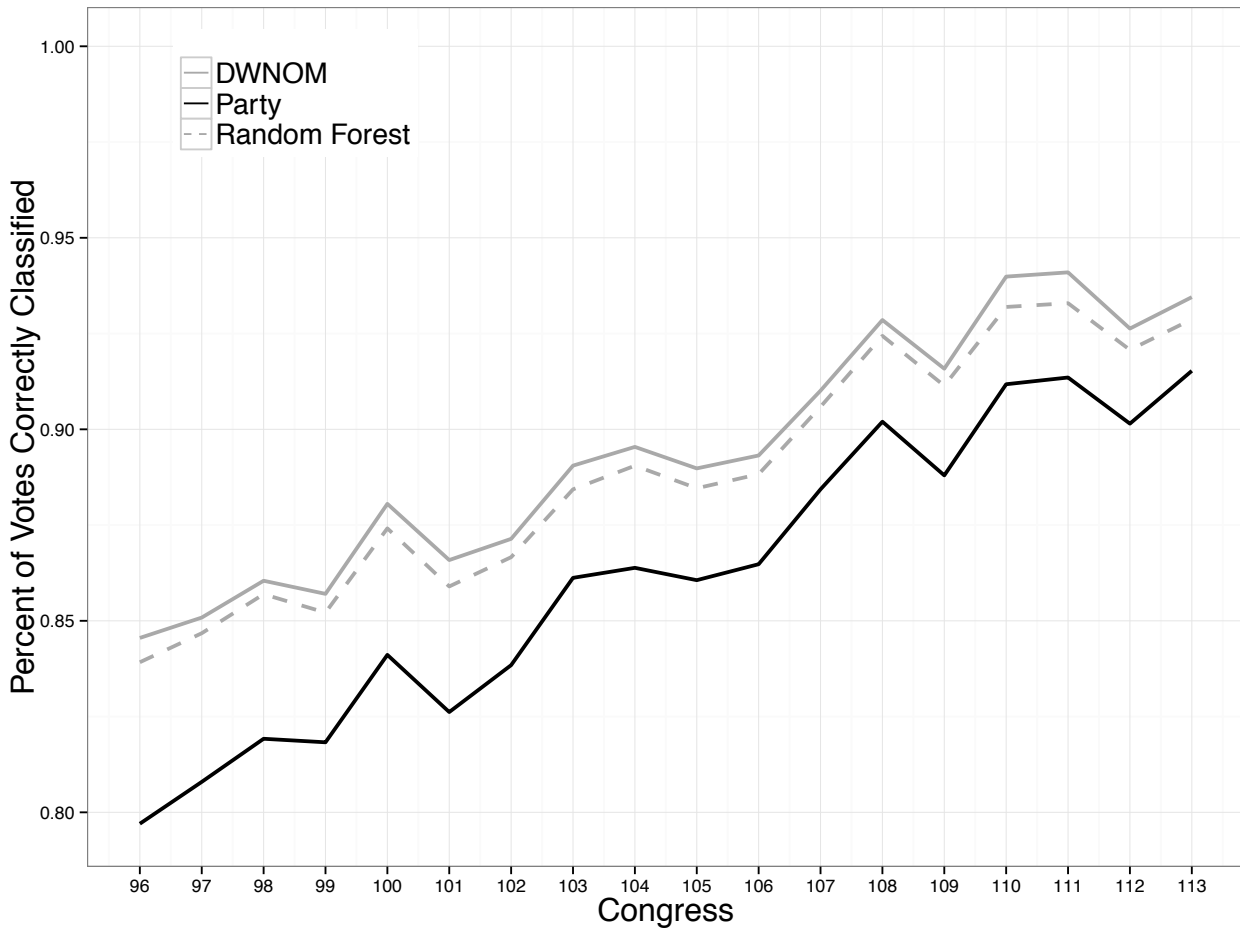
*Note: Aggregate proportional reduction in error (APRE) is in parentheses.*

it is far from static during the period of analysis. The correct classification rate for the House associated with the partisan model increased from 0.80 to 0.92 during the 96-113th Congresses. The increase was even more pronounced in the Senate, growing from 0.76 to 0.91 over the same period. Meanwhile the boost in classification associated with DW-NOMINATE over the partisan model has shrunk from 0.045 to 0.018 in the House and from 0.065 to 0.033 in the Senate.

