

Estimating Candidates' Political Orientation in a Polarized Congress*

Chris Tausanovitch[†]
Department of Political Science
UCLA

Christopher Warshaw[‡]
Department of Political Science
Massachusetts Institute of Technology

August 2016

Abstract: Over the past decade, a number of new measures have been developed that attempt to capture the political orientation of both incumbent *and* non-incumbent candidates for Congress, as well as other offices, on the same scale. These measures pose the tantalizing possibility of being able to answer a host of fundamental questions about political accountability and representation. In this paper, we examine the properties of six recent measures of candidates' political orientations in different domains. While these measures are commonly viewed as proxies for ideology, each involves very different choices, incentives, and contexts. Indeed, we show that there is only a weak relationship between these measures within party. This suggests that these measures are capturing domain-specific factors rather than just candidates' ideology. Moreover, these measures do poorly at distinguishing between moderate and extreme roll call voting records within each party. As a result, they fall short when it comes to facilitating empirical analysis of theories of accountability and representation in Congress. Overall, our findings suggest that future research should leverage the conceptual and empirical variation across these measures rather than assuming they are all synonymous with candidates' ideology.

*We are grateful for feedback about this project from Gregory Huber, Seth Hill, Howard Rosenthal, Adam Bonica, Walter Stone, Boris Shor, Nolan McCarty, Jon Rogowski, Pablo Barbera, Adam Ramey and participants at the 2015 American Political Science Association Conference. We are grateful to Adam Bonica, Walter Stone, Boris Shor, Nolan McCarty, Jon Rogowski, Pablo Barbera for making publicly available their measures of candidate positions. All mistakes are our own.

[†]Assistant Professor, Department of Political Science, UCLA, ctausanovitch@ucla.edu.

[‡]Assistant Professor, Department of Political Science, Massachusetts Institute of Technology, cwarshaw@mit.edu.

1. INTRODUCTION

Over the past decade, a number of new measures have been developed that attempt to capture the political ideology of both incumbent *and* non-incumbent candidates for Congress, as well as other offices, on the same scale. These measures pose the tantalizing possibility of being able to answer a host of fundamental questions about political accountability and representation. Are ideologically extreme candidates punished at the ballot box (Black, 1948; Downs, 1957; Enelow and Hinich, 1984; Canes-Wrone, Brady, and Cogan, 2002; Hall, 2015)? How much do the policy preferences of a district influence the political orientation of candidates that run for Congress (Ansolabehere, Snyder Jr, and Stewart, 2001)? How much does the available pool of candidates affect the degree of legislative polarization (Thomsen, 2014)? Does variation in electoral rules in primaries affect the policy positions of candidates that run for office (Kousser, Phillips, and Shor, 2015; Ahler, Citrin, and Lenz, Forthcoming; Rogowski and Langella, 2014)? Do ideologically extreme candidates raise less money than centrist candidates (Ensley, 2009)?

In this paper, we examine the properties of six recent measures of candidates' political orientations in different domains. Specifically, we examine the properties of measures based on candidates' NPAT responses from Project Vote Smart (Ansolabehere, Snyder Jr, and Stewart, 2001; Montagnes and Rogowski, 2014), their state legislative voting records (Shor and McCarty, 2011), their Twitter followers (Barberá, 2015), their campaign donor networks (Bonica, 2013b, 2014), the perceptions of survey respondents (Aldrich and McKelvey, 1977; Hare et al., 2014; Ramey, 2016), and expert assessments (Stone and Simas, 2010; Joesten and Stone, 2014; Maestas, Buttice, and Stone, 2014).

It is commonly assumed that these new measures all represent a common dimension of candidates' political ideology. However, this assumption has never been rigorously evaluated. In fact, there are a number of reasons to believe that each of these types of political behavior are driven by domain-specific political orientations rather than ideology. Most importantly,

each of these domains involves very different choices, incentives, contexts, and actors. In order to examine the dimensional structure of these measures, we first examine the correlation between each domain-specific measure of candidates' political orientation. We find that each domain is highly correlated with candidates' party. However, there is only a modest correlation across measures within each party. This suggests that a common dimension of political ideology is probably not the primary driver of each domain-specific measure of political orientation.

Next we evaluate how well each domain-specific measure of political orientation predicts candidates' roll call positions in Congress. Roll call votes are an important mechanism for political accountability and representation. If elections are a meaningful constraint, they must constrain what legislators do, not just what legislators say during the campaign. For this reason, roll call behavior has been used as a standard for convergent validity for nearly all of the measurement models of political orientation that we assess (Barberá 2015, 82, Bonica 2014, 370-371, Hare et al. 2014, 769-770, Joesten and Stone 2014, 745).

Our findings, however, indicate that none of these measures of political orientation are good predictors of candidates' roll call voting patterns within their party. For instance, Republican congressman Dave Reichert's DW-Nominate score places him among the most liberal members of his party in 2010, while measures of his political orientation based upon his campaign finance (CF) contributions (Bonica, 2014) place him in the conservative wing of his party. Peter King's roll call record also places him among the most liberal members of his party. Both survey respondents and his Twitter score, however, place him in the more conservative half of the Republican party. On the Democratic side, Henry Waxman's DW-Nominate score places him among the most liberal of Democrats. But his CF-Score places him in the more conservative half of the Democratic Party. Chris Van Hollen's DW-Nominate score places him in the middle of his party, but experts rated him as one of the most liberal members of the Democratic caucus. Overall, the measures we examine only marginally

improve on candidates' party identification for predicting their roll call behavior.¹

In the penultimate sections, we demonstrate that substituting these measures for a measure based on roll call votes can lead to misleading conclusions to important research questions. As examples, we examine the conclusions that would be reached by using these measures to study polarization and the tendency of extreme candidates to lose their reelection. We find that the measures often lead to inconsistent conclusions. While it may be the case that these measures capture interesting variation in political orientations within their specific domains, future research should seek to explain this variation rather than equating these measures with ideology or roll call voting.

2. BACKGROUND

Measuring the preferences and behavior of political officeholders and candidates is central to the study of American Politics. Many of our most important explanations of political phenomena involve choices in an underlying policy space, with Black's median voter theorem being the canonical example (Black, 1948). Although few scholars think that the median voter theorem should be taken literally, its basic intuition is the bedrock of studies of accountability and representation. Simply put: given voters with certain policy desires, they must have a tendency to select candidates who will push policy in their direction. Otherwise elections will not have any tendency to select candidates who represent the policy desires of voters.

A fundamental part of this notion of preferences is the idea of an underlying policy space in which policies can be closer or farther away from what someone desires. If political actors have a shared understanding of this policy space and the space is low-dimensional, then the literature on measuring policy preferences refers to it as an "ideological space." All of the measures under examination in this paper use the word "ideology" in their abstract,

¹It is important to note that we focus our analysis on recent Congresses. It is possible that these measures perform better in earlier, less polarized, Congresses.

and typically use it throughout as a shorthand for “preferences.” A political actor’s “ideal point” is the unique point in this underlying space that represents that set of policies that they prefer above all others.² A person’s “ideal point” is often said to be a measure of their “ideology” or equivalently, their preferences.

This terminology is problematic because it implies that the measures that arise from various applications of “ideal point estimation” are all measures of the same underlying phenomenon: ideology. Take the most influential work in this literature, which concerns legislative roll call voting (e.g., Poole and Rosenthal, 2011). Roll call voting is central to theories of representation and accountability because roll call voting forces legislators to take a public stand on an issue, and to reap the consequences. Roll call voting is constitutionally linked to policy outcomes: if legislators change their votes, different policy can result. Legislators can then be held accountable for these votes by their constituents.

However, a legislator’s ideal point in the space of roll call votes arises from a complex set of calculations which are specific to that context. It is not necessarily the case that this same set of factors defines the policy space that underlies other political choices. Roll call voting only captures one component of legislators’ broader political orientations.

The challenge for empirical scholars of representation and accountability is that roll call behavior is only available for incumbents. In order to test a number of important theories, scholars need measures of the positions of both incumbents and non-incumbents. For example, in order to examine whether polarization in Congress is increasing, it is important to know whether the ideological positions of Democratic and Republican candidates in each district are diverging over time (Ansolabehere, Snyder Jr, and Stewart, 2001). To examine theories of spatial voting in elections, we need to know whether voters are more likely to vote for the more spatially proximate candidate, which requires measures of the ideological positions of both Democratic and Republican candidates (e.g., Jessee, 2012; Joesten and

²Note that it need not be the case that ideal points actually exist for this concept to be useful. People typically do not know what policies they prefer above all others, but their political decisions can still have a center of gravity- a point that they are the most likely to choose in any pairwise choice.

Stone, 2014; Shor and Rogowski, 2015).

