

# Inferring Roll-Call Scores from Campaign Contributions Using Supervised Machine Learning

**Adam Bonica** Stanford University

**Abstract:** *This article develops a generalized supervised learning methodology for inferring roll-call scores from campaign contribution data. Rather than use unsupervised methods to recover a latent dimension that best explains patterns in giving, donation patterns are instead mapped onto a target measure of legislative voting behavior. Supervised models 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, forecasts based on fundraising as a nonincumbent predict future voting behavior as accurately as in-sample forecasts based on votes cast during a legislator's first 2 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 data successfully distinguish between members of the same party.*

**Replication Materials:** The data, code, and any additional materials required to replicate all analyses in this article are available on the *American Journal of Political Science* Dataverse within the Harvard Dataverse Network, at: <https://doi.org/10.7910/DVN/7EWDRF>.

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 (Clinton, Jackman, and Rivers 2004; Poole and Rosenthal 2007). Much of the appeal of these measures lies in 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 the current understanding of Congress as fundamentally unidimensional. This has in turn aided in testing a variety of theories about representation, accountability, and legislative behavior and has fostered their widespread adoption.

A well-known limitation of roll-call-based measures of ideology is that they are confined to voting bodies. This precludes estimating scores for nonincumbent candidates prior to taking office, which is arguably where such predictions would be most valuable (Tausanovitch and Warshaw 2017). Only 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 ever more varied types of data, including voter evaluations of candidates (Hare et al. 2015; Maestas, Buttice, and Stone 2014), legislative speech (Beauchamp 2012; Lauderdale and Herzog 2016), social media follower networks (Barberá 2015; Bond and Messing 2015), and campaign contributions (Bonica 2013, 2014; Hall 2015).

As the most widely used measure in the literature, DW-NOMINATE is widely viewed as a benchmark measure of ideology. Although comparisons with an established measure can be useful for demonstrating face validity, this practice 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>1</sup> The output of roll-call scaling models is most accurately understood as a relative ordering of individuals along a *predictive* dimension that best explains voting behavior in a given voting body. Although widely understood as measures of ideology, this is an

Adam Bonica is Assistant Professor, Department of Political Science, Stanford University, Encina Hall West, Office Number 307, 616 Serra Street, Stanford, CA 94305 (bonica@stanford.edu).

<sup>1</sup>Partial exceptions include Gerrish and Blei (2012), Lauderdale and Clark (2014), and Bonica (2016a), which use semi-supervised methods to identify the dimensionality of roll calls based on issue weights from topic models.

*American Journal of Political Science*, Vol. 00, No. 00, xxxx 2018, Pp. 1–19

©2018, Midwest Political Science Association

DOI: 10.1111/ajps.12376

interpretation given by the researcher and not reflective of any defined objective built into the model.

In a recent article, Tausanovitch and Warshaw (2017) evaluate several alternative measures of ideology recovered from survey data, campaign contributions, and social media data 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 to question the usefulness of these measures for testing theories of representation and legislative behavior or for predicting how nonincumbent candidates would behave in office.

In addition to the obvious methodological implications for researchers, this also 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 (2016) cast doubt on the ability of voters to discern ideological differences between 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 (Carpenter and Milyo 2012; Primo 2013). 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 serve to further undermine an important policy rationale for campaign finance disclosure laws.

This article 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 primary contribution of this approach is the ability to generate the equivalent of roll-call scores for actors

who lack voting records. This is accomplished with a high degree of accuracy and with minimal identifying assumptions. This approach, however, is not limited to roll-call scores. It is readily extended to measures that capture preferences on specific issues (i.e., abortion) and other target dimensions of interest. In doing so, the article contributes to the literature on bridging methods for recovering comparable preference estimates across institutions and types of actors.

## 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, 207) 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>2</sup> According to the theory, 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. Given a shared understanding of how issues map onto 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. Put slightly differently, it is a shared method of systematically simplifying politics with the “knowledge of what goes with what” (Poole 2005, 12).

<sup>2</sup>As explained by Poole (2005, 13), “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, that she is a *conservative* and almost certainly a Republican.”

In recent years, a trend has emerged toward 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 or aptitude that in turn generate responses to test questions. What distinguishes the most intelligent individuals is an enhanced cognitive ability that allows them to 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  $n \times k$  matrix of issue positions of  $n$  individuals on  $k$  issue dimensions and  $\mathbf{X}$  be the  $n \times s$  matrix of individuals' ideal points on the  $s$  ideological dimensions. The presence of issue constraint implies that all issue positions can be represented as  $\mathbf{X}\boldsymbol{\beta} \Rightarrow \mathbf{Y}$ , where  $\boldsymbol{\beta}$  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 with which issues are voted on should matter to how ideal points project onto issue dimensions, which strictly depends on  $\mathbf{X}\boldsymbol{\beta}$ . 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 are reasons to doubt such an interpretation. The crux of the problem is that the sources of constraint remain 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,  $\mathbf{Y} \Rightarrow \mathbf{X}\boldsymbol{\beta}$ . The objective of these models 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. 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,  $\mathbf{X}$  will adjust accordingly. Simply put, an issue that is voted on a hundred times will have a greater influence on the dimension recovered from a scaling model than an issue that is only voted on once, or not at all.

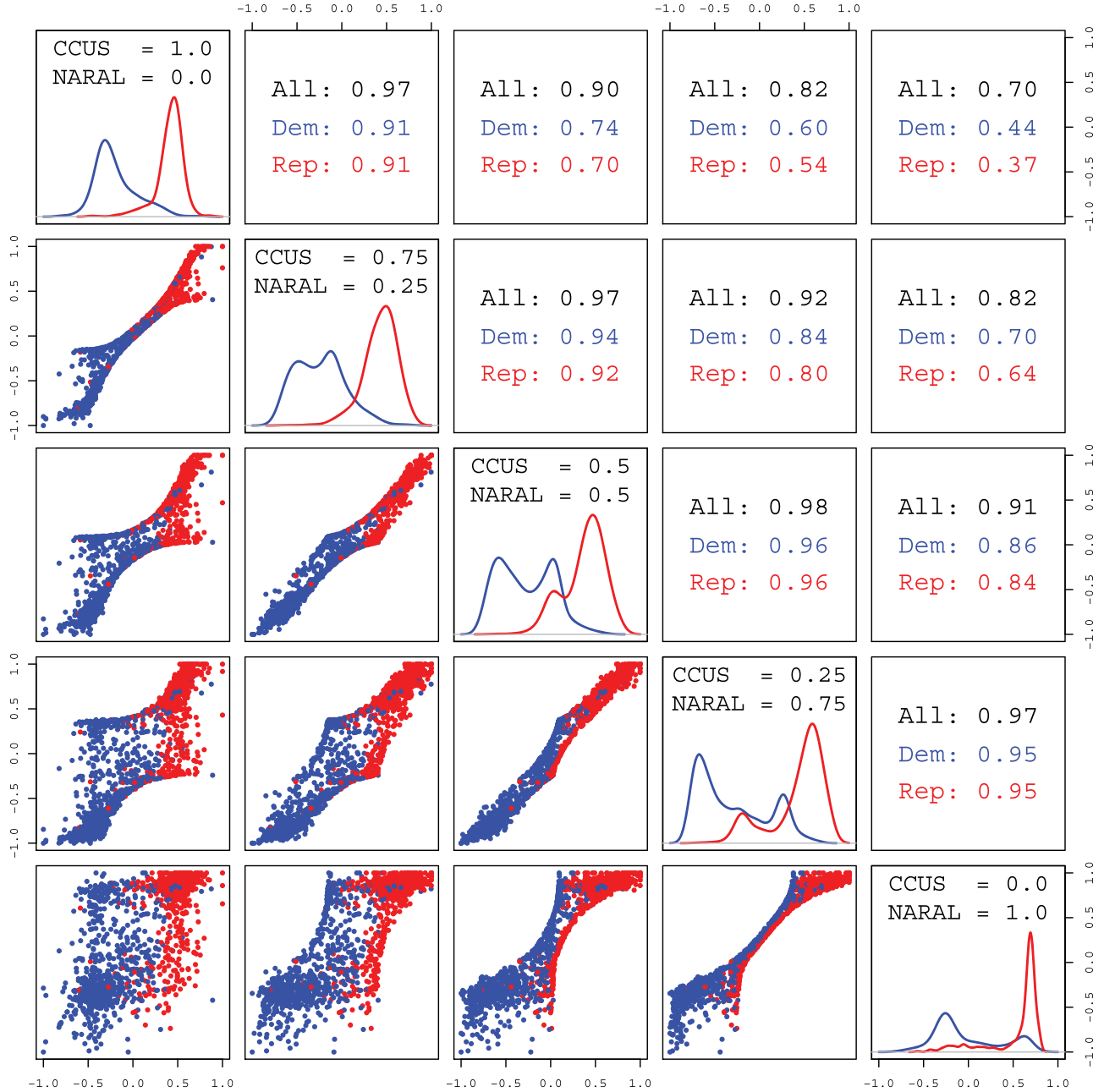
## Implications for Validation and Prediction