## 6 Forecasting Congressional Roll Call Measures

A core objective of the supervised learning approach is to forecast future voting behavior of non-incumbents based on data generated observed prior entering Congress. Bonica (2014) finds that scores assigned to non-incumbents based on their fundraising prior to entering office are highly correlated with scores assigned based on fundraising after entering office. This suggests that fundraising before and after entering office conveys much of the same information about candidate locations.

Since the availability of DW-NOMINATE scores is restricted to candidates who have served in Congress, model performance is assessed based on the relationship with future DW-NOMINATE scores for successful candidates. To facilitate comparisons, I separate out contributions made

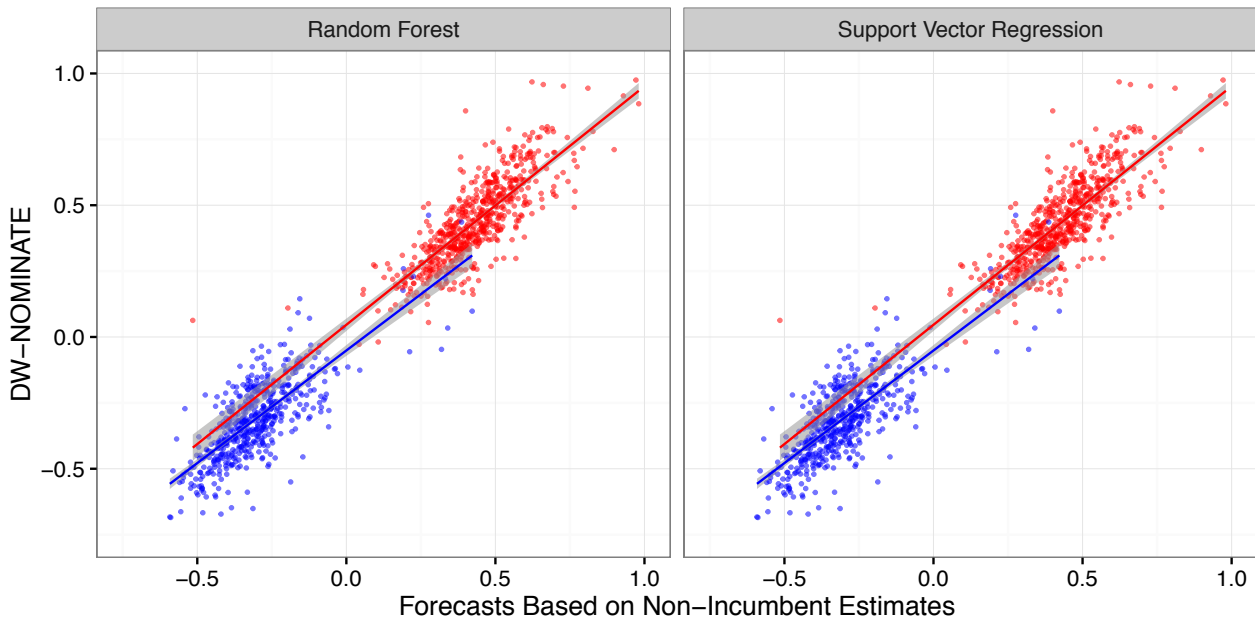


**Figure 3:** Correct Classification by Congress

to candidates before and after they entered Congress. In this setup, candidates who transition from non-incumbents to incumbents enter the data twice as independent row observations. I then retrained the models on fundraising by incumbents, with the rows for non-incumbents held completely out of sample. The non-incumbent scores were then inferred from the model trained on incumbents.

Figure 5 plots the predictions for the held-out sample of non-incumbents against their future DW-NOMINATE scores. Table 3 reports the same fit statistics as above for the held-out sample of non-incumbents. The first row reports the fit for the out-of-sample predictions from the supervised models. The results are in line with those presented in Table 1. They show that fundraising prior to entering office can accurately predict future DW-NOMINATE scores. The overall correlation is 0.97 for both measures. Again, this compares favorably with the Nokken-Poole first term estimates.

Examining the residuals for outliers proves informative. Among the largest outliers are Greg Laughlin (D-TX), Zell Miller (D-GA), and Ben Nighthorse Campbell (D-CO). Laughlin and Nighthorse Campbell both were originally elected as Democrats before joining the Re-



**Figure 4:** Non-incumbent estimates of candidate and future DW-NOMINATE scores

publican Party. Zell Miller ran for unsuccessfully for the Senate during the early 1980's, later served as governor of Georgia, and was appointed to the Senate in 2000 by his successor. He is perhaps best known for his role as a keynote speaker at the 2004 Republican National Convention. These examples are of the type that we should expect to deviate from predictions made from contributions raised as non-incumbents.