In order to meet this challenge scholars have developed a variety of models that estimate ideal points using disparate data sets. These are usually called models of “ideology.” We use the term “political orientation” because it captures the idea that individuals may have different considerations which come in to play in different political contexts. Voters are presumed to have only one ideology, but their choice, for instance, of which politicians to follow on social media may reflect a different political orientation than their decision of who to vote for.

Measures of candidates’ ideal points fall into three broad categories, shown in Table 1. These measurement models all assume that some observed behavior is primarily generated by candidates’ unobserved, latent political orientation.³ Roll calls voting is an insufficient data source because non-incumbents don’t get to vote. In each case, the political orientation of a given individual is summarized by a single number, which we will denote as the variable x_i where i indexes candidates.⁴

³All of the measures are unidimensional.

⁴The choices in question often have features that are taken into account as well- choices will be indexed by j . In order to contrast the models used, we will attempt below to harmonize the notation, and depart from that used by the original authors.

⁵We downloaded Shor and McCarty’s data from the Dataverse (Shor and McCarty, 2014), and manually matched the estimates of state legislators’ ideal points to their ICPSR numbers that Poole and Rosenthal use to index their DW-Nominate scores. Because state legislative ideal points are only available before legislators take office, we use them in the validation below for non-incumbents.

⁶Jon Rogowski generously shared an expanded version of the data used in Montagnes and Rogowski (2014).

⁷In our evaluation, we focus on the estimates from Ramey (2016), which uses 109,935 survey responses from 2010 and 2012 to estimate the positions of House and Senate candidates. We downloaded the replication data from the dataverse (Ramey, 2015), and used this to analyze the ability of Aldrich-McKelvey scores to predict contemporaneous roll call positions. However, the replication data for Ramey (2015) does not include estimates for non-incumbent candidates. So we used what we believe to be the same data, from the 2010 and 2012 Cooperative Congressional Election Studies, and the same method, to compute our own estimates based on an identical measurement model.

⁸We downloaded the replication data from the Dataverse (Maestas, Buttice, and Stone, 2013). We use the `incl_pc09` variable for incumbent placements, `dlc_pc10` for Democratic candidates’ placements, and `rlc_pc10` for Republican candidates’ placements.

⁹We downloaded the replication data from the Dataverse (Barbera, 2014).

¹⁰We downloaded each congressional candidates’ dynamic and static CF-Score data from Adam Bonica’s DIME website (Bonica, 2013a). We use the dynamic CF-Scores in each of the analyses that follow. However the results are very similar using static CF-Scores.

Table 1: Methods for estimating candidate preferences

Paper	Data	Statistical Model
<u>Model based on Roll Call Voting (No non-incumbents)</u>		
Poole and Rosenthal (2011)	Congressional roll call votes	Spatial choice model (dynamic)
<u>Models based on Candidates' Political Positions Outside Congress</u>		
Shor and McCarty (2011) ⁵	State legislature roll call votes	Spatial choice model
Shor and Rogowski (2015) ⁶	NPAT Responses	Spatial choice model
<u>Models based on Perceptions of Candidate Position</u>		
Ramey (2016) ⁷	Survey respondent perceptions	Measurement error model
Joesten and Stone (2014) ⁸	Expert perceptions	Party-adjusted average
<u>Models based on Spatial Model of Citizen Behavior</u>		
Barberá (2015) ⁹	Followers on Twitter	Spatial choice model
Bonica (2014) ¹⁰	Campaign contributions	Correspondence analysis

2.1. *Models of Political Orientation Based on Political Positions Outside Congress*

One potential approach for measuring the political orientation of candidates is to use information from their political positions outside of Congress. For instance, we could estimate the political orientation of state legislators that run for Congress based on their roll call votes in state legislatures (e.g., Shor and McCarty, 2011) or their responses to political questionnaires from interest groups (e.g., Ansolabehere, Snyder Jr, and Stewart, 2001; Montagnes and Rogowski, 2014).

In this model, legislators choose the outcome on each bill, j , that gives them greater utility: either the status quo, a_j or the policy that would be enacted if the bill were passed, b_j . Their utility for any outcome is a function of the distance between their ideal point, x_i , and the outcome in question, a_j or b_k , plus a random error that represents idiosyncratic or random features of the legislator's utility. If the status quo point is 'closer' to what the legislator wants, then she votes nay. If the bill is closer, she votes yea. The only exception

is if the random shock to her utility is enough to make her prefer the less close option more. This will be more likely when the legislator is close to indifferent between the two options. If we make a few simplifying assumptions, we can write the probability that a legislator votes in favor of a bill (yea) as follows:¹¹

$$P(y_{ij} = Yea) = P((x_i - b_j)^2 - (x_i - a_j)^2 + \epsilon_{ij} > 0) \quad (1)$$

The probability of a vote against (nay) is one minus the probability of a vote in favor. The likelihood of the model is simply the product of the likelihoods of every vote. This model is often referred to as the quadratic utility item response model. The “ideal point” summarizes a legislator’s preferences in the sense that legislators will tend to prefer bills that are closer to their ideal points on average. Observing simply the y matrix of vote choices, we can estimate the latent x ’s that underly those choices.

Shor and McCarty (2011) estimate the ideal points of each state legislature separately, and then bridge together the ideal points of state legislators in different states using Project Vote Smart’s National Political Awareness Test (NPAT) of legislators from 1996 to 2009. In total, they estimate the positions of 18,000 state legislators from the mid-1990s to 2014.¹² A related approach is to use *only* candidates’ responses to questionnaires about their positions. The most widely used questionnaire is the National Political Awareness Test (NPAT) survey conducted by Project Vote Smart.¹³ Ansolabehere, Snyder Jr, and Stewart (2001) use factor analysis to estimate candidates’ spatial positions based on the NPAT survey. More recently, Montagnes and Rogowski (2014), Shor and Rogowski (2015), and others use a spatial utility model similar to equation 1 to estimate candidates’ ideal points based on their NPAT responses. These estimates have been widely used in the applied, empirical literature for

¹¹Poole and Rosenthal (2011) put flesh on this model by assuming a normal curve as the shape of the utility functions, and errors ϵ_{ij} that are logistically distributed. A much simpler formula results if we use quadratic utility with normal errors (Shor and McCarty, 2011). Clinton, Jackman, and Rivers (2004) show that the results of this model are almost identical to the results of Nominate.

¹²It is important to note that these measures are important in their own right for the study of polarization, representation, and accountability in state legislatures, regardless of their ability to predict congressional candidates’ positions.

¹³This is the survey that Shor and McCarty (2011) use to link legislators from different states.

studies on polarization, spatial voting, elections, and other topics.

Of course, only a fraction of state legislators become candidates for Congress, and even fewer win election to Congress. Moreover, a changing constituency in Congress may lead candidates to adapt their behavior (Stratmann, 2000). Finally, the institutional context of Congress is very different from state legislatures in a variety of ways, depending on the state. For example, partisan gatekeeping power varies substantially (Anzia and Jackman, 2013), and the level of party pressure may vary as well (Lee, 2009; Bateman, Clinton, and Lapinski, 2015).

2.2. Models of Political Orientation Based on Perceptions of Candidate Positions

Another approach is to estimate candidates' political orientation from survey respondents' or experts' perceptions of candidates' ideological positions. This approach has the benefit of providing estimates for candidates that did not serve in the state legislature or complete Project Vote Smart's questionnaire. Indeed, conceptually one could imagine survey respondents or experts rating thousands of candidates for all levels of office.

Stone and Simas (2010) and Joesten and Stone (2014) pioneered the use of experts to rate candidates' ideological positions. These studies survey a sample of state legislators and party convention delegates and ask them to place their congressional candidates on a 7-point scale.¹⁴ These "expert informants" can label candidates as either very liberal, liberal, somewhat liberal, moderate, somewhat conservative, conservative, or very conservative. The resulting scores are adjusted by subtracting/adding the average difference between partisans and independents. Averaging responses is a sensible approach if we assume that errors in perceptions are symmetrically distributed.