The apparent disconnect between ideological measurement in theory and in practice has led to some confusion about how to interpret and validate alternative measures of ideology. In practice, the output of scaling models is the dimension that best explains variation in the patterns of behavior observed in the data. In this sense, these models are primarily descriptive in nature as opposed to explicit measures of a well-defined target concept. This does not necessarily imply these measures should be viewed as less valid or useful—to the contrary, they often reveal something very meaningful about how political preferences are structured in a given context and can be remarkably powerful tools for describing and summarizing the data at hand. But this can make benchmark comparisons between different measures of ideology difficult to interpret if neither the mapping function nor the issue weights are observed. As a result, it is difficult to determine whether discrepancies between measures result from measurement error or systematic differences in how issues are mapped onto the latent dimensions.

To illustrate why issue salience matters when interpreting ideal point estimates, consider a simplified issue space composed of two issue dimensions, one relating to economic policy and the other to social issues. To represent legislator preferences on each dimension in a realistic manner, I make use of adjusted interest group ratings. Ideal points on the economic and social dimensions are based on legislator ratings issued by the U.S. Chamber of Congress (CCUS) and the National Abortion and Reproductive Rights League (NARAL), respectively. The ratings, which are derived from voting patterns, have been adjusted using a technique developed by Groseclose, Levitt, and Snyder (1999) that places legislators across legislative periods on a single scale to allow for over-time comparisons.<sup>3</sup> For each issue dimension, I generate a matrix of simulated vote choices drawn conditional on the adjusted interest group scores. This yields two distinct sets

<sup>3</sup>The adjusted interest group scores cover legislators who served between 1979 and 2008 and were downloaded from <http://timgroseclose.com/adjusted-interest-group-scores/>. Legislators serving in more than one congress are assigned a single score based on their mean adjusted scores.

**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.

of simulated roll-call votes, each capturing preferences on a distinct issue domain.

By combining roll-call votes simulated from each issue dimension, I generate hypothetical scenarios by varying the proportion of bills drawn from each issue dimension. Given a weighting profile where all 1,000 bills are drawn from a single issue dimension, scaling the

resulting roll-call matrix will recover the set of adjusted interest group ratings for that dimension. Scaling weighted profiles of bills drawn proportionately from both issue dimension provides a sense of how the recovered ideal points respond to changes in issue salience.

Figure 1 compares ideal points recovered from scaling an  $n \times 1,000$  vote matrix corresponding to one of

five hypothetical weighting profiles. A one-dimensional NOMINATE model is used to perform the scaling (Poole et al. 2011). The two corner scenarios assume that 100% of the bills are related to a single issue dimension. In two other scenarios, 75% of the bills are drawn from one issue dimension, whereas 25% are drawn from the other. A fifth scenario weights each issue dimension equally. Comparing ideal points recovered under different weighting profiles illustrates just how sensitive unsupervised scaling models can be to issue salience. Of particular interest are the resulting distributions of ideal points grouped by party shown along the diagonal. Depending on the issue weights, the distributions of ideal points on the latent dimension can look very different.<sup>4</sup>

### Implications for Bridging Applications

Bridging applications are one example of where failing to account for issue salience can be highly problematic. A common identification strategy uses legislators who served in multiple legislatures as bridge observations—for example, state legislators who later serve in Congress. Linear projections are used to rescale ideal points recovered from voting patterns in state legislatures to the same actors' ideal points recovered from voting patterns 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 voting in one legislature is identical to the dimension that best explains roll-call voting in the other and that, after rescaling, any differences in ideal points recovered from each voting body are simply a matter of measurement error. If issues are weighted differently across voting bodies, the shared dimensionality assumption will be violated.

It is doubtful that the shared dimensionality assumption holds 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 that ultimately make it to the floor are the products of a highly strategic and closely managed agenda-setting process (e.g., see Cox and McCubbins 2006). Moreover, the types of issues considered by a legislature are largely contingent on its jurisdiction.

<sup>4</sup>Scaling the model allowing for two dimensions does little to address the problem. The inclusion of a second dimension results in an appreciable increase in correct classification, but the first dimension recovered remains nearly identical to that recovered from a one-dimensional model. See the supporting information for results from the first dimension of a two-dimensional model.

Many issues that are central to state legislatures, such as education policy, receive much less attention in Congress. On the flip side, issues related to defense, foreign policy, and trade are almost strictly the domain of Congress.

This problem is complicated even further for 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. Political action committees (PACs) and ballot committees that focus on social issues consistently draw large numbers of donors. The likely consequence of this is that social issues will receive more weight when scaling contributions and less weight when scaling congressional roll calls.

One way researchers have made progress on the comparability problem is by using communal data sets such as the National Political Awareness Test (NPAT) candidate surveys as intermediaries between voting bodies (Shor and McCarty 2011). In this arrangement, state legislators and members of Congress are first jointly scaled using their NPAT responses. Congress and state legislatures are then scaled independently using roll-call data and projected onto the NPAT common space with an error-in-variables regression model. While this greatly increases the number of available bridge observations and streamlines the process with the introduction of a common data source, identification still fundamentally rests on the shared dimensionality assumption. For this assumption to be satisfied, the unsupervised scaling models applied to state legislatures must all recover preferences along a shared predictive dimension. This seemingly heroic assumption is only made at all plausible in light of the intensely unidimensional nature of contemporary American politics. But such an assumption is made much more tenable and transparent by a supervised approach.

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.

### Supervised Learning Algorithms for Predicting Congressional Voting

This section outlines the methodology for inferring DW-NOMINATE scores for candidates based on alternative 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 incumbents and nonincumbents. Nearly all candidates engage in fundraising, making contribution data ideal for this purpose. 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 (Grimmer and Stewart 2013; Hainmueller and Hazlett 2014). By far, the most common application of 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 offers an efficient means of topic coding large corpora 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 is associated with a legislator (Diermeier et al. 2012; Yu, Kaufmann, and Diermeier 2008). Similar techniques have been used to measure the personality traits of legislators from their speech (Ramey, Klingler, and Hollibaugh 2016).

