Parties and Agenda Setting: A Final Passage Approach to Ideal Point Estimation*

Joshua Y. Lerner

Department of Political Science, Duke University

May 10, 2023

Abstract

Canonical approaches to the study of ideal points in Congress are all built off of a theoretical Congress that treats casting a vote on any given roll call as an individualized independent action. Ignoring the structure of votes leads to the creation of measures that obstruct how parties structure votes in such a way to ensure the legislative outcomes they desire: positive and negative agenda control. Building off of a growing literature on the problems with existing measures of legislative preferences, I propose a more straightforward approach to ideal point estimation that privileges final passage votes over others and therefore takes seriously the use of agenda control by the majority party to shape outcomes. In doing so, I find the effects of party on individual roll calls conditional on the majority status of the MC, as well as significant discrepancies in how these votes impact our existing measurements of ideal points. I further introduce a series of tests from the psychometrics literature on the viability of latent traits to identify which of these measures are most consistent and generalizable. I find that partisan centric ideal points are more accordant with these tests than existing measures. With these final-passage generated ideal points, I find more evidence of MC success and failures consistent with theories of both positive and negative agenda control.

*Prepared for presentation at the Annual Meeting of the American Political Science Association, September 10-13, 2020.

In the past two decades, unidimensional spatial models of lawmaking have come to play a central role in empirical studies of the U.S. Congress. Such models typically assume that legislators have (1) spatial preferences over a unidimensional set of policy outcomes, and (2) complete and perfect information about each others' preferences and the link between policy choices and outcomes. Unidimensional spatial models are particularly amenable to empirical testing because each legislator's policy preferences are fully characterized by his most preferred policy or ideal point. Accordingly, political methodologists have developed several powerful statistical methods such as DW-NOMINATE (Poole & Rosenthal 2000, 2011) and IDEAL (Clinton *et al.* 2004) for estimating ideal points from Congressional roll call data.

One of the most well-known theoretical models in political science is the spatial model of voting. Although scholars across many of the political science subfields have applied the model in their areas of research, the model plays an especially prominent role in the foundational theories in the congressional literature (Cox & McCubbins 1993; Krehbiel 1998; Aldrich & Rohde 2000; Cox & McCubbins 2005). Soon after the introduction of the model, scholars struggled to develop ways to measure the voting behavior of elected officials in order to place them on an ideological scale. For many, Poole & Rosenthal (1985, 2000) went a long way towards solving the problem with the introduction of NOMINATE. Few developments in the last quarter century have had as much impact on congressional research, in particular, as the creation of the NOMINATE procedure for roll-call analysis. Poole and Rosenthal employed the technique, based on a simple spatial model with probabilistic voting, to recover legislators' positions and roll-call outcomes across the entire history of congressional voting, derived in a simple latent measure.

As a measurement question, latent trait models (LTM) make up vast majority of methods work in Congress. These measurement models inform all of spatial modeling and, therefore, most theoretical work in Congress. As an example, Clinton (2007) has proposed an interesting hypothesis for the failure of ideal point-driven tests to adjudicate between partisan and nonpartisan pivot theories. Clinton argues in effect that pivot theories predict legislative behavior that should result in estimated ideal points for pivotal legislators that are statistically indistinguishable, and therefore that pivot theories are also subject to the endogeneity critique. Connecting evaluative standards from latent trait modeling to the assumptions underlying the spatial theory of voting (and therefore ideal point measurement), I show that these are inextricably linked in the underlying theory of measurement as well as addressed most directly by *limiting* the roll calls used to estimate ideal points. Pursuant to this end, I estimate ideal points separating out final passage votes from non-final passage votes and evaluate derived latent measures of spatial preferences. As will be shown in this paper, limiting the rolls used in estimating ideal points can increase the consistency of ideal points and shed new light on questions of agenda setting in Congress.