Although Joesten and Stone (2014) correct for the average "bias" from partisanship, they do not attempt to correct for the fact that individuals often use scales differently. For

¹⁴Maestas, Buttice, and Stone (2014) improve on the measurement model in Joesten and Stone (2014). However, we will focus here on Joesten and Stone (2014) for simplicity.

instance, some individuals may think that “very liberal” is an appropriate term for anyone who is not a Republican whereas others may reserve the term for revolutionary socialists. When individuals are asked to rate a variety of politicians and political entities, their own tendencies in the use of the scale can be accounted for. This observation led Aldrich and McKelvey (1977) to the following model:

$$\tilde{x}_{ij} = w_j(x_i - c_j) + \epsilon_{ij} \tag{2}$$

where \tilde{x}_{ij} is person j 's placement of candidate i ; w_j and c_j are coefficients that capture person j 's individual use of the scale, which can be estimated because each person places multiple candidates and political entities; and x_i is again the actual, latent political orientation of candidate i . Hare et al. (2014) and Ramey (2016) use a Bayesian variant of this model to estimate candidate locations based on the perceptions of survey respondents.¹⁵

However, there are a number of non-ideological factors that could influence survey respondents' and experts' perceptions of candidates' ideology. For instance, more bombastic candidates could be perceived as more ideologically extreme, and candidates with more establishment support could be perceived as more moderate. In addition, candidates' campaign positions, press releases, and television advertisements might not be an accurate reflection of legislators' roll call record, or their underlying ideology (see, e.g., Cormack, 2015; Grimmer, 2013; Henderson, 2013; Rogowski, 2014). All of these factors could lead voters and experts to perceive candidates to be more liberal or conservative than the candidates actually are. Thus, it is important to evaluate how well these measures actually capture legislators' roll call behavior and other measures of candidates' political orientation.

¹⁵Ramey (2016) allows the variance of the error to have a candidate-specific component, and we follow this specification. There are many possible extensions. For instance, Hare et al. (2014) allow the error variance to have both a candidate-specific and a rater-specific component.

2.3. Models of Political Orientation Based on Spatial Models of Citizen Behavior

Another approach is to measure candidates' political orientation based on the idea that some set of behavior by voters or citizens is driven by a spatial model which is a function of candidate positions. For instance, we could assume that citizens donate to spatially proximate candidates. Likewise, we could assume that social network users follow spatially proximate candidates on Facebook and Twitter.

In Barberá (2015), the choice by Twitter users of whether or not to follow political candidates is assumed to be a function of the policy distance between the Twitter user and the candidate.¹⁶ The Twitter user follows the candidate if the utility of doing so is greater than some threshold, t , where utility is once again quadratic. Barberá uses a logistically distributed random error, which is very similar to the normal distribution. So the probability that user j follows candidate i is:

$$P(y_{ij} = \textit{Follow}) = P(-(x_i - \theta_j)^2 + \epsilon_{ij} > t) \quad (3)$$

In order to allow for arbitrary levels of sensitivity to this distance, Barberá (2015) adds a scaling parameter, γ , as well as two different intercepts, recognizing that any given user can only follow so many accounts, and that many candidates have limited name recognition and thus few followers. The term α_i captures candidate i 's overall popularity with users, and β_j captures user j 's propensity for following people on Twitter. These intercepts are arbitrarily scaled, so we can replace the threshold t with an arbitrary fixed number, in this case 0. The following specification results:

$$P(y_{ij} = \textit{Follow}) = P(\alpha_i + \beta_j - \gamma(x_i - \theta_j)^2 + \epsilon_{ij} > 0) \quad (4)$$

Based on this model, Barberá (2015) estimates the latent ideology of several hundred House and Senate candidates using data on 301,537 Twitter users from November of 2012.

¹⁶Twitter is a social media platform that allows users to send brief messages to other users who choose to receive these messages or "follow" them.

In a related paper, Bonica (2014) uses correspondence analysis to estimate candidates' political orientation based on their campaign contributors.¹⁷ The main difference between Barberá (2015)'s model and the correspondence analysis model in Bonica (2014) is that when it comes to campaign contributions, donors must choose both who to give to and how much to give. Bonica recodes all contribution amounts as categories of \$100s of dollars, and uses correspondence analysis to recover ideal points.¹⁸

However, there is no reason to necessarily believe that politicians' political orientation on Twitter, or the profile of their Twitter followers or donors, are based on the same underlying factors that drive legislative behavior. Moreover, there is no reason to necessarily believe that the ideal points recovered by these models are a linear mapping of individuals' underlying ideology. There are a number of factors that could influence donations and social media following. For instance, ideologically extreme voters may be more likely to donate to, and follow on Twitter, more outspoken candidates. Similarly, ideologically extreme voters may be more likely to donate to candidates who specialize in attacking the opposing party. There are many other possible non-ideological drivers of these behaviors.

3. DO THESE DOMAIN-SPECIFIC MEASURES CAPTURE A COMMON DIMENSION OF POLITICAL IDEOLOGY?

Empirical scholars have typically assumed that the measures of political orientation that emerge from these models are synonymous with candidates' underlying political ideology. However, each of these forms of political behavior involves very different choices, incentives, contexts, and actors. There is no reason that they are necessarily equivalent to ideology.

We find that all of the domain-specific measures of political orientation accurately capture

¹⁷Bonica (2014) uses correspondence analysis to estimate the political orientation of virtually every House and Senate candidate between 1980 and 2012 based on over 100 million contributions to political campaigns from 1979 to 2012.

¹⁸The correspondence analysis in Bonica (2014) is meant to approximate an IRT model similar to the one in Barberá (2015). It builds off of an earlier paper, Bonica (2013b), which actually estimates such a model and even includes a role for potentially confounding variables. However, due to the very large size of the donation data, Bonica (2014) opts for this simpler method.

candidates’ partisan identification. In fact, they typically classify candidates into the correct party over 95% of the time. However, this does not necessarily indicate that the measures are structured by a common ideological dimension. Instead, it is possible that they are structured by a combination of partisanship and domain-specific factors. To assess this possibility, Tables 2 and 3 show the within-party correlations between the six domain-specific measures for Democrats and Republicans.¹⁹ In general, the within-party correlation between the various measures of political orientation is low. The average correlation between the various measures is .58 for Democrats and .43 for Republicans. This means that two measures can be expected to explain only 34% of the variance in one another for Democrats and 19% for Republicans. This leads to two potential conclusions. One is that a common dimension of political ideology may not be the primary component underlying the various measures. This would indicate that the individual measures are best viewed as capturing domain-specific political orientations rather than a common dimension of political ideology. Alternatively, it is possible that each measure is capturing an underlying dimension of ideology, but with a great deal of measurement error. In other words, either the accuracy or the precision of the measures is in question.

Table 2: Within-Party Correlation Between Measures of Political Orientations in Different Domains for Democrats

	NPAT	State Leg.	Twitter	CF-Score	Survey	Experts
NPAT	1.00					
State Leg.	0.77	1.00				
Twitter	0.81	0.23	1.00			
CF-Score	0.46	0.53	0.72	1.00		
Survey	0.54	0.68	0.72	0.39	1.00	
Experts	0.66	0.74	0.41	0.26	0.81	1.00

¹⁹In order to generate this analysis, we took the average value of each measure for each candidate (the only exception was CF-Scores, where we used the static CF-Score for each candidate). Then we took the correlation between the measures within each party.

Table 3: Within-Party Correlation Between Measures of Political Orientations in Different Domains for Republicans

	NPAT	State Leg.	Twitter	CF-Score	Survey	Experts
NPAT	1.00					
State Leg.	0.65	1.00				
Twitter	0.20	0.38	1.00			
CF-Score	0.31	0.29	0.44	1.00		
Survey	0.44	0.48	0.45	0.44	1.00	
Experts	0.56	0.48	0.59	0.64	0.04	1.00

4. THE USAGE OF DOMAIN-SPECIFIC POLITICAL ORIENTATION AS A PROXY FOR ROLL CALL VOTES IN CONGRESS

Despite the low correlations between these measure, it is possible that some of them are good predictors of candidates’ roll call votes in Congress. It is important to examine whether they are accurate predictors of roll call votes because these votes are the most obvious means by which legislators can fulfill the obligations of accountability. If elections are a meaningful constraint, they must constrain what legislators do, not just what legislators say during the campaign.²⁰ This is probably the reason why roll call behavior is used as a metric of validation by nearly all of the existing measurement models of candidate positions that we assess below (Barberá 2015, 82, Bonica 2014, 370-371, Hare et al. 2014, 769-770, Joesten and Stone 2014, 745).

Clearly, it is not enough to equate roll call voting itself with ideology. We know already that the measures in question do not have much convergent validity, but roll call voting itself may encapsulate many considerations as well (Poole and Rosenthal, 2011). For instance, many of the most important debates in Congress may be rooted in partisan competition for power rather than ideological disagreements (Lee, 2009). The endogeneity of the choice set (e.g., the roll call agenda) may also influence the mapping between legislators’ ideology and their revealed behavior in different domains (Lee, 2008). It is worth noting, however, that

²⁰Of course, it need not be the case that a legislator’s roll call behavior agrees with the image that she tries to portray of herself, or her own “true” preferences. Indeed, there is research showing that legislators often try to give an impression of themselves that does not reflect their voting records (Cormack, 2015; Henderson, 2013).