The results demonstrate that fundraising prior to entering office provides a highly informative signal about future voting behavior. Impressively, it is nearly as predictive of future voting as the votes cast during the first two-years in Congress.

## 7 Feature Analysis

The random forest model has a built-in algorithm that ranks variables with respect to their importance to the model. The variable importance scores can help provide insight into which types donors are most important in mapping candidates onto the target variable. Table 4 lists the top 20 federal PACs ranked by their importance to the model.<sup>11</sup> It also reports the number

<sup>11</sup>Note that several individual donors made it onto the list but were excluded from the table.

	All Non-Incumbents		Dem Non-Incumbents		Rep Non-Incumbents	
	R	RMSE	R	RMSE	R	RMSE
Random Forest	0.97	0.09	0.81	0.09	0.84	0.09
Support Vector Regression	0.97	0.10	0.77	0.10	0.81	0.10

**Table 3:** Forecasting DW-NOMINATE Scores: Cross validated fit statistics for held-out sample of non-incumbent candidates.

	Variable Importance	N. Recips.	Pct to Reps	Avg. DWNOM	Std. Dev. DWNOM
Council For Citizens Against Gov. Waste	82.4	113	0.01	-0.46	0.10
Consumer Federation of America	82.4	121	1.00	0.61	0.14
Blue Dog Democrats	68.0	119	0.01	-0.19	0.11
American Security Council	67.8	320	0.70	0.25	0.25
NRCC	65.4	751	1.00	0.36	0.15
VFW PAC	59.2	827	0.54	0.08	0.34
National Education Association	59.2	1008	0.06	-0.31	0.20
Democrats Win Seats	56.7	157	0.00	-0.22	0.11
AFL CIO	52.1	788	0.18	-0.28	0.25
Boll Weevil PAC	47.6	61	0.00	-0.12	0.11
National Rural Letter Carriers	47.5	1069	0.25	-0.16	0.32
National Alliance For Political Action	45.8	94	0.03	-0.45	0.20
Active And Retired Federal Employees	45.1	1091	0.19	-0.20	0.31
Intl Union of Bricklayers and Allied Craftsmen	43.1	214	0.02	-0.37	0.15
Victory Now PAC	43.0	153	0.00	-0.23	0.11
Harvest PAC	42.4	69	0.03	-0.15	0.10
United Brotherhood of Carpenters And Joiners	42.0	988	0.12	-0.26	0.25
Railway Clerks Political League	41.9	787	0.06	-0.31	0.20
DRIVE PAC (Teamsters)	41.9	976	0.06	-0.30	0.23
Hoyer For Congress	40.7	336	0.00	-0.27	0.14
Brady Campaign To Prevent Gun Violence	40.6	427	0.08	-0.32	0.22
Grassroots Organizing Acting & Leading	40.4	142	0.00	-0.24	0.11
Democrats For The '80's	38.9	338	0.00	-0.30	0.14
National Right To Life	38.1	730	0.97	0.38	0.21
Right To Work	38.0	676	0.02	-0.31	0.12
Conservative Victory Fund	37.7	236	0.99	0.47	0.14

**Table 4:** Random Forest Variable Importance

of distinct recipients supported by the PAC, the mean and standard deviation of the their recipients' DW-NOMINATE scores by amount, and the percentage of contribution dollars made to major party candidates going to Republicans.

Most of the organizations on the list tend to donate primarily to candidates from one or the other party. The top two features are organizations that locate to the extremes of the parties, with the Council for Citizens against Government Waste on the right and the Consumer Federation of America on the left. The mean score for each party during the period is -0.32 ( $sd = 0.15$ ) for Democrats and 0.39 ( $sd = 0.18$ ) for Republicans. Many of the highest ranked features appear to discriminate within party. Tellingly, among the highest ranked features are the PACs setup to support the Blue Dog Democrats—the most prominent organizations of moderate Democrats—and the Boll Weevils, a direct predecessor to the Blue Dogs comprised of conservative southern Democrats who earned their name by providing crucial support for several of President Ronald Reagans major policy initiatives in the 1980s. Appearing further down the list (ranking at 39th overall) is a PAC founded to support the Tuesday Group, a Republican counterpart of the Blue Dogs caucus.

## 8 Contributor Estimates

One might also be interested in estimating scores for individual donors. Neither of the supervised models produce directly interpretable estimates of contributor ideal points. However it is relatively straightforward to project contributors onto the same ideological dimension as candidates. This can be done using an intuitive technique developed by McCarty, Poole, and Rosenthal (2006) to recover ideal point estimates for contributors based on the dollar weighted-average of the DW-NOMINATE scores of recipient legislators.