The supervised machine learning task undertaken here can be thought of in a similar vein. The candidate-contributor matrix takes on a structure nearly identical 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 who 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 is 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 scores are unobserved. 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 represent 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  $f(\cdot)$ . Then  $\hat{f}(\cdot)$  is 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 is attractive for its computational efficiency and speed, whereas random forests tend to be less susceptible to overfitting and perform best in cross-validated tests.

## 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 of 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,

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

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

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

where  $k(\cdot)$  is the kernel function and  $b$  is the “bias” term. A linear kernel  $k(x_i, x_j) = x_i^T x_j$  is used here in part due to the availability of an extremely fast implementation of the algorithm for large-scale data sets (Fan et al. 2008). The SVM algorithm solves the constrained optimization problem by introducing nonnegative optimizers  $\alpha_i$  and

$\alpha_i^*$  and constant  $C > 0$ . The dual formula for the optimization problem is

$$\begin{aligned} \min_{\alpha_i, \alpha_i^*} & \frac{1}{2} \sum_{i,j=1}^N (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) k(x_i, x_j) \\ & + \epsilon \sum_{i=1}^N (\alpha_i^* + \alpha_i) + \sum_{i=1}^N y_i (\alpha_i^* - \alpha_i), \end{aligned} \quad (5)$$

subject to

$$\sum_{i=1}^N (\alpha_i^* - \alpha_i) = 0, \quad \alpha_i, \alpha_i^* \in [0, C].$$

## Random Forests

Random forests are a widely used decision-tree-based ensemble method for supervised learning. The method operates by constructing many random decision trees from the input data and aggregating over the output to generate predictions. For each tree, a sample of observations is drawn at random, with the remaining observations held “out-of-bag.” For the initial node of each tree, a sample of  $m$  predictor variables is selected at random. The variable that best classifies the data is used to split data at that node. This process is repeated at each subsequent node to generate the tree. The model averages predictions over trees to generate estimates. The main advantages of random forests are efficiency with large data sets, resistance to overfitting, and built-in estimates of variable importance. (See Breiman 2001 for an overview.)

## 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 1st–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 are from the Database on Ideology, Money in Politics, and Elections (DIME; Bonica 2016b). The DIME data cover a period from 1980 to 2014 and contain records for 72,065 candidates from state and federal elections (1,718 of whom have DW-NOMINATE scores). Following Bonica (2014), I exclude corporate and trade PACs from the feature set due to their tendency to mix

ideological and strategic motives. In addition, indicator variables for three basic candidate traits—party, home state, and gender—are included in the feature matrix.

## Feature Selection and Extraction

Were every donor to be retained as individual predictors, the resulting matrix would contain several million features. This makes a strategy for dealing with high dimensionality. As an initial step, I filter on donors who have given to at least 15 distinct candidates included in the training set (e.g., who have DW-NOMINATE scores). This reduces the number of features to 67,844. The threshold of 15 distinct candidates reflects the trade-off between improved model performance and computational costs.

While filtering reduces sparsity in the contribution matrix, it also discards potentially useful information from millions of excluded donors. Much of the relevant information can be retained via feature extraction. Specifically, I construct an  $n \times q$  matrix that summarizes the percentage of funds a candidate raised from donors that fall within  $q = 10$  ideological quantiles. This is done by calculating contributor coordinates based on the dollar-weighted average of the DW-NOMINATE scores of recipient candidates and then binning the contributor coordinates into deciles. I then calculate the proportions of contribution dollars raised by each candidate from each decile of donors. The decile shares are constructed in conformance with the cross-validation scheme. For each fold of the data, the decile shares are recalculated using the training set as part of the cross-validation loop. Thus, the decile shares do not incorporate any information from the test set. The decile shares are calculated using a ten-fold cross-validation scheme. This results in a slightly less efficient use of the training data but ensures that donations to the candidate are not used to inform the decile shares.

## 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.

Note that DW-NOMINATE scores are treated as known quantities despite being measured with error. This

makes interpreting model fit slightly less straightforward, as it is unclear the extent to which cross-validation error actually reflects measurement error in the target variable. Measurement error in the target variable is relatively common for supervised machine learning exercises, especially those that rely on human coding to generate a training set. Although measurement error of this sort can lead to overfitting, regularized kernel-regression methods and random forests, in particular, are less prone to overfitting in the presence of low levels of measurement error.

## 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 (Ansolahehere, 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 cross-validated 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 scores based on the money-weighted average of their recipients' DW-NOMINATE scores. The process is then reversed and scores for candidates are calculated based on the money-weighted average of their contributors.<sup>5</sup> The scores reported below are taken from out-of-sample predictions.

The two alternative congressional 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 using a scaling model similar to that of DW-NOMINATE but

incorporate additional data on position taking by nonlegislative 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 2-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 a legislator's career and then as a fixed score based on a legislator's first term in Congress.<sup>6</sup> The first-term Nokken-Poole DW-NOMINATE scores are a particularly informative benchmark for assessing predictive accuracy. It tells us how well voting patterns observed during the first 2 years in Congress predicts voting behavior over the course of a legislative career.

Table 1 reports comparisons with DW-NOMINATE for the supervised methods and alternative measures of ideology. For the supervised models, cross-validated (i.e., out of sample) and in-sample fit statistics are reported separately. (For the remainder of the article, 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.<sup>7</sup>

Note that model fit is narrowly defined here in terms of strength of association with the first dimension of DW-NOMINATE. While this is an appropriate metric for assessing the performance of supervised models, it is less appropriate as a stand-alone metric for unsupervised models. Nonetheless, for the purpose of this exercise, comparisons between supervised and unsupervised models are informative.

The supervised models perform well in predicting DW-NOMINATE scores, overall and within party, with random forests performing best. As expected, supervised models predict DW-NOMINATE scores with greater accuracy than either the common-space CFscores or PAC-based IRT CFscores, which are similarly based on campaign contributions but unsupervised. The fit statistics can also be compared against other vote-based measures.

<sup>6</sup>Only first-term scores for legislators who served in multiple congresses are included.

<sup>7</sup>Each model requires that certain parameters be tuned. The optimal values were selected via a grid search over repeated cross-validation runs. Random forest regression tunes a single parameter that specifies the number of variables to be randomly sampled at each split. Its optimal value is 22,615, or a third of the total number of predictors. The optimal values for the support vector regression are  $C = 0.05$  for the penalty parameter and  $\epsilon = 0.01$  for the tolerance margin.

<sup>5</sup>This scaling model is similar to the one used by Hall (2015). There are two key differences in how the scores are estimated. First, tenfold cross-validation is used here, whereas Hall uses a leave-one-out scheme. Second, all donors giving to two or more distinct recipients are included here, whereas Hall includes only those donors who had given to 20 or more distinct House candidates.