Hirsch (2011) shows that some important aspects of the legislative process do not bias ideal point estimates based on roll call data.

We examine how much each measure increases the predictive classification of candidates' roll call votes compared to their party identification alone. We focus on each measure's within-party explanatory power for two reasons. First, a good measure of candidates' political orientation should also be able to outperform measures that are much simpler and more parsimonious. In recent years, over 90% of the variation in roll call behavior can be predicted by the party identification of the legislator. Polarization in Congress has been on the rise since the 1970s (Poole and Rosenthal, 2011). As the parties have become more extreme and more homogeneous, across-party prediction of roll call behavior has become easier and within-party prediction more difficult. Thus, many measures are able to report very high correlations with DW-Nominate and other scaled measures of roll call behavior because they have very high correlations with party ID. The empirical problem with such a measure is not just that it might as well be replaced with party identification. Moreover, understanding within-party variation in political orientation is vitally important for understanding polarization, accountability, and spatial voting. Polarization is a process by which extreme legislators are replacing moderates within each party. In order to identify instances of this process, we need measures of preferences that can accurately identify which candidates in nomination contests are more extreme than other candidates *within* their party. Likewise, spatial voting involves judgements about which candidates are closer in some sense to particular voters, which requires accurate measures of the spatial location of candidates within their party.

For our primary empirical analysis, we run univariate logistic regressions for every vote case in the House of Representatives from 2001 to 2012. We focus on this period because many empirical studies focus on recent congresses, and these congresses may be particularly hard to predict because they are so polarized. Moreover, many of the measures that we evaluate are only available for recent Congresses. For instance, Twitter scores are only

available for the 112th Congress (Barberá, 2015, 82).

For each measure, we calculate predicted votes and compare them to the actual votes. Next, we calculate the Percent Correctly Predicted (PCP) which is simply the percent of all non-missing votes that are correctly predicted. Then, we calculate how much each model improves over a naive model that only uses a dummy variable for the party identification of the legislator to predict their roll call votes. The “Improvement over Party” is the percent reduction in error where the error from the party model is in the denominator.

$$\text{Improvement over Party} = \frac{\sum_{\text{votes}} \text{Party Model Errors} - \text{Errors From This Model}}{\sum_{\text{votes}} \text{Party Model Errors}} \quad (5)$$

We validate each measure against both candidates’ contemporaneous roll call behavior and, for non-incumbents, their future roll call behavior after they win election to Congress. Accurate measures of non-incumbents’ positions are crucial for studies of spatial voting and representation. Indeed, we already have good estimates of incumbent legislators’ behavior based on their roll call positions. Thus, the most common use of the estimates from the recent wave of models is to provide estimates of *non-incumbents’* spatial positions. Few of the existing papers validate their measures of non-incumbents’ positions against their *future* roll call positions.²¹

There are a variety of reasons to think that pre-election measures of candidates’ political orientation may not be accurate predictors of their future roll call records. Although candidates make commitments and promises during their campaigns, these commitments are rarely enforceable (Alesina, 1988). Incumbent legislators are widely believed to be in a highly advantageous position to win reelection (Gelman and King, 1990; Lee, Moretti, and Butler, 2004), so punishing legislators for unkept promises may be difficult, and may even

²¹An exception is Bonica (2014, 371), which validates campaign-finance (CF) scores against candidates’ future DW-Nominate scores across parties. Also, Bonica (2013b, 298-299) validates campaign-finance (CF) scores for non-incumbents against the same candidate’s future CF-Score. But it does not validate them against candidates’ future *roll call behavior*.

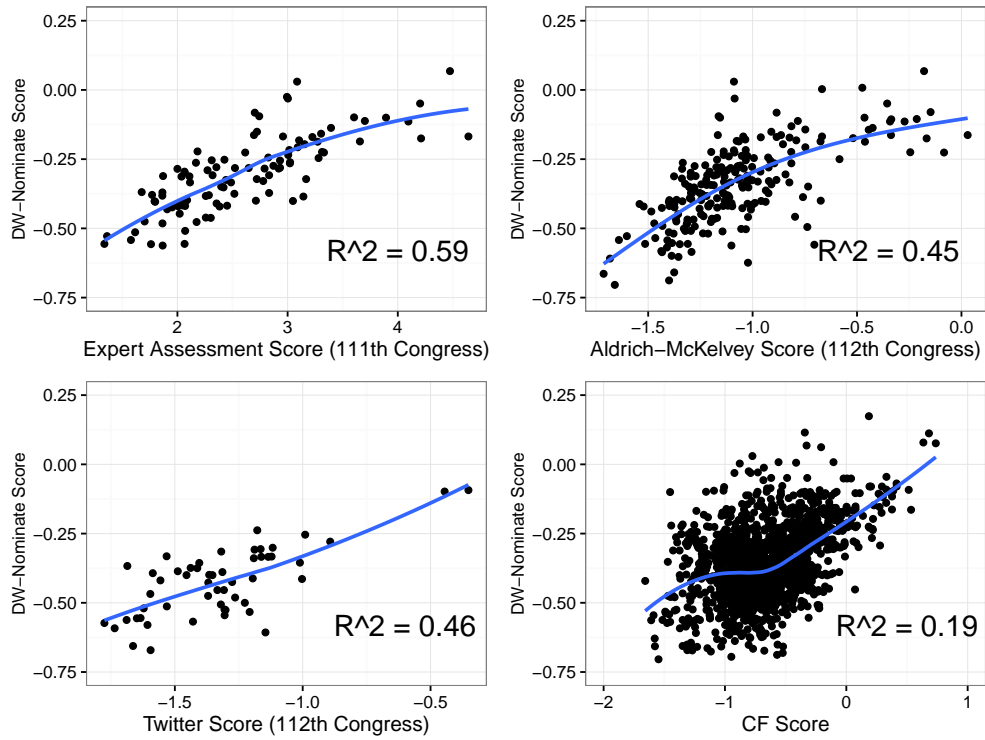
risk electing a legislator from the opposite party. The quirks of political geography are also important in shaping candidate’s support bases. Social media commentators, donors, and the public are limited in the choice of viable candidates to support in any particular district. Information gleaned from these relationships may be a feature of the limited choice set rather than true similarity. As a result, we should not *assume* that measures based on these sources will ultimately reflect actual legislative behavior.

4.1. *Roll Call Votes in U.S. House*

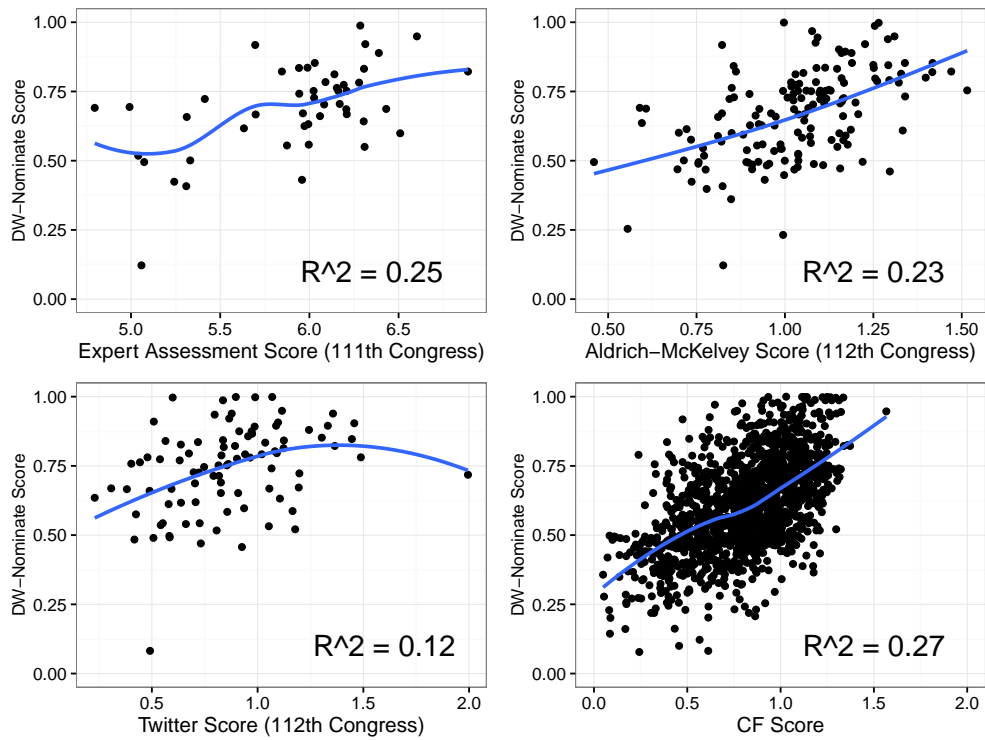
In order to visualize the relationship between each measure and candidates’ contemporaneous roll call behavior, we first examine the correlation between each measure and legislators’ DW-Nominate scores. Figure 1 shows the relationship between each measure and DW-Nominate scores. Each panel contains a scatterplot of individual measurements as well as a loess line to allow a more flexible comparison between the measure and DW-Nominate.