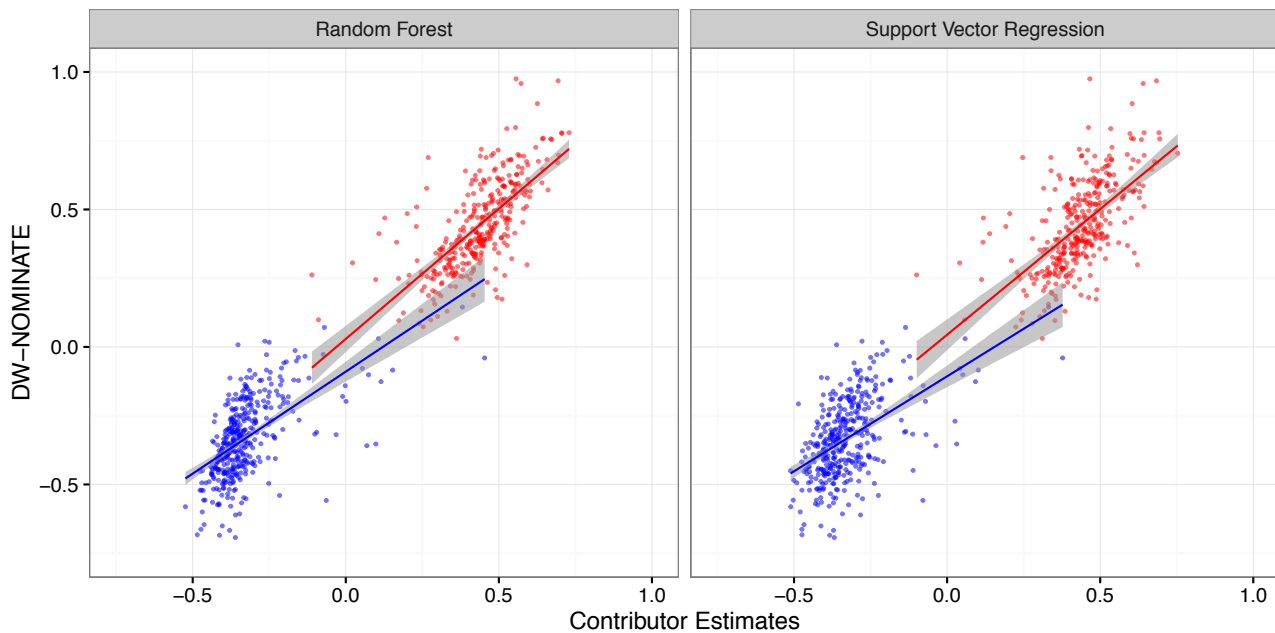
The contributor scores presented here are based on a slightly modified version of this technique. Rather than calculate the weighted averages based on DW-NOMINATE, the cross-validated estimates from the random forest model are instead used. Incorporating the predicted scores for non-congressional actors from the supervised models greatly increases the number of candidates that can be referenced in locating donors. This in turn greatly increases the number of donors for which scores can be estimated. The score for donor  $i$  is calculated as,

$$\theta_i = \frac{\sum_j \delta_j w_{ij}}{\sum_j w_{ij}}. \quad (6)$$

where  $\delta$  is a vector of recipient ideal point estimates and  $w_i$  is a vector of contribution amounts. Left unadjusted, the weighted means will have the effect of shrinking the contributor scores towards the center of the space. I take advantage of a distinctive characteristic of contribution data to adjust for shrinkage. A large percentage of candidates appear in the data both as individual donors and as recipients and thus simultaneously enter in the data as row and column observations. This makes it possible to identify contributor scores with respect to candidate scores (Bonica, 2014).

Figure 5 plots the relationship between the projected donor scores from the supervised models and DW-NOMINATE scores for candidates. Only candidates that have personally donated to five or more distinct candidates are included in the analysis. Both sets of estimates strongly correlate with DW-NOMINATE at  $r = 0.95$ . The within-party correlations are above  $r = 0.50$  for Democrats and above  $r = 0.60$  for Republicans.

A candidate's personal donations, fundraising, and voting behavior all provide consistent signals about her ideological location. Estimates recovered from personal political donations are only marginally less predictive of congressional voting behavior than scores derived from candidate fundraising.