**TABLE 1 Predicting DW-NOMINATE Scores: Fit Statistics for Supervised and Unsupervised Measures of Ideology**

	All Candidates			Democratic Candidates			Republican Candidates		
	R	RMSE	N	R	RMSE	N	R	RMSE	N
<b>Cross-Validated</b>									
Random Forest	0.98	0.09	1,763	0.86	0.08	903	0.87	0.09	860
Support Vector Regression	0.97	0.10	1,763	0.80	0.10	903	0.82	0.11	860
Supervised CFscores	0.94	0.14	1,763	0.61	0.14	903	0.75	0.13	860
<b>In-Sample</b>									
Random Forest (in-sample)	1.00	0.03	1,763	0.98	0.03	903	0.98	0.04	860
Support Vector Regression (in-sample)	0.99	0.06	1,763	0.94	0.06	903	0.94	0.06	860
<b>Roll-Call Measures</b>									
Nokken-Poole (Career)	0.98	0.09	9,522	0.92	0.06	5,065	0.85	0.11	4,457
Nokken-Poole (First Term)	0.97	0.10	1,597	0.88	0.08	824	0.86	0.11	773
Bailey Scores	0.92	0.16	12,574	0.78	0.15	6,456	0.61	0.17	6,118
Bailey Scores (Mean)	0.91	0.17	1,644	0.79	0.16	835	0.63	0.18	809
Turbo-ADA	0.90	0.17	1,439	0.68	0.18	756	0.59	0.16	683
Shor-McCarty	0.93	0.17	227	0.58	0.18	106	0.55	0.16	121
<b>Alternative Measures</b>									
Common-Space CFscores	0.91	0.17	1,763	0.44	0.19	903	0.67	0.14	860
IRT CFscores	0.89	0.18	1,299	0.58	0.19	682	0.53	0.17	617
NPAT (1996)	0.92	0.16	250	0.75	0.16	110	0.62	0.17	140

All three supervised models increase overall fit relative to Turbo-ADA and Bailey scores. This highlights that roll-call measures can be sensitive to even relatively minor adjustments to modeling assumptions, even if the underlying data being scaled are identical. Of the included roll-call measures, the Shor-McCarty scores are the only measures based on noncongressional roll-call data. They also exhibit the weakest within-party correlations with DW-NOMINATE, speaking to the challenges inherent in bridging across institutions even when we observe bridge actors engaging in the same type of behavior in both settings.

Perhaps most telling is that the supervised machine learning models perform on par with the Nokken-Poole first-term estimates in predicting DW-NOMINATE scores. This demonstrates that *it is possible to infer legislators' DW-NOMINATE scores from their contribution records just as accurately as we can from observing how they vote during their first 2 years in Congress*. This result is all the more impressive given that the first-period Nokken-Poole scores are estimated in-sample.

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

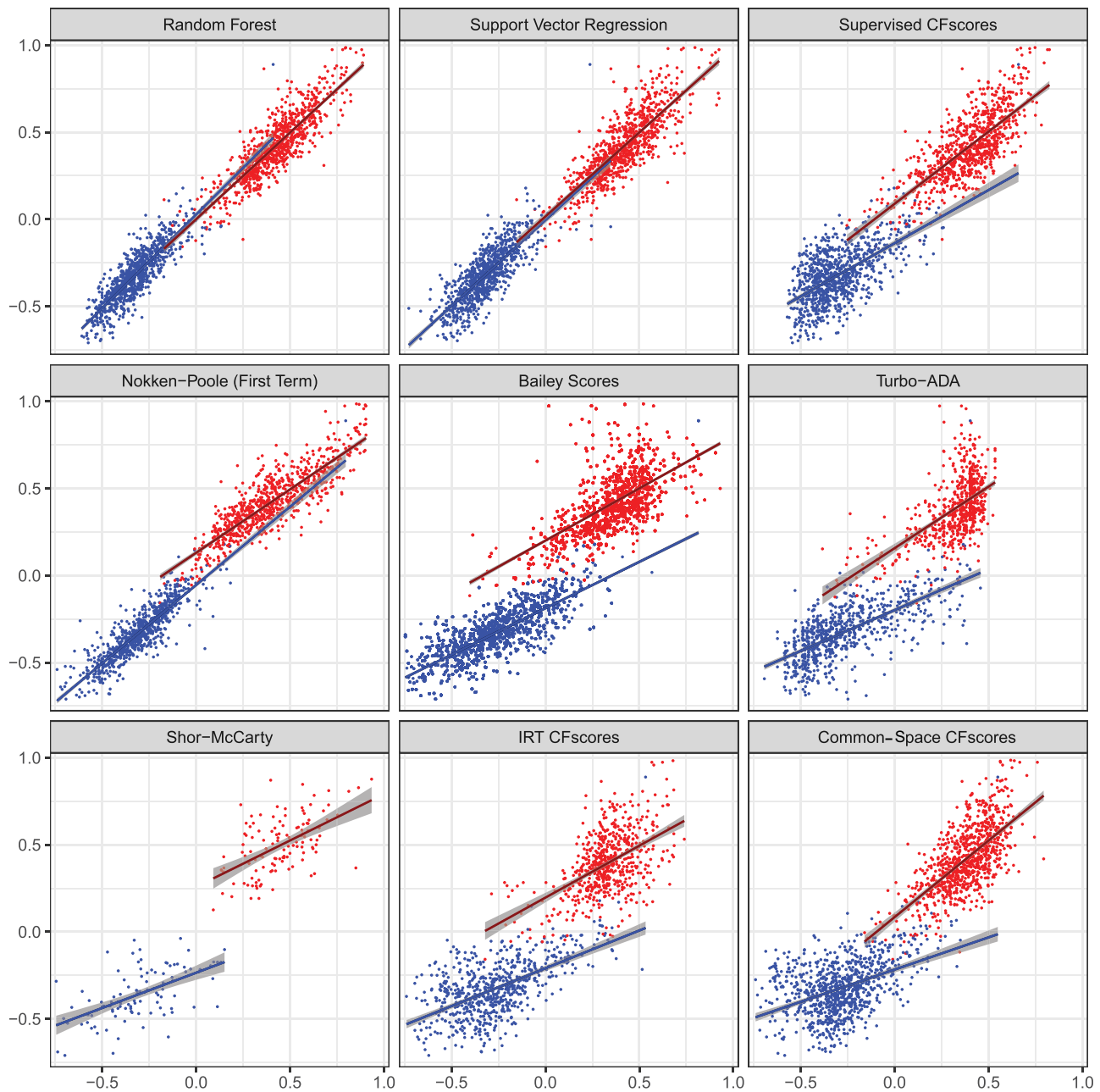
of partisan overlap.<sup>8</sup> This suggests that DW-NOMINATE may tend to overstate the extent to which the parties in Congress have polarized. In contrast, both supervised measures appear to successfully capture the gap between parties present in DW-NOMINATE, which helps to explain their higher overall correlations. Note that this is less true of the supervised CFscores, which rely on a more simplistic mapping function.

As noted above, the supervised learning approach is readily extended to other preference measures. This is demonstrated using the adjusted interest group ratings from CCUS and NARAL used in Figure 1. Results reported in the supporting information show that positions on specific issue areas can be predicted from campaign contributions with comparable accuracy.

### 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 an optimal cutting-line

<sup>8</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).

**FIGURE 2 Comparing Measures of Legislator Ideology against DW-NOMINATE Scores**

*Note:* The scales for unsupervised methods have been rescaled for purposes of comparison. Linear trend lines are fit separately for each party.

procedure (Poole and Rosenthal 2007). 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 96th–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

**TABLE 2 Percentage of Votes Correctly Classified (96th–113th Congresses)**

	House	Senate
DW-NOMINATE	0.899 (0.701)	0.887 (0.662)
Random Forest	0.895 (0.688)	0.882 (0.646)
Support Vector Regression	0.893 (0.682)	0.878 (0.635)
Nokken-Poole (First Term)	0.892 (0.680)	0.877 (0.632)
Supervised CFscores	0.884 (0.657)	0.877 (0.630)
Bailey Scores (Mean)	0.884 (0.654)	0.876 (0.629)
Common-Space CFscores	0.882 (0.650)	0.874 (0.621)
Turbo-ADA	0.873 (0.622)	0.858 (0.575)
Party	0.871 (0.616)	0.849 (0.549)

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

assigns separate ideal points based on votes cast before and after a legislator switched parties, but most other measures do not.) Following Poole and Rosenthal (2007), lopsided votes with winning margins greater than 97.5% are dropped.