In this paper, I first evaluate existing ideal point models as latent traits, applying the set of criteria from psychometrics to evaluate them on consistency and reliability. Then, I take a new approach to ideal point modeling that models final passage and non-final passage votes separately. Comparing the ideal points estimated under those conditions to DW-NOMINATE, I test how majority status impacts distributions of preferences, and how that impacts within-member changes to their rankings and scores. Finally, I use these score to test agenda setting power in Congress. Namely, I reevaluate the propensity of individual members of Congress to be rolled, and find that models that are built on final passage based ideal points clearly support more robustly partisan agenda control theories of Congressional policymaking, and find strong evidence for both positive and negative agenda setting powers.

Role of Latent Trait Measures in Spatial Models

Spatial models of politics are all built along the same basic framework. Scholars of legislative politics have long noted how proposals near the median along a left-right ideological spectrum gain the support of a majority. Variants of such models have accounted for proposal power (Romer & Rosenthal 1978), committees as gatekeepers (Denzau & Mackay 1983), bicameralism and super-majority rules (Krehbiel 1998), and agenda-setting parties (Cox & McCubbins 2005). Throughout all the various changes to the model over the years, all spatial theories of voting have the same basic components: policy space, actors, preferences, behavioral assumptions, institutions (the rules of the game), and information. Roll calls appear to offer an ideal opportunity for measuring the preferences of political elites because we observe elites registering their individual support on many proposed policies.[1]

The source of most of the scholarly debate has been the ideal point model. All roll call models assume that legislators have a most-preferred point in the policy space that they would implement if they could. Legislators' policy preferences are satiable, and the ideal point reflects some combination of personal, constituency, and other heterogeneous policy preferences. In fact, an entire literature explores why legislators take the positions they do. Whether preferences are defined over policy outcomes or public positions is unclear, and the meaning likely varies by application.

Given a most-preferred outcome or position in the policy space, the model of individual choice in roll call analyses assumes that legislators vote for the closest expected alternative in every pairwise comparison. We can assume that legislator i votes yes on vote t (i.e. $y_{i,t} = 1$) if the policy position associated with voting yes ($\theta_{y(t)}$) is closer to the most-preferred ideal point x_i than the position associated with voting no ($\theta_{n(t)}$). The model of individual choice underlying the measurement assumes that the only relevant information to understanding a given vote is the relative distance between a legislator's most-preferred policy and the alternative being considered (and the status quo or reversion point).

All spatial models assume the existence of a true ideal point x_i, but creating a model of roll calls only recovers the preferences given the observed votes. It is also restrained by the

[1] Although the term "roll call" technically refers to the recording of a vote in a deliberative body such as a legislature, the methods that were originally developed to analyze votes in the US Congress have been applied to decision-making behavior by non-legislators in executive and judicial branches throughout the world (e.g., the United States Supreme Court justice, public positions taken by a president, votes in the United Nations' General Assembly), as well as in groups not formally affiliated with government (e.g., interest groups) (Clinton 2012).

assumed model of individual choice as implemented via a statistical model. It is therefore possible that the preferences revealed in the pattern of votes differ from true preferences. If legislators vote for alternatives because of non-spatial considerations, the model of individual choice that is assumed by common estimators is incorrect. In essence, an ideal point is a particular type of latent trait that is meant to capture the revealed preferences of a given policy actor.

The key issue is whether the implicit division of labor between the production and consumption of roll call estimates obscures consequential connections between how the matrix of roll calls is analyzed to produce estimates of elite preferences and how those estimates are used in subsequent analyses to evaluate models of the political process. Because every parameter estimated from roll calls depends on the assumed model of individual choice (and perhaps also on how the model is statistically implemented), interpreting the roll call estimates requires understanding the consequences of the many modeling decisions that are made when analyzing roll call votes (Clinton 2012). Measurement issues regarding the latent traits are swallowed up by theories underlying the data generating process. Testing theories of partisan effects in Congress (see Rohde 1991; Cox & McCubbins 1993; Krehbiel 1998; Chiou & Rothenberg 2003; Cox & McCubbins 2005; Clinton 2007, among many others) can be problematic in that the assumptions built into the measurement of the latent trait bias results.