**Figure 5:** Contributor estimates against DW-NOMINATE scores for members of Congress.

## 9 Benchmarking Unsupervised Ideal Point Measures

The results in the previous section speak to a recent debate about whether donor-based measures accurately measure individual-level ideology (Barber, Canes-Wrone, and Thrower, 2015; Hill and Huber, 2015). The results are consistent with Barber, Canes-Wrone, and Thrower (2015) who find that donation behavior is ideologically conditioned even among co-partisans. At least for the sample of candidates, there is strong evidence that individual donors can discriminate the ideology of members of the same party. Whether this generalizes beyond political candidates to the donor population at large remains to be seen, especially for one-off donors. However, there is some evidence that political donors behave more like candidates and other political elites than does the typical voter. For example, Barber and Pope (2016) find that a single dimension explains a much higher proportion of variance in the preferences of CCES respondents who self-reported as donors than for those who did not.

At the same time, the results here are inconsistent with the conclusion drawn by Hill and Huber (2015) that political donations fail to discriminate between members of the same party. They base their claim on a set of comparisons using the CFscores for survey respondents that have been matched against the DIME data. The CFscores for respondents are compared with a corresponding set of ideal point measures that were estimated by applying factor analysis to

responses to nine policy items from the CCES.<sup>12</sup> As discussed by the authors, there are several factors specific to the analysis that likely contributed to the weaker within-party correlations.<sup>13</sup>

For many of the reasons discussed above, relying strictly on comparisons between unsupervised models applied to data generated in different contexts can lead to flawed inferences about measure validity. Drawing inferences about measure validity from comparisons between unsupervised scaling models introduces a special set of challenges. The inability of one unsupervised measure to predict another unsupervised measure can be informative, but such comparisons are limited in what they can tell us. Upon observing that two measures do not perfectly align, how do we assess which is more valid?

As the importance of party has grown, any measure that successfully distinguishes the parties will be strongly correlated with other measures of ideology. In response the literature has shifted towards evaluating measures based on their ability to explain within-party variation, usually in reference to DW-NOMINATE or some other benchmark measure chosen by the researcher (e.g. Tausanovitch and Warshaw (2016)). Implicit in this approach is the assumption that the chosen benchmark measure successfully captures within-party variation in policy preferences. In doing so, it is often taken for granted that a benchmark measure improves fit over a partisan model.

Establishing that a measure represents a meaningful improvement over party requires a way to define and operationalize what it means to outperform party. I propose two complementary strategies for doing so. The first is to compare the relative contributions of between group and within-group variation with respect to party. The second is a bounding strategy. This involves constructing lower and upper bounds for how well a single dimension predicts outcomes of interest (i.e. roll call vote choices). A partisan model represents a natural choice for the lower bound, which serves as a baseline. The upper bound can then be determined by fitting an unsupervised IRT or factor analytic model on the target dataset. This will identify the latent dimension that best explains variation in the data. This provides an intuitive way to gauge model

---

<sup>12</sup>The reported within-party correlations are .10 for Democrats and .49 for Republicans. The overall correlation is not reported.

<sup>13</sup>The contributor scores used in the paper are recalculated based only on donations made during the 2011-2012 election cycle. The majority of estimates were based on a single donation, often to presidential candidates. Partly as a result, the estimates for co-partisans exhibited less heterogeneity than is observed the raw DIME scores for contributors. Because the DIME scores were recalculated based solely on donations made during a single presidential election cycle, the matched sample clusters around the ideal points of their respective party's presidential nominee. This effect is especially severe for Democratic donors. Moreover, a small amount of random noise was added to the DIME scores for matched donors to protect anonymity of respondents which likely introduced additional attenuation bias.

performance relative to a partisan baseline while taking into account that the total variation that can be explained with a single dimension is limited.

To illustrate, I downloaded the 2012 CCES data and replicated the factor model described in Hill and Huber (2015).<sup>14</sup> An initial examination suggests the measure does poorly in explaining within-group variation among co-partisans. When the Hill and Huber factor analysis model is run separately for respondents from each party, it explains a mere 4.5 and 6.6 percent of total variation for Democrats and Republicans, respectively. The survey-based measures do not fare much better when compared with self-reported ideology. For neither party does the factor model explain more than 20 percent of the variance in self-reported ideology. When restricting the sample to the middle 80 percent of respondents from each party (between the 10th to 90th percentiles), the measure fails to explain more than 8 percent of the variance in self-reported ideology within either party.