The table orders measures with respect to their success in classifying roll-call outcomes, from best to worst. 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.

Figure 3 tracks correct classification (jointly 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 is consistent relative to DW-NOMINATE over the period. Also of note is that while the partisan model provides a natural baseline, 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 96th–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.

## Feature Analysis

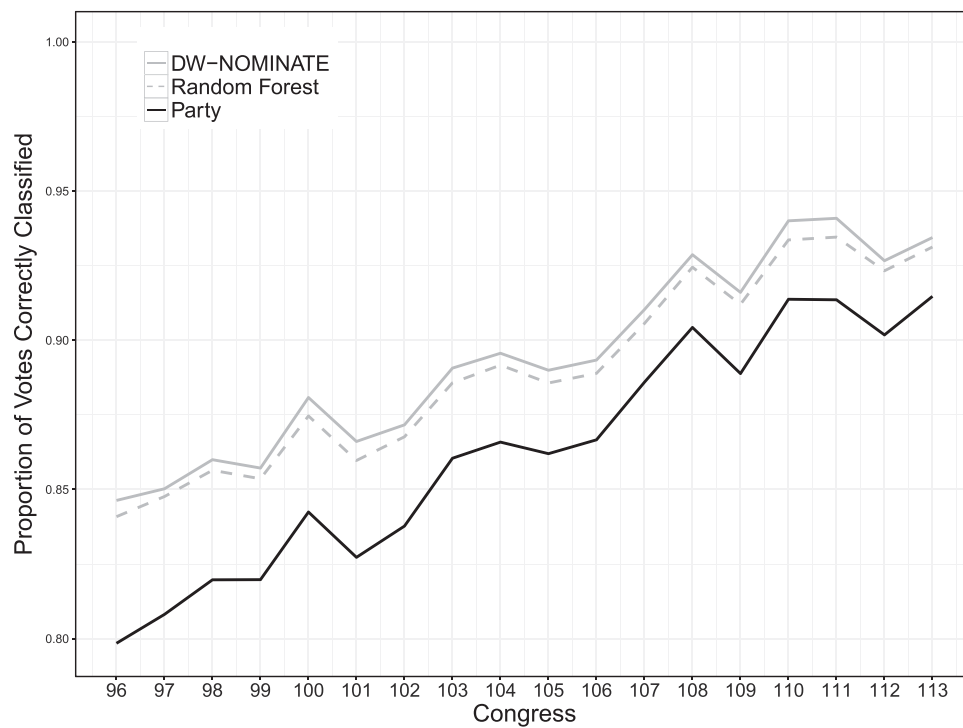
The random forest model has a built-in algorithm that ranks variables with respect to their importance to the model. Variable importance is calculated using the permutation method. It is determined by the average increase in out-of-bag (OOB) error caused by randomly permuting the values of a given variable. If permuting a variable does not increase OOB error rates, the variable is deemed to be unimportant to the model. The larger the increase in OOB error resulting from permuting a variable, the greater its importance. The variable importance scores offer insight into which types of donors matter most to the mapping function and which the model can do without.<sup>9</sup>

Table 3 lists the top 50 federal PACs ranked by variable importance.<sup>10</sup> The values reported in the first column are the percentage increase in mean squared error (MSE) over the baseline of 0.008 associated with permuting the variable. (A value of 10 would indicate that permuting the variable increases MSE from the baseline of 0.0082 to 0.00902.) Also reported are the number of distinct recipients supported by the PAC, the mean and standard deviation of their recipients' DW-NOMINATE scores, and the proportions of total contribution dollars going to Republicans and nonincumbents.

Most of the organizations listed tend to donate primarily to candidates from one or the other party. This suggests that many of the most important features contribute to the model by discriminating within party. The top feature is the Council for Citizens against Government Waste, which consistently supports the most conservative Republicans. Two other highly ranked features are the Blue Dog Democrats, a prominent organization of moderate to conservative Democrats, and the New Democrat Coalition, a congressional affiliate of the pro-business Democratic Leadership Council, which passed the Blue Dog Caucus's membership in 2013 to become the leading coalition of moderate Democrats. Not far down the list are the Republican Main Street PAC—which was set up

<sup>9</sup>A low importance score does not necessarily imply that a feature is unrelated to the target variable but rather that the information provided by the variable does not improve predictive accuracy. This may be because the variable is completely uninformative or that the information it provides is not unique.

<sup>10</sup>Note that several individual donors made it onto the list but were excluded from the table.

**FIGURE 3 Correct Classification by Congress**

to support moderate Republicans—and the Boll Weevil PAC, a direct predecessor to the Blue Dogs composed of conservative southern Democrats who earned their name by providing crucial support for several of President Ronald Reagan’s major policy initiatives in the 1980s.

### Partial Dependence

Partial dependence plots, which describe the mean marginal value associated with a variable averaged across the observed values for other variables, can be used to visualize the relationships between predictor variables and the target outcome. Partial dependence generalizes to predictive models regardless of complexity and can be interpreted in a manner similar to regression coefficients (Friedman 2001).

Figure 4 displays partial dependence plots for 20 PACs from Table 3. Consistent with the notion that the most important features tend to discriminate within party, the partial dependence plot for the Blue Dog Democrats and the New Democrat Coalition shows that the marginal effect of their contributions is in a conservative direction. As both give exclusively to Democrats, their main contribution to the model is to help identify Democrats with moderate voting records and shift them toward the center. Two Democratic leadership PACs that rank among the

top 10 most important features—Democrats Win Seats (Debbie Wasserman Schultz) and Victory Now (Chris Van Hollen)—exhibit similar patterns. The partial dependence plots for labor unions are generally much flatter. This suggests that it is an interaction with other variables that makes these features important to the model.

### Forecasting Roll-Call Measures for Nonincumbents

Much of the appeal of supervised learning is in forecasting the future voting behavior of nonincumbents based solely on data generated before entering Congress. Bonica (2014) finds that scores assigned to nonincumbents based on their fundraising prior to entering office are highly correlated with scores assigned based on fundraising after entering office. This suggests that fundraising activity before and after entering office conveys consistent information about candidate positions.

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 to candidates before and after they