Latent Trait and Spatial Model Theory and Assumptions

We now turn to exactly those sets of assumptions that underlie both the spatial theory of voting and the measurement theory of latent trait models, specifically for IRTs (but generally true for all measures of underlying latent abilities).There is substantial overlap in the measurement theory underlying latent traits and spatial models, and that the areas in which problems arise bias both the estimation stage of fitting the models and in interpreting them or using those scores to test theories.

What are the key assumptions behind a singular latent trait model? Assuming that it is a reflective construct—which all ideal points are by definition—then there are three assumed consistencies: the model is constructed with both item and group invariance and there is a consistency in directionality of the latent trait and the parameterization of the underlying indicators (or items for an IRT) (Stenner *et al.* 2008).[2]

The first major assumption of an IRT model is item invariance. For the estimation of a consistent trait, the estimation of the latent trait must be independent of the items selected; a consistent measure of the latent trait is only possible if subsections of the same items produce essentially the same measure (Baker 2001; Baker & Kim 2004; De Boeck & Wilson 2004b; van der Linden & Hambleton 2013). What that means in practice is that there cannot exist a systematic way in which you can divide the "questions" in an exam that will give substantially different scores. If that happens, it is indicative of an implicit bias in question choice and would indicate rewriting the questions or separating the test into multiple different instruments. This property of test design is why most standardized tests operate on a test bank of questions; knowing the relative difficulty of the questions, all we have to do to get consistent scoring estimates for test taker ability is select any random subsample and the scores will be consistent (measured with error of course). Questions that systematically bias results—ones that only in conjunction with each other produce different difficult or discrimination scores—are to be avoided and dropped. How one should think about this for roll calls is that estimation of a consistent ideal point, where it to be consistent with the underlying logic of an IRT, should be invariant to the types of votes we subsample from, whether that is by vote type, policy area, etc. An item-invariant measure

[2]We can think of constructs as being innately either reflective or formative. Reflective traits exist independent of how they are measured. The components of the trait are caused by the construct itself — roll call votes are caused by MC's spatial preferences. The canonical reflective trait would be intelligence; we do not believe that test questions on an intelligence test make up intelligence uniquely, but instead reflect the test takers underlying intelligence and their responses (measured with some error) indicate their latent ability. Formative measures are measures that are formed by multiple different component that cause the construct. For formative measures, the canonical example is socioeconomic status, which exists only insofar as it is a combination of education, wealth, and occupational prestige. It is not that these elements of the construct are caused by SES, but rather that SES is the combined influence of all of these components (Stenner *et al.* 2008).

would produce substantially similar results—in particular, they would be of similar rank ordering—regardless of the subsample of data used.

The next assumption about the stability of latent traits would be group invariance: the distribution of the item parameters do not differ when comparing different groups of test takers. This is another straightforward test: do the estimation of the item parameters, which are essential to linking questions between tests, vary when dividing the test takers based on group characteristics (Baker 2001; van der Linden & Hambleton 2013). This is evidence of implicit bias in the construction of the test: if you know the relative latent ability of a test taker, they should consistently get questions correct based on the known properties of the test question. If you have questions that high skill test takers always get wrong if they are of a certain racial group, for example, then there is evidence that the questions themselves are pulling from something beyond that latent trait *for those members who perform worse* and that secondary dimension is driving performance (Baker & Kim 2004; De Boeck & Wilson 2004b). For roll calls, it is worthwhile to see if certain types of votes get different distributions of item parameters when one party is in the majority versus another. It is also a consideration to think through how the different types of votes map onto the latent trait when comparing majority and minority party results. Inconsistencies between question types by majority status could be indicative of problems of group invariance.