The top panel of Figure 1 shows that none of the measures explain more than 60% of the variation in DW-Nominate scores within the Democratic party, and most of the measures perform much worse than that. The bottom panel of Figure 1 indicates that none of the measures explain more than about a third of the variation in Republicans’ DW-Nominate scores. These figures do not inspire much confidence in the ability of the measures we examine to explain roll call behavior. Of course, our ultimate object is to predict roll call votes, rather than a scaled measure of roll call votes such as DW-Nominate. So, next, we conduct a more detailed evaluation of each model’s ability to provide accurate estimates of the roll call votes of incumbents in the U.S. House between 2001-2012 (Table 4). This table also shows the number of legislator-sessions analyzed, which varies due to the availability of the measures in question. For comparison, we examine how much each measure improves on party ID as a predictor of roll call votes.²²

²²In each row, we calculate how well party predicts roll call votes using the same legislator-sessions that are available for each measure. For example, for the evaluation of Dynamic CF-Scores, we calculate the



(a) Democrats



(b) Republicans

Figure 1: The relationship between DW-Nominate and various measures of candidate positions in the House between 2001-2012

Unsurprisingly, the results of Table 4 mirror the earlier graphs using DW-Nominate scores. Despite the very high importance of party in recent years, DW-Nominate scores substantially improve the classification of votes. This is why we use DW-Nominate scores as a general measure of legislator behavior. It should be noted, however, that DW-Nominate is based on a parametric model. We also include Poole’s Optimal Classification (OC), which maximizes that number of votes correctly classified. OC provides an upper bound for how well a single dimension can classify roll call votes (Poole, 2000). In each case, DW-Nominate performs close to as well as OC.

In contrast, the remaining measures vary significantly in their explanatory power, which is often close to 0, and sometimes even negative. No measure besides DW-Nominate and OC substantially reduces error above and beyond party. Twitter scores and survey-based Aldrich McKelvey scores explain 7.8% and 8.4% of the variation left unexplained by party in the one congress where they are available. This is still only 60% of the reduction in error achieved by DW-Nominate, and less half of the reduction in error achieved through Optimal Classification (Poole, 2000). Moreover, these are contemporaneous comparisons. We will show that they likely overestimate the predictive power of these measures for non-incumbents’ future roll call voting behavior.

Next, we repeat the statistics from Figure 1 and Table 4, but this time each measure is taken from a candidate for the House of Representatives who has not previously held office. Their roll call votes are from the next Congress after they win election. This enables us to assess a counterfactual: how well do these measures capture how non-incumbent candidates would vote in Congress if they were sitting legislators?

Figure 2 shows that none of the measures that were taken before legislators served in Congress predict more than about a third of the variation in DW-Nominate scores within their party after they took office, and most of the measures perform much worse than that. Table 5 shows the predictive results for individual roll call votes. For the observations we

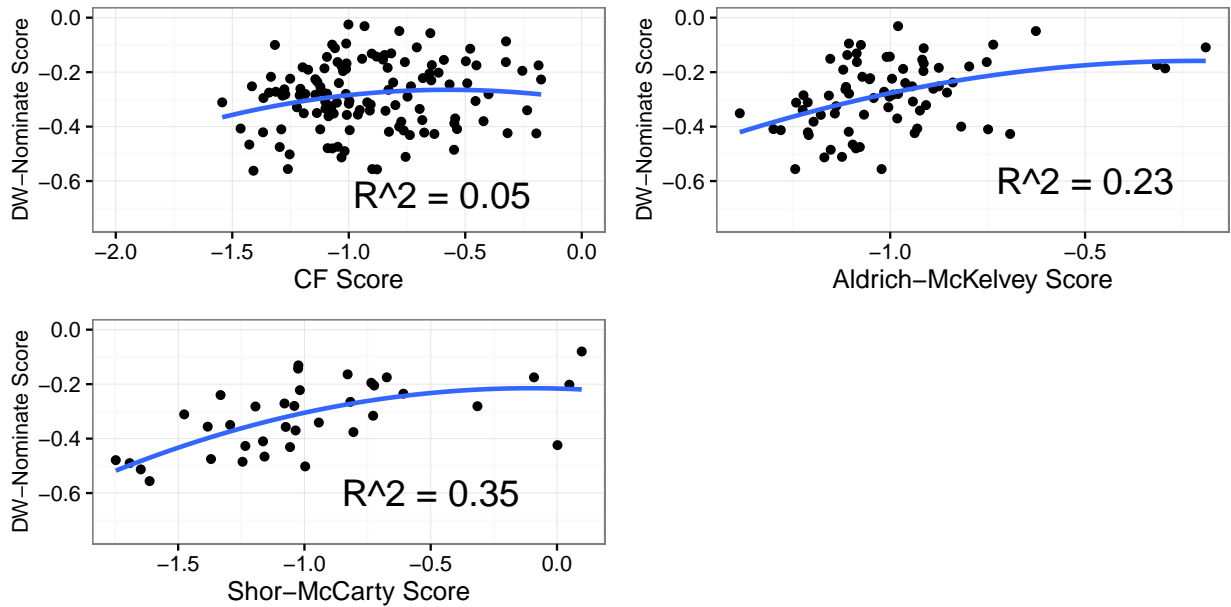
percentage of votes that party id correctly predicts using the 2605 legislator-sessions where dynamic CF-scores are available.

Table 4: Accuracy of Various Models at Predicting Contemporaneous Roll Call Votes in the U.S. House (107-113 Congresses)

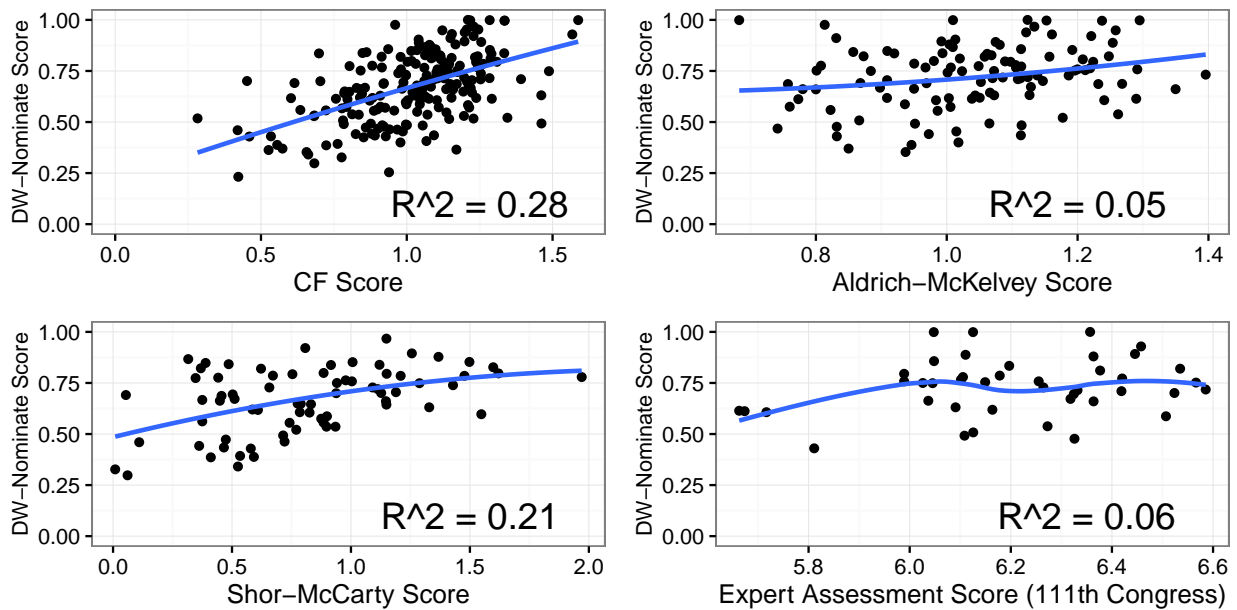
	PCP (Model)	PCP (Party ID)	Improvement over party	Legislator- Sessions
Party ID	91.7%	91.7%	0%	3094
Models based on Spatial Model of Citizen Behavior				
Dynamic CF-Score	91.8%	91.7%	1.1%	2605
Twitter	92.4%	91.8%	7.8%	147
Models based on Perceptions of Candidate Position				
Experts	93.4%	92.9%	7.6%	148
Aldrich-McKelvey	91.7%	90.9%	8.4%	429
Models based on Roll Call Votes				
DW-Nominate	92.9%	91.7%	14.6%	3094
Optimal Classification	93.3%	91.7%	18.4%	3094

do have, the results are much weaker than they were for the contemporaneous comparisons. The exception is Shor and McCarty’s estimates of the ideal points of state legislators. It makes sense that these scores are reasonably good predictors of subsequent roll call behavior in Congress since they are themselves based on roll call behavior. However, they are only available for the very small number of legislators that served in a state legislature prior to sitting in Congress.