**TABLE 3 Top Organizational Donors Ranked by Variable Importance**

	<b>Importance (% Increase in MSE)</b>	<b>No. Distinct Recipients</b>	<b>Avg. DW-NOM</b>	<b>Std. Dev. DW-NOM</b>	<b>Prop. to Republicans</b>	<b>Prop. to Non- incumbents</b>
Council for Citizens Against Government Waste	7.05	121	0.61	0.16	0.99	0.02
Democrats Win Seats	6.77	157	−0.23	0.13	0.00	0.24
NRCC	6.73	759	0.37	0.19	1.00	0.35
National Automotive Dealers Association	6.71	1,482	0.18	0.39	0.66	0.12
National Rifle Association	6.65	1,223	0.26	0.36	0.75	0.19
New Democrat Coalition	6.44	820	−0.23	0.38	0.10	0.23
National Federation of Independent Business	5.77	1,097	0.36	0.33	0.90	0.24
National Education Association	5.33	1,018	−0.31	0.28	0.07	0.18
Victory Now PAC	5.26	160	−0.23	0.14	0.00	0.19
Laborers' International Union of North America	4.76	980	−0.30	0.28	0.11	0.09
National Association of Letter Carriers	4.65	1,092	−0.28	0.31	0.11	0.13
American Security Council	4.63	331	0.25	0.27	0.72	0.06
Friends of Jim Clyburn	4.42	229	−0.27	0.15	0.00	0.23
DCCC	4.36	753	−0.28	0.22	0.02	0.44
Railway Clerks Political League	4.26	809	−0.32	0.26	0.07	0.08
Blue Dog Democrats	4.07	114	−0.19	0.13	0.04	0.08
American Postal Workers Union	3.96	960	−0.34	0.24	0.05	0.10
Machinists & Aerospace Workers	3.92	855	−0.36	0.21	0.02	0.14
Sierra Club	3.78	688	−0.33	0.22	0.04	0.18
American Federation of Government Employees	3.76	914	−0.32	0.26	0.06	0.11
DRIVE PAC (Teamsters)	3.68	982	−0.31	0.29	0.09	0.12
United Brotherhood of Carpenters	3.60	1,005	−0.26	0.29	0.14	0.12
Human Rights Campaign	3.46	596	−0.33	0.22	0.10	0.18
International Union of Operating Engineers	3.39	1,013	−0.26	0.29	0.14	0.08
Safari Club International	3.37	567	0.37	0.31	0.84	0.15
American Federation of Teachers	3.37	810	−0.36	0.21	0.02	0.15
National Committee for an Effective Congress	3.34	601	−0.33	0.16	0.01	0.21
American Nurses Association	3.34	957	−0.29	0.32	0.13	0.14

(Continued)

TABLE 3 Continued

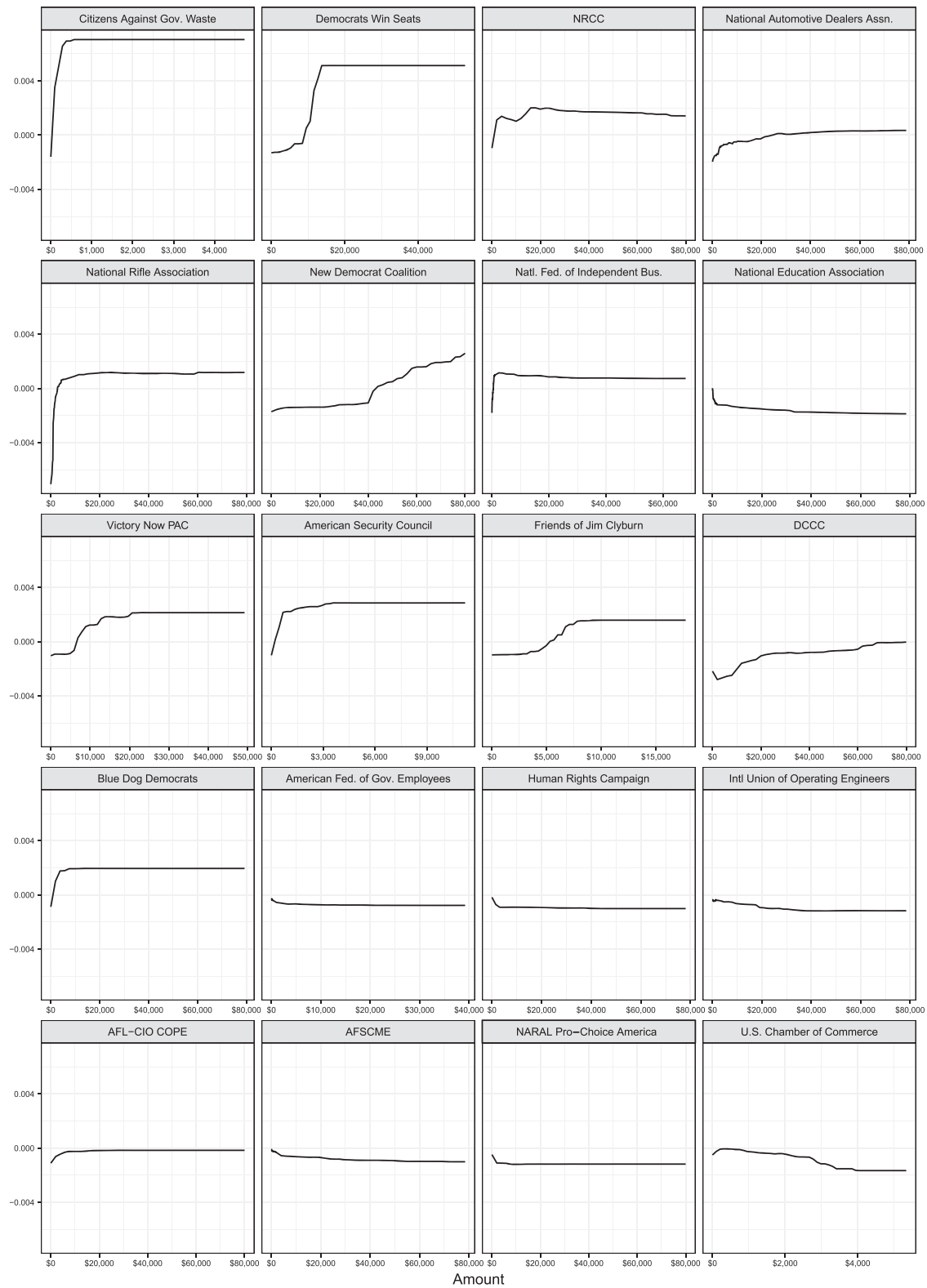
	Importance (% Increase in MSE)	No. Distinct Recipients	Avg. DW-NOM	Std. Dev. DW-NOM	Prop. to Republicans	Prop. to Non- incumbents
Consumer Federation of America	3.20	113	−0.47	0.10	0.01	0.03
International Brotherhood of Boilermakers	3.14	814	−0.31	0.25	0.06	0.09
Majority Congress Committee	2.97	374	−0.28	0.20	0.04	0.25
RNC	2.91	595	0.37	0.18	1.00	0.25
Democrats for the 80s	2.91	342	−0.30	0.17	0.01	0.21
Republican Main Street PAC	2.77	208	0.32	0.14	1.00	0.16
JustinPAC	2.64	20	0.83	0.12	1.00	0.15
Growth and Prosperity PAC	2.64	275	0.43	0.16	1.00	0.16
Northrup Grunman PAC	2.50	1,162	0.12	0.39	0.58	0.03
AFL-CIO COPE	2.49	687	−0.31	0.19	0.02	0.22
National Rural Letter Carriers	2.40	1,101	−0.17	0.37	0.25	0.08
Chevron PAC	1.99	861	0.28	0.37	0.78	0.14
NARAL Pro-Choice America	1.99	540	−0.33	0.21	0.06	0.21
Right to Work PAC	1.90	231	0.50	0.16	1.00	0.28
National Organization For Women	1.86	309	−0.38	0.18	0.05	0.33
AFSCME	1.73	990	−0.35	0.27	0.03	0.15
National Treasury Employees PAC	1.72	842	−0.31	0.26	0.07	0.11
U.S. Chamber of Commerce	1.60	566	0.32	0.23	0.96	0.24
United Technologies Corporation	1.45	1,632	−0.14	0.40	0.28	0.11
National Committee to Preserve Social Security	1.45	915	−0.26	0.31	0.13	0.12
Liberty PAC	1.14	48	0.70	0.27	1.00	0.12
American Federation of Musicians	0.76	444	−0.32	0.24	0.07	0.05

entered Congress. Under this arrangement, candidates who transition from nonincumbents to incumbents enter the data twice as independent row observations. The model is then retrained using data on fundraising by incumbents, with the rows for nonincumbents held completely out of sample. The nonincumbent scores are then inferred from the model fit to the training set.