The final assumption is that the item parameters will have measurement consistency: unidimensional models requires α directional consistency. Essentially, IRT models require consistency in the estimation of their item parameters to have stable estimation of the latent trait. This is a discussed more thoroughly in Lerner *et al.* (2021), but serious miss-estimation of the latent trait can occur if we have individual items that are inversely predicted by that latent trait. There is no scenario under which, in canonical item response theory modeling, that a person would consistently be more likely to get a question wrong if they have greater ability, if there was an inverse relationship between the probability of a correct answer on any given question and an overall increase in their latent ability. If this relationship comes up, it can be assumed that it is either because of a poorly written question, or the addition of an extra latent dimension driving answering the question that is orthogonal to the latent trait being estimated. One can think of these three assumptions as being an indirect way of getting at the primary dimension: *all items in an IRT should be IID, as should the latent trait of any test taker, expect thought the latent trait itself.*

On the other hand, we also have a set of assumptions from the spatial theory of voting that are rather important to understanding how these intersect, and why this assumption of IID items might be too difficult given the different types of voting in Congress. When we think of any given vote in Congress, there are several contexts that make voting a more complex environment. For establishing consistency in a spatial voting model, we need to make assumptions about the underlying agenda model, utility structure, indifference curve, tie-breaking rule, decisiveness rule, and the status quo rule. Each of these components of the spatial voting model need to be the same for the votes to mean the same exact thing; different tie-breaking procedures or status quo reversion rules, for example, would make comparing the votes very challenging, and would require an entirely different set of assumptions to be made about what members of Congress are considering when faced with these votes. It is not wrong to think of this as the data generating process of votes, and that consistency along the data generating process for votes (and the response we would expect from members of Congress in response) would be analogous to IID assumptions necessary for consistency of latent trait modeling.[3]

Problems arise for voting when we think about the different types of votes in Congress–

[3] There are other assumptions that go beyond the scope of this paper that may be problematic, in particular dimensionality (Aldrich *et al.* 2014; Dougherty *et al.* 2014; Roberts *et al.* 2016), identification (Londregan 1999; Jackman 2001) and the impossibility of ascertaining status quo, given what we see (votes) we cannot uncover both legislator position and status quo/proposal position simultaneously. Given the general critiques of Ho & Quinn (2010) as well—written for courts but entirely applicable to legislatures as well—there are some general reservations we should have about the overall usage of ideal points in spatial models at all. Going through these assumptions and correcting for what we can, seems like an appropriate starting point. Relatedly, there is some concern about the instability of ideal points given that additional votes come in non-randomly (Carrubba *et al.* 2008; Tahk 2010; Lynch & Madonna 2013; Lynch *et al.* 2016), as well as concerns about the underlying model of decision making the ideal point measure is supposed to capture a choice or skill (Poole 2005) or the shape of the utility function in general (Clinton & Jackman 2009; Carroll *et al.* 2013).

procedural, amendments and final passage votes–in terms of agenda structure, status quo reversion, and the structure of a given members utility. While discussed more directly in the following section, it is hard to imagine the same data generating process would produce all three sets of votes—and indeed, there is almost certainly deep variations in the process for different types of procedural votes and for amendments as well. What makes these elements all constant is looking at final passage votes alone.[4]

Splitting the Roll Call by Final Passage Votes

Now that I have established what a latent trait (in particular a measure based on an IRT requires), the next step is to apply what was learned from this to estimating ideal points from roll calls. In particular, I am going to evaluate the consistency of the item-parameters for two separate models: a Bayesian IRT ideal point model on final passage votes and the same model run over all other votes. Then I'm going to compare the resulting ideal points estimated from these models and show how these compare with canonical DW-NOMINATE scores over the same period. Differences that emerge between the relative scores will be explained entirely by partisan forces, indicating that final passage based ideal points are more consistent with a partisan theories of agenda control.