Overall, the average accuracy of the six models we examine at explaining within-party variation in roll call votes in Congress is very low. In fact, no model performs much better than a model that assumes one ideal point per party. At the very least, this degree of measurement error should give applied researchers pause. Moreover, this measurement error could be even more problematic if these measures are biased, rather than just noisy. We will revisit the potential for bias in the applications below.



(a) Democrats



(b) Republicans

Figure 2: The relationship between DW-Nominate and various measures of candidate positions in the House in the election before their first term in the House between 2001-2012

Table 5: Accuracy of Various Models at Predicting Prospective Roll Call Votes in the U.S. House (107-113 Congresses)

	PCP (Model)	PCP (Party ID)	Improvement over party	Legislator- Sessions
Party	92.4%	92.4%	0%	461
Models based on Roll Call Behavior Outside Congress				
NPAT	93.9%	93.4%	7.3%	39
State Leg.	94.2%	93.6%	9.3%	110
Model based on Spatial Model of Citizen Behavior				
CF-Score	92.7%	92.4%	3.5%	327
Models based on Perceptions of Candidate Positions				
Experts	92.5%	92.1%	5.2%	44
Aldrich-McKelvey	92.7%	92.6%	1.4%	260

4.2. Roll Call Votes in U.S. Senate

Of course, it is possible is that these measures perform poorly for the House of Representatives because it is inherently difficult to predict the voting records of House members. House members tend to have lower visibility to donors, members of the public, and experts. Some House candidates are political novices, and may not have formed their own views on a variety of issues. The experience of operating in a chamber where majority party control is the norm may alter candidate positions once they begin serving.

In contrast, the United States Senate is a much more visible body, and candidates for the Senate tend to have longer experience in the public eye. Once elected, Senators participate in a legislative body that is noted for its individualism rather than overbearing party control. For these reasons we might expect non-roll call based measures to have better accuracy in the Senate than in the House of Representatives.

The disadvantage of the Senate is a greatly reduced sample size. There are fewer total Senators (100 instead of 435), fewer Senatorial elections (each Senator is up for election every six years instead of 2), and lower turnover. We lack enough data from two of the models

(NPAT and Experts) to test these models at all. For the other measures, we have lower sample sizes for the contemporaneous comparison. For the predictive comparison involving candidates who win, we will not be able to test the Twitter-based measure either.

Table 6: Accuracy of Various Models at Predicting Contemporaneous Roll Call Votes in the U.S. Senate (107-113 Congresses)

	PCP (Model)	PCP (Party ID)	Improvement over party	Legislator- Sessions
Party	89.7%	89.7%	0%	723
Models based on Spatial Model of Citizen Behavior				
CF-Score	88.9%	89%	-1.5%	573
Twitter	89.4%	88.6%	6.4%	66
Model based on Perceptions of Candidate Position				
Aldrich-McKelvey	90.3%	88.9%	12.4%	91
Models based on Roll Call Votes				
DW-Nominate	91.4%	89.7%	17%	723
Optimal Classification	91.6%	89.7%	18.8%	723

Table 6 shows the contemporaneous comparison for the Senate. In most cases, the fit is substantially higher for these measures than in the case of the House of Representatives. Aldrich-McKelvey scores perform particularly well. However, the overall predictive power of these measures is still limited.

Table 7 repeats the analysis above using the candidate scores for candidates who have not yet held Senate seats and their later roll call behavior as Senators. Unfortunately, due to the small amount of turnover in the Senate during this period, we have very small sample sizes to work with. We begin with only 87 new senators. Of these, 67 have CF-Scores. Once again, CF-Scores have limited predictive value. Aldrich-McKelvey and Shor-McCarty scores show more promise, but with only 48 and 10 observations, respectively, we cannot draw any firm conclusions. Legislators who appear in these data are not necessarily representative of the broader set of Senators.

Table 7: Accuracy of Various Models at Predicting Prospective Roll Call Votes in the U.S. Senate

	PCP (Model)	PCP (Party ID)	Improvement over party	Legislator- Sessions
Party	92.5%	92.5%	0%	100
Model based on Roll Call Behavior Outside Congress				
State Leg.	97.3%	96.4%	24.4%	10
Model based on Spatial Model of Citizen Behavior				
CF-Score	92.9%	92.1%	9.7%	75
Model based on Perceptions of Candidate Positions				
Aldrich-McKelvey	93.8%	92.2%	20.3%	48

4.3. Polarization

There is a vast literature that examines changes in polarization over time among legislators and candidates. In their authoritative study, McCarty, Poole, and Rosenthal (2006) show that legislators' roll call records have polarized asymmetrically, with virtually all of the polarization occurring among Republicans. In line with this finding, the upper panel of Figure 3 shows that between 1980 and 2012 the two parties' DW-Nominate scores drifted steadily apart, with nearly all of the polarization in DW-Nominate scores occurring among Republicans. The middle panel shows polarization in incumbents' CF-Scores over this time period. A number of recent empirical studies have used CF-Scores to examine the causal factors that drive polarization in state legislatures and Congress (e.g., Ahler, Citrin, and Lenz, Forthcoming; Rogowski and Langella, 2014; Thomsen, 2014). This panel indicates that unlike their roll call records, the two parties' campaign donor scores did not start moving apart until the mid-1990s. Moreover, the bulk of the polarization appears to have occurred among Democrats. Finally, the bottom panel shows polarization in NPAT-scores. It indicates that virtually all of the polarization in NPAT-scores occurs among Democrats. Moreover, the two parties' NPAT-scores only modestly moved apart during the time period when these data are available.

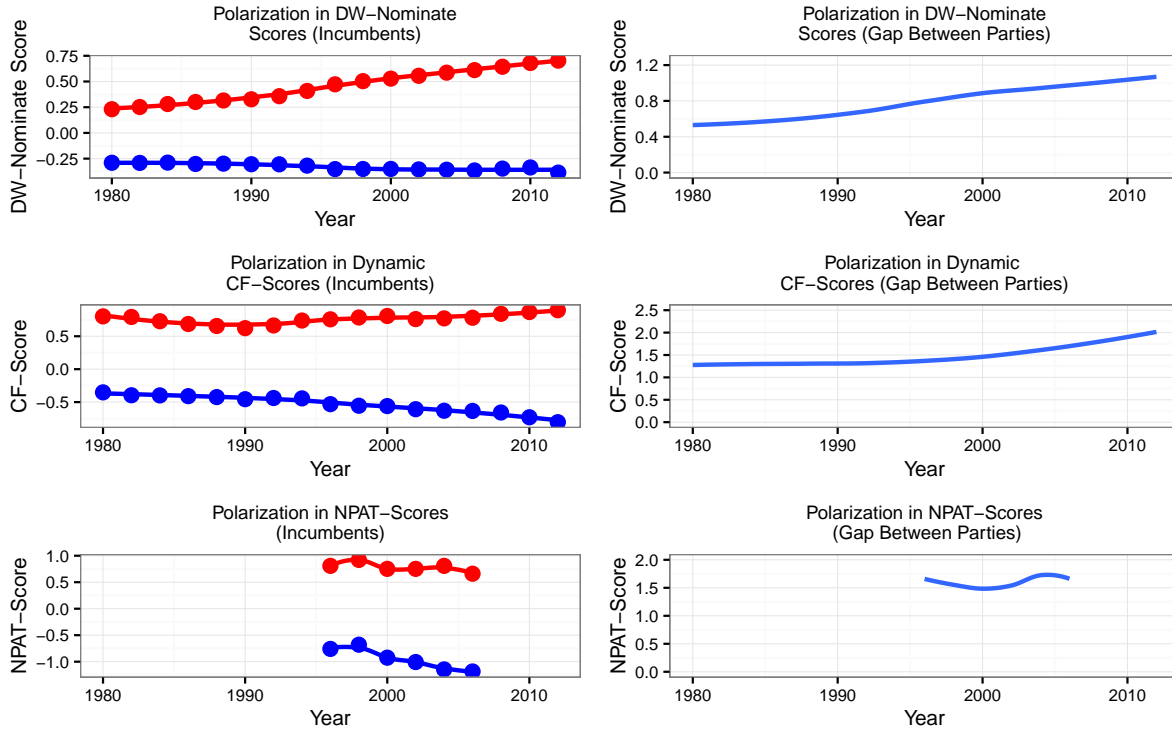


Figure 3: The evolution of DW-Nominate and various measures of candidate positions for Democrats and Republicans in the House between 1980-2012 (NPAT scores are only available from 1996-2006). Blue dots show the mean spatial position of Democrats and red dots show the mean spatial position of Republicans.