Figure 5 plots the predictions for the held-out sample of nonincumbents against their future DW-NOMINATE scores. Table 4 reports the same fit statistics as above for

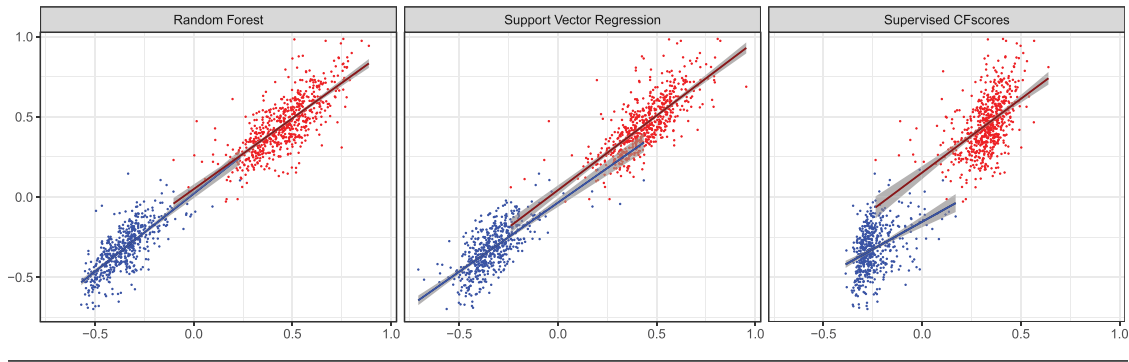
the held-out sample of nonincumbents. The results are in line with those presented in Table 1. They show that fundraising prior to entering office can accurately predict future DW-NOMINATE scores. The overall correlation is 0.97 for the random forest estimates and 0.96 for the support vector regression estimates. Again, this compares favorably with the Nokken-Poole first-term estimates.

Examining the residuals for outliers from the random forest model proves informative. Among the largest outliers are Greg Laughlin, Nathan Deal, and Ben

**FIGURE 4 Partial Dependence**

*Note:* The values for the x-axis represent total amounts given from a PAC to a candidate since 1979.

**FIGURE 5 Forecasting Future DW-NOMINATE Scores of Nonincumbents Based on Fundraising Prior to Entering Office**



**TABLE 4 Forecasting DW-NOMINATE Scores: Cross-Validated Fit Statistics for Held-Out Sample of Nonincumbent Candidates**

	All Candidates			Democratic Candidates			Republican Candidates		
	R	RMSE	N	R	RMSE	N	R	RMSE	N
Random Forest	0.97	0.11	1,162	0.74	0.10	522	0.75	0.12	640
Support Vector Regression	0.96	0.11	1,162	0.70	0.11	522	0.75	0.12	640
Supervised CFscores	0.93	0.15	1,164	0.41	0.14	524	0.51	0.15	640

Nighthorse Campbell. Laughlin, Deal, and Campbell were initially elected as Democrats but switched parties while in office. These examples are of the type that we should expect to deviate from predictions made from contributions raised as nonincumbents. Indeed, their career scores similarly deviate from their voting patterns during their first terms in office.

The results demonstrate that fundraising prior to entering office provides a highly informative signal about future voting behavior. Impressively, it is as predictive of future voting as the votes cast during the first 2 years in Congress.

## Contributor Estimates

Neither of the supervised models produces 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 noncongressional actors from the supervised models greatly increases the number of candidates who can be referenced when locating donors. This in turn greatly increases the number of donors for whom scores can be estimated. The score for donor  $i$  is calculated as

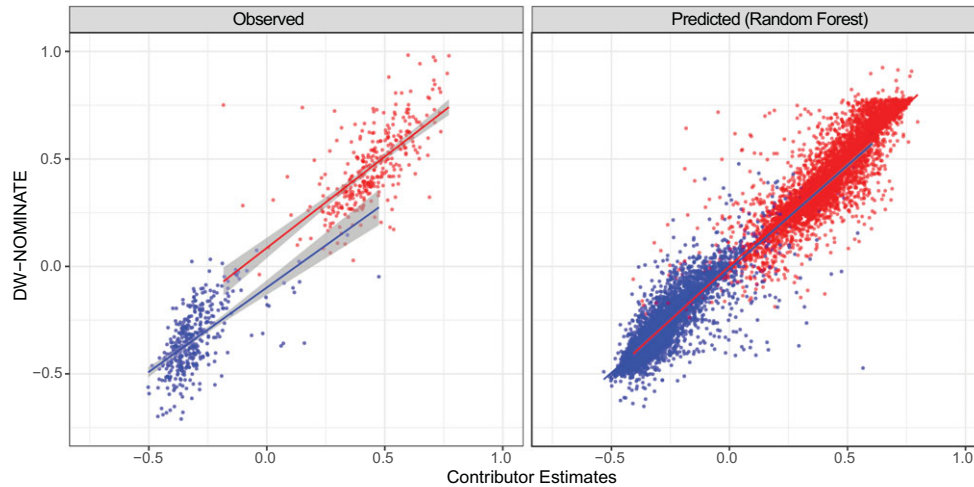
$$\theta_i = \frac{\sum_j \delta_j w_{ij}}{\sum_j w_{ij}}, \quad (6)$$

where  $\delta$  is a vector of recipient estimates and  $w_i$  is a vector of contribution amounts. Left unadjusted, the weighted means will have the effect of shrinking the contributor scores toward 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 6 plots the contributor scores against DW-NOMINATE scores for members of Congress and then against the predicted  $\delta$ s for all other candidates. Only candidates who have personally donated to five or more distinct candidates are included in the analysis.



**FIGURE 6 Contributor Estimates against DW-NOMINATE Scores for Members of Congress and Predicted Scores for Candidates**



The contributor estimates strongly correlate with DW-NOMINATE at 0.95. The within-party correlations are 0.61 for Democrats and 0.66 for Republicans. This suggests that donating, fundraising, and voting all provide consistent signals of candidate positions.

These results speak to an emerging debate in the literature about whether donation patterns accurately measure individual-level ideology. The results presented are consistent with findings based on a survey of donors that shows donation behavior to be ideologically conditioned even among co-partisans (Barber, Canes-Wrone, and Thrower 2017). Hill and Huber (2017) reach a different conclusion based on their analysis of a sample of Cooperative Congressional Election Study (CCES) respondents that had been matched against donor records in DIME. They find that contributor CFscores successfully sort individuals by party but are less successful at discriminating between co-partisans. This claim is supported by results showing that the contributor CFscores only weakly correlate within party with a survey-based measure that the authors constructed by scaling responses to CCES policy items using factor analysis.<sup>11</sup>

This debate helps to illustrate the problems inherent in validating measures of ideology based on comparisons with other unsupervised measures. The conclusion reached in Hill and Huber (2017) rests on the assumption that the measure they constructed offers an appropriate benchmark. For this to be the case, the measure should (1) be internally consistent (reliability), (2) predict outcomes on external criteria (concurrent validity), and (3)

weight issues in a reasonable manner (construct validity). Hill and Huber provide few details about how the CCES factor scores were validated. However, a replication analysis shows the measure to be lacking, especially on the first two counts.<sup>12</sup>

A more productive validation strategy would be to use the contributor CFscores to predict responses to policy-related items. When the contributor CFscores are used to predict responses to a battery of 19 CCES policy items while controlling for party, the estimated coefficient on contributor CFscore is significant for each of the 19 policy items. When this analysis is repeated with donors grouped by party, the coefficient on contributor CFscore is significant for 16 items for Republicans and 18 items for Democrats. (See the supporting information for results.) These results offer a sharp contrast to the inferences drawn based on comparisons with the CCES factor scores. This should caution against relying too heavily on unsupervised measures to validate other measures of ideology,

<sup>11</sup>The reported within-party correlations are 0.22 for Democrats and 0.42 for Republicans. The overall correlation between measures is 0.89.