The bounding strategy for benchmarking unsupervised ideal point models outlined above evaluates measures based on their predictive power relative to a baseline partisan model. When possible, it is best to assess predictive accuracy with respect to outcomes that were generated independently of the data used to construct the measures. One way the CCES factor model can be compared against a partisan model is to assess its accuracy in predicting responses for other policy-based survey items not included in the factor analysis. For this task, I make use of a battery of 10 CCES items that correspond with roll call votes taken in Congress. The responses to these questions are binary outcomes indicating support or opposition for a stated policy. The first step is to estimate a roll call model using the items as a way to determine the upper bound for predicting responses with a single dimension. As before, the partisan model represents the lower bound.

	Donors	Full Sample
IRT Model	0.813 (0.478)	0.771 (0.294)
Hill & Huber Factor Model	0.781 (0.387)	0.719 (0.215)
Partisan Model	0.776 (0.378)	0.715 (0.206)

**Table 5:** Predicting CCES Roll Call Items For Self-Reported Donors Using Party and Survey-based Measures of Ideology (Percentage of Votes Correctly Classified).

Table 5 reports the correct classification and APRE statistics for the CCES roll call items

<sup>14</sup>I thank Seth Hill for providing the replication code.



across four measures. For purposes of comparison, the first column reports results from a restricted sample of CCES respondents limited to self-reported donors that self-identify with either of the major parties. The second column reports results for the entire sample. For the restricted sample, the IRT model increases correct classification over the partisan model by 3.8 percentage points, representing the distance between the lower and upper bounds. The factor model improves correct classification over a partisan model by a modest 0.4 percentage points. This tells us that the factor model accounts for about 12 percent of the difference between lower and upper bounds. To compare, the results presented in Table 2 show that most measures account for well over half of the increase in classification between the partisan model and DW-NOMINATE. The numbers are no more reassuring when the full sample of respondents is included. The IRT model improves classification over a partisan model by 5.6 percentage points, widening the distance between the lower and upper bounds. However, the factor model again improves classification by just 0.4 percentage points, accounting for only 7 percent of the distance between bounds.

Table 5 calls into question the conclusions drawn by Huber and Hill regarding the reliability of donor-based measures of ideology. Given that the Hill and Huber factor model explains very little in the way of within-party variation in a related dataset of stated policy preferences, it offers a poor benchmark for validation. This highlights the challenges inherent in drawing inferences about measure validity from comparisons between unsupervised scaling models. At the same time, it speaks to the need for a more grounded approach to comparing and benchmarking ideal point measures along the lines of those proposed here.

## **10 Conclusion**

Spurred by recent efforts to extend ideal point estimation beyond the confines of legislatures, researchers have developed new and innovative ways to measure preferences from increasingly diverse sources of data. Supervised learning methods stand out as a powerful tool for prediction and measurement. The results presented here demonstrate supervised learning to be a highly effective strategy for forecasting future voting behavior based on fundraising activity prior to entering office.

One limitation of the supervised learning approach is that it reveals relatively little about the details of the mapping process. The feature analysis performed in Section 7 provides some initial insight into the process. Future work may take advantage of supervised learning methods

such as kernel regularized least squares (Hainmueller and Hazlett, 2014) that allow for improved interpretability of the target function than the more standard machine learning methods used here.

Ideology is a fluid construct, used to describe a seemingly self-organizing system of preferences that evolves over time and adapts to different institutions and political behaviors. In that sense, ideal point estimates recovered from a specific institution of political behavior should be understood as contextual. Researchers will have to decide whether to define ideology as being linked to a specific institution or more broadly as the dimension that explains the entirety of political preferences at the societal level. If ideology is defined more narrowly in the context of a single institution, it suggests a supervised learning approach similar to the one used here is most appropriate and that ideal point measures should be judged with respect to their success in predicting outcomes in the specified institution. If ideology is instead defined more broadly, ideal point measures should be viewed as complementary rather than as competing models and suggests value in combining information across measures.