Overall, these plots indicate that DW-Nominate scores, NPAT-scores, and CF-scores each show a different story regarding the relative changes in polarization in recent congresses. Most problematically, the results show that NPAT and CF-scores are not just noisy measures of candidates' roll call positions. They are actually biased. For instance, in contrast to DW-Nominate, Democrats' NPAT responses and campaign finance scores are growing more extreme much more rapidly. This suggests that scholars should use caution in using non-roll call based measures of candidate political orientation to make inferences about hypotheses that implicate actual legislative behavior.

4.4. The Effect of Political Orientation on Elections

An important question in the study of representation and accountability is how various metrics of political orientation are associated with election results. There is a large literature that argues that legislators that take extreme roll call positions are penalized at the ballot box (e.g., Canes-Wrone, Brady, and Cogan, 2002). However, there is less work concerning how other metrics of political orientation are related to election results (c.f., Hall, 2015).

	<i>Dependent variable:</i>				
	General Election Percentage				
	(1)	(2)	(3)	(4)	(5)
CF-Score	5.613*** (0.518)				
NPAT-Score		3.426*** (0.664)			
Twitter			-3.751 (3.046)		
Survey				0.322 (1.608)	
Experts					2.296*** (0.613)
Rep. Presidential Vote	-10.811*** (0.448)	-18.810*** (0.696)	-11.455*** (3.163)	-10.973*** (1.926)	-15.128*** (0.750)
Constant	40.314*** (1.175)	52.779*** (1.154)	52.306*** (3.851)	54.797*** (2.074)	48.616*** (0.576)
Observations	3,552	672	51	160	104
R ²	0.209	0.557	0.315	0.225	0.846
Adjusted R ²	0.207	0.553	0.287	0.215	0.843

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 8: Relationship between political orientation and election results for Democratic candidates.

In this section, we examine whether the conclusions of this literature vary depending on which measure of political orientation is used. Following the literature, we examine how

measures of political orientation in different domains are correlated with election results after controlling for constituency preferences. We standardize all the independent variables to make the results easily comparable across models. All independent variables are oriented such that higher values are more conservative. Tables 8 shows the results for Democrats and Table 9 shows the results for Republicans.

	<i>Dependent variable:</i>				
	General Election Percentage				
	(1)	(2)	(3)	(4)	(5)
CF-Score	-3.484*** (0.547)				
NPAT-Score		-2.273*** (0.655)			
Twitter			-1.919 (1.472)		
Survey				0.185 (0.835)	
Experts					-0.719 (2.251)
Rep. Presidential Vote	7.228*** (0.450)	14.029*** (0.648)	13.785*** (2.114)	9.778*** (1.197)	7.191** (3.436)
Constant	37.474*** (0.455)	47.844*** (0.620)	57.358*** (1.361)	57.225*** (0.895)	62.945*** (2.370)
Observations	3,524	517	81	213	51
R ²	0.079	0.486	0.353	0.269	0.095
Adjusted R ²	0.078	0.484	0.337	0.262	0.057

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 9: Relationship between political orientation and election results for Republican candidates.

In general, candidates with more extreme political orientations across the various domains suffer at the ballot box, and more moderate political orientations are rewarded, following the conclusions of the literature. For instance, a one-standard deviation to the right in

Democrats' CF-Scores (i.e., their donor network) corresponds to a 5.6% increase in their general election vote share. For Republicans, a one-standard deviation to the left in their CF-Scores (i.e., their donor network) corresponds to a 3.5% increase in their general election vote share.

However, the effect magnitude varies substantially across measures. CF-Scores always have the largest effect despite the fact that they are consistently among the weakest predictors of roll call voting. The effects of scores based on Twitter, surveys, and expert ratings are inconsistent in terms of sign and significance, although smaller sample sizes are one source of this uncertainty. Once again there is not enough evidence to justify the use of these scores in a study of accountability without abundant caution in interpretation. Most of these measures appear to be related to election outcomes, suggesting that they may be measuring something of political importance. The question for future research is how these phenomena affect politics given that they are only weakly related to legislative behavior or a common ideological space.

5. CONCLUSION

Despite the development of a variety of innovative strategies for measuring the political positions of candidate for Congress, existing measures do not measure the same underlying dimension, and have only limited predictive power in terms of the voting records that candidates establish once elected. Even contemporaneous measures, which use data on legislators as they are currently serving in Congress, typically fail to explain even half the variation in legislator' roll call voting, and usually closer to a third. The performance of these measures varies across parties, with no measure clearly dominant. As a result, the usage of these measures of candidate positions could lead to serious inferential errors for substantive, applied research. For instance, we have shown that different measures of candidate positions lead to dramatically different inferences for studies of polarization and representation. These

findings have important implications for academic research, as well as for our understanding of democracy. Prospective voting requires voters, not just political scientists, to know what candidates will do if elected, and these results suggest that this predictive exercise is very difficult.

Overall, our findings call into question the usefulness of these measures for examining questions that depend on the relative spatial distance between candidates, such as tests of spatial voting theories or the causes of Congressional polarization.²³ There are a variety of explanations that may account for the fact that constituents' implicit (e.g., campaign finance donations or twitter following) and explicit (e.g., survey responses) perceptions of candidates' political orientation are both only weakly associated with candidates' roll call behavior inside of Congress. Although candidates make commitments and promises during their campaigns, these commitments are rarely enforceable (Alesina, 1988). Moreover, candidates have a variety of reasons to distort their positions during the campaign. This may weaken the relationship between candidates' campaign platforms and their roll call positions (Rogowski, 2014). The ability of constituents to predict roll call behavior may be further distorted by political geography. Indeed, social media commentators, donors, and the public are limited in the choice of viable candidates to support in any particular district. Information gleaned from these relationships may be a feature of the limited choice set rather than true similarity. Finally, there are a variety of factors that could influence candidates' roll call votes (e.g., lobbying, agenda-control, party leaders, etc).

While these measures perform poorly at predicting legislators' roll call positions, they may have other valuable uses. They could potentially be used to impute the partisanship of candidates and voters when other information on their partisanship is not readily available (Hill and Huber, 2015). If interpreted and validated properly, they may be useful for examining potential explanations for the mismatch between survey respondents' perceptions

²³Whether or not these measures are useful depends on the application in question. Even relatively weak proxy measures can sometimes produce orderings that are correct a substantial fraction of the time. However, comparisons of relative distances can be highly inaccurate.

and candidates' actual roll call positions (e.g., Cormack, 2015; Grimmer, 2013; Henderson, 2013). These measures also have a number of potential applications for specific substantive questions outside the realm of legislative behavior. For instance, CF-Scores could be used to examine the campaign finance behavior of bureaucrats (Bonica et al., 2015) and Barberá (2015)'s measures of candidates' Twitter followers could be used to examine the effect of candidates' roll call positions on their followings on social networks. However, even in these applications the authors must answer some difficult questions for the results to be meaningful. What is it that these measures are capturing, why is this kind of measure necessary, and why is the underlying latent variable important to study?

It is important to note that our findings do not imply that it is impossible to find a better measure of candidates' spatial positions. On the contrary, we hope that new data sources and statistical tools will facilitate more accurate estimates of the positions of candidates than extant measures (e.g., Bonica, 2016). However, future researchers should measure success by a high standard given the importance of these measures for a variety of fundamental substantive questions.

REFERENCES

- 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 .
- Aldrich, John H, and Richard D McKelvey. 1977. “A Method of Scaling with Applications to the 1968 and 1972 Presidential Elections.” The American Political Science Review 71(1): 111–130.
- Alesina, Alberto. 1988. “Credibility and Policy Convergence in a Two-Party System with Rational Voters.” American Economic Review 78(4): 796–805.
- Ansolabehere, Stephen, James M Snyder Jr, and Charles Stewart. 2001. “Candidate Positioning in US House Elections.” American Journal of Political Science 45(1): 136–159.
- Anzia, Sarah F, and Molly C Jackman. 2013. “Legislative organization and the second face of power: Evidence from US State Legislatures.” The Journal of Politics 75(01): 210–224.
- Barbera, Pablo. 2014. “Replication data for: Birds of the Same Feather Tweet Together. Bayesian Ideal Point Estimation Using Twitter Data.” <http://dx.doi.org/10.7910/DVN/26589>.
- Barberá, Pablo. 2015. “Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data.” Political Analysis 23(1): 76–91.
- Bateman, David A, Joshua Clinton, and John Lapinski. 2015. “A House Divided? Roll Calls, Polarization, and Policy Differences in the US House, 1877–2011.”
- Black, Duncan. 1948. “On the Rationale of Group Decision-Making.” The Journal of Political Economy 56(1): 23–34.