<sup>12</sup>With respect to internal consistency, a Cronbach's alpha of 0.79 suggests the scores are only moderately reliable. A single factor accounts for just 14.5% of total variance. When the factor model is run separately with respondents grouped by party, a single factor accounts for just a mere 4.5 and 6.6% of total variation for Democrats and Republicans, respectively. The factor scores are similarly poor predictors of policy items not used in constructing the measure. The CCES includes an additional battery of 10 policy items that correspond to specific roll-call votes taken in Congress, which can be predicted out of sample. A baseline partisan model correctly classifies 71.5% of responses. Fitting an IRT model on the CCES roll-call items does significantly better, correctly classifying 77.1% of responses. By comparison, the factor scores barely improve correct classification over a partisan model by a modest 0.4 percentage points to 71.9%. Lastly, six of the 13 questions used to construct the scores focus on immigration policy, thus assigning much greater weight to immigration than other issues.

especially if the chosen benchmark measure has not been extensively validated.

## 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 before 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 provides some initial insight into the process. Future work might 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 over 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 it suggests value in combining information across measures.

## References

- Ahler, Douglas J., Jack Citrin, and Gabriel S. Lenz. 2016. "Do Open Primaries Improve Representation? An Experimental Test of California's 2012 Top-Two Primary." *Legislative Studies Quarterly* 41(2): 237–68.
- Ansolabehere, Stephen, Jr., James M. Snyder, and Charles Stewart III. 2001. "Candidate Positioning in U.S. House Elections." *American Journal of Political Science* 45(1): 136–59.
- Bailey, Michael A. 2013. "Is Today's Court the Most Conservative in Sixty Years? Challenges and Opportunities in Measuring Judicial Preferences." *Journal of Politics* 75(3): 821–34.
- Barber, Michael J., Brandice Canes-Wrone and Sharece Thrower (2017), "Ideologically Sophisticated Donors: Which Candidates Do Individual Contributors Finance?." *American Journal of Political Science*, 61(2): 271–288.
- Barberá, Pablo. 2015. "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data." *Political Analysis* 23(1): 76–91.
- Beauchamp, Nick. 2012. "Using Text to Scale Legislatures with Uninformative Voting." Mimeo, New York University [http://nickbeauchamp.com/work/Beauchamp\\_scaling\\_urrent.pdf](http://nickbeauchamp.com/work/Beauchamp_scaling_urrent.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(1): 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): 367–87.
- Bonica, Adam. 2016a. "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." *Russell Sage Foundation Journal of the Social Sciences* 2(7): 11–32.
- Bonica, Adam. 2016b. "Database on Ideology, Money in Politics, and Elections: Public Version 2.0" [Computer file]. Stanford, CA: Stanford University Libraries [distributor]. <http://data.stanford.edu/dime>.
- 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." Unpublished manuscript, Carnegie-Mellon University.
- Carpenter, Dick, and Jeffrey Milyo. 2012. "The 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." *Fordham Urban Law Journal* 40(2): 603–37.
- Clinton, Joshua D., Simon Jackman, and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *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 E. Apter. New York: Free Press.
- Cox, Gary W., and Matthew D. McCubbins. 2006. *Setting the Agenda*. Cambridge: 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(1): 31–55.
- Drucker, Harris, Chris J. C. Burges, Linda Kaufman, Alex Smola, and Vladimir Vapnik. 1997. "Support Vector Regression Machines." *Advances in Neural Information Processing Systems* 9: 155–61.

- Enelow, James M., and Melvin J. Hinich. 1984. *The Spatial Theory of Voting: An Introduction*. Cambridge: Cambridge University Press 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(Aug): 1871–74.
- Friedman, Jerome H. 2001. "Greedy Function Approximation: A Gradient Boosting Machine." *Annals of Statistics* 29(5): 1189–1232.
- Gerrish, Sean, and David M. Blei. 2012. "How They Vote: Issue-Adjusted Models of Legislative Behavior." In *Advances in Neural Information Processing Systems*, ed. F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger. 2753–61. <http://papers.nips.cc/book/advances-in-neural-information-processing-systems-25-2012>.
- Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21(3): 267–297.
- Groseclose, Tim, Steven D. Levitt, and James M. Snyder Jr. 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): 143–68.
- Hall, Andrew B. 2015. "What Happens When Extremists Win Primaries?" *American Political Science Review* 109(1): 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): 759–74.
- Hill, Seth J., and Gregory A. Huber. 2017. "Representativeness and Motivations of the Contemporary Donorate: Results from Merged Survey and Administrative Records." *Political Behavior* 39(1): 3–39.
- Hinich, Melvin J., and Michael C. Munger. 1996. *Ideology and the Theory of Political Choice*. Ann Arbor: 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, and Tom S. Clark. 2014. "Scaling Politically Meaningful Dimensions Using Texts and Votes." *American Journal of Political Science* 58(3): 754–71.
- Lauderdale, Benjamin E., and Alexander Herzog. 2016. "Measuring Political Positions from Legislative Speech." *Political Analysis* 24(3): 374–94.
- 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): 354–73.
- McCarty, Nolan M., Keith T. Poole, and Howard Rosenthal. 2006. *Polarized America: The Dance of Ideology and Unequal Riches*. Cambridge, MA: MIT Press.
- Nokken, Timothy P., and Keith T. Poole. 2004. "Congressional Party Defection in American History." *Legislative Studies Quarterly* 29(4): 545–68.
- Poole, Keith, Jeffrey Lewis, James Lo, and Royce Carroll. 2011. "Scaling Roll Call Votes with wnominate in R." *Journal of Statistical Software* 42(14): 1–21.
- Poole, Keith T. 2005. *Spatial models of parliamentary voting*. Cambridge University Press.
- Poole, Keith T., and Howard Rosenthal. 2007. *Ideology and Congress*. 2nd rev. ed. New Brunswick, NJ: Transaction.
- Primo, David M. 2013. "Information at the Margin: Campaign Finance Disclosure Laws, Ballot Issues, and Voter Knowledge." *Election Law Journal* 12(2): 114–29.
- Ramey, Adam J., Jonathan D. Klingler, and Gary E. Hollibaugh. 2016. "Measuring Elite Personality Using Speech." *Political Science Research and Methods* 1–22. <https://doi.org/10.1017/psrm.2016.12>.
- 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): 305–17.
- 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(3): 417–48.
- Shor, Boris, and Nolan M. McCarty. 2011. "The Ideological Mapping of American Legislatures." *American Political Science Review* 105(3): 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. 2017. "Estimating Candidates' Political Orientation in a Polarized Congress." *Political Analysis* 25(2): 167–87.
- Windett, Jason H., Jeffrey J. Harden, and Matthew E. K. Hall. 2015. "Estimating Dynamic Ideal Points for State Supreme Courts." *Political Analysis* 23(3): 461–69.
- Yu, Bei, Stefan Kaufmann, and Daniel Diermeier. 2008. "Classifying Party Affiliation from Political Speech." *Journal of Information Technology & Politics* 5(1): 33–48.

## Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Appendix A:** Issue Salience Example with Two-Dimensional Model

**Appendix B:** Comparing Predictive Accuracy for Models with Excluded Features

**Appendix C:** Supervised Learning Applied to Interest Group Legislators Ratings

**Appendix D:** Regressing Donor CFscores on CCES Policy Items