## Bibliography

- Ahler, Douglas J., Jack Citrin, and Gabriel S. Lenz. Forthcoming. "Do Open Primaries Improve Representation? An Experimental Test of California's 2012 Top-Two Primary." *Legislative Studies Quarterly* .
- Ansolabehere, Stephen, Jr. Snyder, James M., and III Stewart, Charles. 2001. "Candidate Positioning in U.S. House Elections." *American Journal of Political Science* 45 (1): pp. 136–59.
- Bailey, Michael A. 2013. "Is today's court the most conservative in sixty years? challenges and opportunities in measuring judicial preferences." *The Journal of Politics* 75 (03): 821–834.
- Barber, Michael J, Brandice Canes-Wrone, and Sharece Thrower. 2015. "Sophisticated Donors: Which Candidates Do Individual Contributors Finance?" Unpublished manuscript. Available for download at <http://scholar.princeton.edu/bcwrone/publications/sophisticated-Donors-Which-Candidates-Do-Individual-Contributors-Finance>.
- Barber, Michael, and Jeremy C Pope. 2016. "Lost in Issue Space? Measuring Levels of Ideology in the American Public." Unpublished manuscript. Available for download at [https://apw.polisci.wisc.edu/APW\\_Papers/scalingvoters\\_1.pdf](https://apw.polisci.wisc.edu/APW_Papers/scalingvoters_1.pdf).
- Barberá, Pablo. 2015. "Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data." *Political Analysis* 23 (1): 76–91.
- Barberá, Pablo, John T Jost, Jonathan Nagler, Joshua A Tucker, and Richard Bonneau. 2015. "Tweeting From Left to Right Is Online Political Communication More Than an Echo Chamber?" *Psychological Science* 26 (10): 1531–1542.
- Beauchamp, Nick. 2012. "Using text to scale legislatures with uninformative voting." New York University Mimeo. Available for download at [http://nickbeauchamp.com/work/Beauchamp\\_scaling\\_current.pdf](http://nickbeauchamp.com/work/Beauchamp_scaling_current.pdf).
- Bond, Robert, and Solomon Messing. 2015. "Quantifying Social Media's Political Space: Estimating Ideology from Publicly Revealed Preferences on Facebook." *American Political Science Review* 109 (01): 62–78.
- 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): 367–387.
- Bonica, Adam. 2016. "A Data-Driven Voter Guide for U.S. Elections: Adapting quantitative measures of the preferences and priorities of political elites to help voters learn about candidates." Forthcoming. *RSF Journal on Big Data and the Social Sciences*.
- Breiman, Leo. 2001. "Random forests." *Machine learning* 45 (1): 5–32.
- Cahoon, Lawrence S, Melvin J Hinich, and Peter C Ordeshook. 1976. "A Multidimensional Statistical Procedure for Spatial Analysis." Manuscript. Carnegie-Mellon University.
- Cantú, Francisco, and Sebastián M Saiegh. 2011. "Fraudulent democracy? An analysis of Argentina's infamous decade using Supervised Machine Learning." *Political Analysis* 19 (4): 409–433.
- Carpenter, Dick, and Jeffrey Milyo. 2012. "Public's Right to Know versus Compelled Speech: What Does Social Science Research Tell us about the Benefits and Costs of Campaign Finance Disclosure in Non-Candidate Elections, The." *Fordham Urb. LJ* 40: 603.
- Clinton, Joshua D., Simon Jackman, and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *The American Political Science Review* 98 (2): 355–70.
- Converse, Philip E. 1964. "The Nature of Belief Systems in Mass Publics. In *Ideology and Discontent*, ed. David Apter. New York: Free Press."
- Cox, Gary W., and Mathew D. McCubbins. 2006. *Setting the Agenda*. Cambridge University Press.
- Diermeier, Daniel, Jean-François Godbout, Bei Yu, and Stefan Kaufmann. 2012. "Language and ideology in Congress." *British Journal of Political Science* 42 (01): 31–55.
- Drucker, Harris, Chris JC Burges, Linda Kaufman, Alex Smola, and Vladimir Vapnik. 1997. "Support vector regression machines." *Advances in neural information processing systems* 9: 155–161.
- Enelow, James M, and Melvin J Hinich. 1984. *The spatial theory of voting: An introduction*. CUP Archive.