- Bonica, Adam. 2013a. "Database on Ideology, Money in Politics, and Elections: Public version 1.0." <http://data.stanford.edu/dime>.
- Bonica, Adam. 2013b. "Ideology and Interests in the Political Marketplace." American Journal of Political Science 57(2): 294–311.
- Bonica, Adam. 2014. "Mapping the Ideological Marketplace." American Journal of Political Science 58(2): 367–386.
- Bonica, Adam. 2016. "Inferring Roll-Call Scores from Campaign Contributions Using Supervised Machine Learning." Unpublished manuscript. Available for download at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2732913.
- Bonica, Adam, Chen Jowei, Johnson Tim et al. 2015. "Senate Gate-Keeping, Presidential Staffing of Inferior Offices, and the Ideological Composition of Appointments to the Public Bureaucracy." Quarterly Journal of Political Science 10(1): 5–40.
- Canes-Wrone, Brandice, David W. Brady, and John F. Cogan. 2002. "Out of Step, Out of Office: Electoral Accountability and House Members' Voting." American Political Science Review 96(1): 127–140.
- Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." American Political Science Review 98(2): 355–370.
- Cormack, Lindsey. 2015. "Extremity in Congress: Communications versus Votes." Unpublished manuscript. Available for download at personal.stevens.edu/~lcormack/extreme_comms_votes.pdf.
- Downs, Anthony. 1957. An Economic Theory of Democracy. New York: Harper & Row.
- Enelow, James M, and Melvin J Hinich. 1984. The Spatial Theory of Voting: An Introduction. Cambridge University Press.

- Ensley, Michael J. 2009. "Individual Campaign Contributions and Candidate Ideology." Public Choice 138(1-2): 221–238.
- Gelman, Andrew, and Gary King. 1990. "Estimating incumbency advantage without bias." American Journal of Political Science 34(4): 1142–1164.
- Grimmer, Justin. 2013. Representational Style in Congress: What Legislators Say and Why It Matters. Cambridge University Press.
- 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. 2014. "Using Bayesian Aldrich-McKelvey Scaling to Study Citizens' Ideological Preferences and Perceptions." American Journal of Political Science .
- Henderson, John Arthur. 2013. "Downs' Revenge: Elections, Responsibility and the Rise of Congressional Polarization." Unpublished PhD Dissertation. Available for download at <http://gradworks.umi.com/36/16/3616463.html>.
- Hill, Seth, and Greg Huber. 2015. "Representativeness and Motivations of Contemporary Contributors to Political Campaigns: Results from Merged Survey and Administrative Records." Unpublished manuscript. Available for download at http://www.sethjhill.com/HillHuberDonorate_062515.pdf.
- Hirsch, Alexander V. 2011. "Theory driven bias in ideal point estimates—A Monte Carlo study." Political Analysis 19(1): 87–102.
- Jessee, Stephen A. 2012. Ideology and Spatial Voting in American Elections. New York, NY: Cambridge University Press.
- Joesten, Danielle A, and Walter J Stone. 2014. "Reassessing Proximity Voting: Expertise, Party, and Choice in Congressional Elections." The Journal of Politics 76(3): 740–753.

- Kousser, Thad, Justin Phillips, and Boris Shor. 2015. "Reform and Representation: A New Method Applied to Recent Electoral Changes." Unpublished manuscript. Available for download at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2260083.
- Lee, David S, Enrico Moretti, and Matthew J Butler. 2004. "Do voters affect or elect policies? Evidence from the US House." The Quarterly Journal of Economics 119(3): 807–859.
- Lee, Frances E. 2008. "Agreeing to disagree: Agenda content and Senate partisanship, 1981–2004." Legislative Studies Quarterly 33(2): 199–222.
- Lee, Frances E. 2009. Beyond ideology: politics, principles, and partisanship in the US Senate. University of Chicago Press.
- Maestas, Cherie D., Matthew K.; Buttice, and Walter J. Stone. 2013. "Replication data for: Extracting Wisdom from Experts and Small Crowds: Strategies for Improving Informant-based Measures of Political Concepts." <http://dx.doi.org/10.7910/DVN/23170>.
- 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, MA: MIT Press Cambridge.
- Montagnes, B Pablo, and Jon C Rogowski. 2014. "Testing Core Predictions of Spatial Models: Platform Moderation and Challenger Success." Political Science Research and Methods Forthcoming.
- Poole, Keith T. 2000. "Nonparametric unfolding of binary choice data." Political Analysis 8(3): 211–237.
- Poole, Keith T, and Howard L Rosenthal. 2011. Ideology and Congress. Transaction Publishers.

- Ramey, Adam. 2015. "Replication Data for: Vox Populi, Vox Dei? Crowd-Sourced Ideal Point Estimation." <http://dx.doi.org/10.7910/DVN/JDPNNU>.
- Ramey, Adam. 2016. "Vox Populi, Vox Dei? Crowdsourced Ideal Point Estimation." Journal of Politics 78(1).
- Rogowski, Jon C. 2014. "Faithful Agents? Campaign Positions, Legislative Behavior, and the Partisan Organization of Congress."
- Rogowski, Jon C, and Stephanie Langella. 2014. "Primary Systems and Candidate Ideology Evidence From Federal and State Legislative Elections." American Politics Research 43(5): 846–871.
- Shor, Boris, and Jon C Rogowski. 2015. "Ideology and the US Congressional Vote." Unpublished manuscript. Available for download at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2650028.
- Shor, Boris, and Nolan McCarty. 2011. "The Ideological Mapping of American Legislatures." American Political Science Review 105(03): 530–551.
- Shor, Boris, and Nolan McCarty. 2014. "Individual State Legislator Shor-McCarty Ideology Data, July 2014 update." <http://dx.doi.org/10.7910/DVN/26805>.
- Stone, Walter J, and Elizabeth N Simas. 2010. "Candidate valence and ideological positions in US House elections." American Journal of Political Science 54(2): 371–388.
- Stratmann, Thomas. 2000. "Congressional voting over legislative careers: Shifting positions and changing constraints." American Political Science Review 94(03): 665–676.
- Thomsen, Danielle M. 2014. "Ideological Moderates Won't Run: How Party Fit Matters for Partisan Polarization in Congress." The Journal of Politics 76(03): 786–797.

A. APPENDIX A: COMPARISON OF OUR RESULTS WITH THE VALIDATION METRICS IN THE SOURCE PAPERS

In this section, we compare our results to the results reported in the original papers that we evaluate. Table 10 shows the percentage of the variation in incumbents’ DW-Nominate scores explained by each model that we report in the main text. It also shows the percentage of the variation in DW-Nominate scores that each of the source papers reports that their model explains.

Table 10: Validation Statistics for Various Measurement Models Against Contemporaneous Nominate Scores in the U.S. House - Comparison with Results in Source Papers

Name	Survey Respond’s	Experts	Twitter	CF-Score
Success in Explaining Within-Party Variation in DW-Nominate Scores for Democrats				
% Variation Explained (our analysis)	0.45	0.59	0.48	0.19
% Variation Explained (source paper)	0.57	0.52	0.52	0.31
Success in Explaining Within-Party Variation in DW-Nominate Scores for Republicans				
% Variation Explained (our analysis)	0.23	0.25	0.08	0.28
% Variation Explained (source paper)	0.29	0.31	0.12	0.44

Overall, our results regarding the amount of within-party variation explained by each model are very similar to the results reported in the source papers.²⁴ The only notable differences between our results and those in the source papers are for CF-Scores (Bonica, 2014) and the Survey-based scores (Ramey, 2016). This difference for CF-Scores likely stems from the fact that we focus on the 107-113 Congresses, while Bonica (2014, 370-371) focuses on the 96-112 Congresses. There appears to have been a tighter relationship between CF-Scores and DW-Nominate scores in earlier Congresses. Also, we use dynamic rather than static CF-Scores. The difference for the survey-based scores likely stems from the fact that Ramey (2016) uses W-Nominate rather than DW-Nominate to validate his estimates, and

²⁴Note that the results from Barberá (2015) that we report in Table 10 were calculated from his replication data based on members of the U.S. House in the 112th Congress. The correlations reported in the paper are somewhat higher, but they include both members of the U.S. House and Senate.

there is a slightly higher correlation between the survey scores and W-Nominate than there is with DW-Nominate.