- Fan, Rong-En, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. "LIBLINEAR: A Library for Large Linear Classification." *Journal of Machine Learning Research* 9: 1871–1874.
- Gerrish, Sean, and David M Blei. 2012. How they vote: Issue-adjusted models of legislative behavior. In *Advances in Neural Information Processing Systems*. pp. 2753–2761.
- Grimmer, Justin, and Brandon M Stewart. 2013. "Text as data: The promise and pitfalls of automatic content analysis methods for political texts." *Political Analysis* .
- Groseclose, Tim, Steven D. Levitt, and Jr. Snyder, James M. 1999. "Comparing Interest Group Scores across Time and Chambers: Adjusted ADA Scores for the U.S. Congress." *American Political Science Review* 93 (1): 33–50.
- Hainmueller, Jens, and Chad Hazlett. 2014. "Kernel Regularized Least Squares: Reducing Misspecification Bias with a Flexible and Interpretable Machine Learning Approach." *Political Analysis* 22 (2): 143–168.
- Hall, Andrew B. 2015. "What Happens When Extremists Win Primaries?" *American Political Science Review* 109 (01): 18–42.
- Hare, Christopher, David A Armstrong, Ryan Bakker, Royce Carroll, and Keith T Poole. 2015. "Using Bayesian Aldrich-McKelvey Scaling to Study Citizens' Ideological Preferences and Perceptions." *American Journal of Political Science* 59 (3): 759–774.
- Hill, Seth J., and Gregory A. Huber. 2015. "Understanding the Representativeness and Motivations of Contemporary Contributors to Political Campaigns: Results from Merged Survey and Administrative Records." Yale University Working Paper. Available for download at [http://www.sethjhill.com/HillHuberDonorate\\_062515.pdf](http://www.sethjhill.com/HillHuberDonorate_062515.pdf).
- Hinich, Melvin J, and Michael C Munger. 1996. *Ideology and the theory of political choice*. University of Michigan Press.
- Kuhn, Max. 2008. "Building predictive models in R using the caret package." *Journal of Statistical Software* 28 (5): 1–26.
- Lauderdale, Benjamin E, and Alexander Herzog. 2015. "Measuring political positions from legislative speech." Unpublished manuscript. Available for download at <http://benjaminlauderdale.net/downloads/WordshoalPaper.pdf>.

- Lauderdale, Benjamin, and Tom S. Clark. 2014. "Scaling Politically Meaningful Dimensions Using Texts and Votes." *American Journal of Political Science* 58 (3): 754–771.
- Laver, Michael, Kenneth Benoit, and John Garry. 2003. "Extracting Policy Positions from Political Texts Using Words as Data." *American Political Science Review* 92 (2): 311–32.
- Maestas, Cherie D, Matthew K Buttice, and Walter J Stone. 2014. "Extracting Wisdom from Experts and Small Crowds: Strategies for Improving Informant-based Measures of Political Concepts." *Political Analysis* 22 (3): 354–373.
- McCarty, Nolan M., Keith T. Poole, and Howard Rosenthal. 2006. *Polarized America: The Dance of Ideology and Unequal Riches*. Cambridge, Mass.: MIT Press.
- Nokken, Timothy P., and Keith T. Poole. 2004. "Congressional Party Defection in American History." *Legislative Studies Quarterly* 29: 545–68.
- Poole, Keith T. 2000. "Nonparametric Unfolding of Binary Choice Data." *Political Analysis* 8 (3): 211–37.
- Poole, Keith T. 2005. *Spatial Models of Parliamentary Voting*. Analytical Methods for Social Research New York: Cambridge University Press.
- Poole, Keith T., and Howard Rosenthal. 2007. *Ideology and Congress*. 2nd rev. ed. New Brunswick: Transaction Publishers.
- Primo, David M. 2013. "Information at the Margin: Campaign Finance Disclosure Laws, Ballot Issues, and Voter Knowledge." *Election Law Journal* 12 (2): 114–129.
- Ramey, Adam. 2016. "Vox Populi, Vox Dei? Crowdsourced Ideal Point Estimation." *The Journal of Politics* 78 (1): 281—295.
- Roberts, Jason M., and Steven S. Smith. 2003. "Procedural Contexts, Party Strategy, and Conditional Party Voting in the U.S. House of Representatives, 1971-2000." *American Journal of Political Science* 47 (2): 305–317.
- Shor, Boris, Christopher Berry, and Nolan McCarty. 2010. "A Bridge to Somewhere: Mapping State and Congressional Ideology on a Cross-Institutional Common Space." *Legislative Studies Quarterly* 35: 417–48.

- Shor, Boris, and Nolan M. McCarty. 2011. "The Ideological Mapping of American Legislatures." *American Political Science Review* 105 (03): 530–51.
- Smola, Alex J, and Bernhard Schölkopf. 2004. "A tutorial on support vector regression." *Statistics and computing* 14 (3): 199–222.
- Tausanovitch, Chris, and Christopher Warshaw. 2016. "Estimating Candidate Positions in a Polarized Congress." MIT Working Paper. Available for download at <http://cwarshaw.scripts.mit.edu/papers/CandidatePositions160213.pdf>.
- Windett, Jason H, Jeffrey J Harden, and Matthew EK Hall. 2015. "Estimating Dynamic Ideal Points for State Supreme Courts." *Political Analysis* 23 (3): 461–469.
- Yu, Bei, Stefan Kaufmann, and Daniel Diermeier. 2008. "Classifying Party Affiliation from Political Speech." *Journal of Information Technology & Politics* 5 (1): 33–48.