Note that this is not the first paper to look at a vote split as a point of contrast. Roberts *et al.* (2016); Crespin & Rohde (2010) and Jessee & Theriault (2014) look at the impact of splitting up the roll call in various ways on the estimation of ideal points. Neither paper, however, takes the criticism far enough; it is sufficient to show that there is no score stability between the splits, which calls into question the underlying measurement strategy. Even though Jessee & Theriault (2014) also look at the final passage vote as a meaningful split (and concludes that the ideal points differ when looking at procedural votes vs. final

[4] This should bring together consistency in theory and bring together consistency empirically as well. For ideal points to be comparable across measures and robust to alternative specifications it has to have constancy in its spatial framework and stability within it's performance for psychometric tests. These are simultaneously *theoretical* and *empirical* propositions.

passage votes), they do not use this opportunity to evaluate the consistency of the new ideal points, they do not compare the underlying data-generating processes and prefer one over another and, because they use W-NOMINATE as their estimation approach, they innately cap the scoring range of the ideal point set at $[-1, 1]$, which makes comparisons between each Congress impossible. This point will be addressed more directly later in the paper, but an unbounded IRT method should provide more direct comparisons year-to-year, so long as the priors are linked and stable.[5]

The motivation for this paper is in many ways an extension of Jessee & Theriault (2014) but committing to the modeling strategy of final passage votes entirely, and deriving a specific IRT method around doing this explicitly. The non-final passage votes (including all procedural votes) are subject to kinds of strategic voting that final passage votes are not; it is not unreasonable to conclude that the public facing nature of final passage votes at least presents the member of Congress the most transparent vote choice (Carrubba *et al.* 2008; Crisp & Driscoll 2012), especially for the majority party since they can assume bills put to final passage vote will pass (Cox & McCubbins 2005). Though the selection of final passage votes will have an implicit bias (see Carrubba *et al.* 2008, for an example), it will be a *consistent* bias, whereas procedural and amendment votes introduce a variety of biases that are harder to track down. Indeed, there is some evidence from Lynch and Madonna (2019) that the increase in polarization is partially driven by the increase in (mostly meaningless) procedural or symbolic votes that do not change the substance of law. At least the outcome of a given final passage vote will always directly reflect the lawmaking preferences of members voting because all passed bills can, and often will, change the policy status quo.

Importantly, we cannot think of procedural votes and amending votes as ever being independent or identically distributed with final passage votes given the nature of how these votes *structure* final passage votes themselves, literally in the case of some procedural votes (voting on the Rules), but metaphorically for the rest, and indeed, amending votes literally

[5](see Jackman 2001; Martin & Quinn 2002; Clinton *et al.* 2004, for a discussion of this point on linking priors and other ways of smoothing dynamic ideal point estimates)

change the substance of the bill being considered by future final passage votes. Considering the extensive work on procedural manipulation and how the use of these tools have changed over time in response to changing partisan conditions (Roberts 2007; Lynch & Madonna 2013; Dougherty *et al.* 2014; Lynch *et al.* 2016) and on the multi-faceted incentive structure for amendment votes (see the sophisticated voting literature on this point), we can be sure that, whatever the utility space of voting on procedural and amendment votes may be, there is no way it is the same as it would be for final passage votes.

Final passage votes remove four distinct sources of inconsistency in establishing spatial voting game: agenda model, tie breaking rule, decisiveness rule, and the status quo rule. We know that the agenda setting model for final passage votes is consistent; this goes back to Romer & Rosenthal (1978), but is elaborated on extensively in Cox & McCubbins (1993). The rules for tie breaking and decisiveness are also unchanging because, again, this is all established beforehand by the rules committee. Final passage votes all follow the same sets of procedures, which other sets of votes (amendments and procedural votes) do not. We also have consistent expectations with regards to the status quo, in that failure of the final passage votes means preservation of the status quo by definition. Final passage votes also have the benefit of being the "public facing" votes, in that it is an unambiguous signal the legislator has about the given proposal (Theriault 2008). The layers of complexity in dealing with amendments and procedural votes means that it is never going to be clear to the public what is going on with an particular vote; final passage votes are what most people would expect voting in Congress to be.

Part of what will tell us that the final passage votes might be better for giving a more comprehensive look than non-final passage votes will be to look at the distribution of vote margins. Returning to the theories of IRT models, we should have better and more precise coverage when outcomes are distributed more evenly along the latent trait; i.e. when there is greater coverage in terms of potential outcomes for assessments (Baker & Kim 2004; van der Linden & Hambleton 2013). For the testing literature, that usually means having more

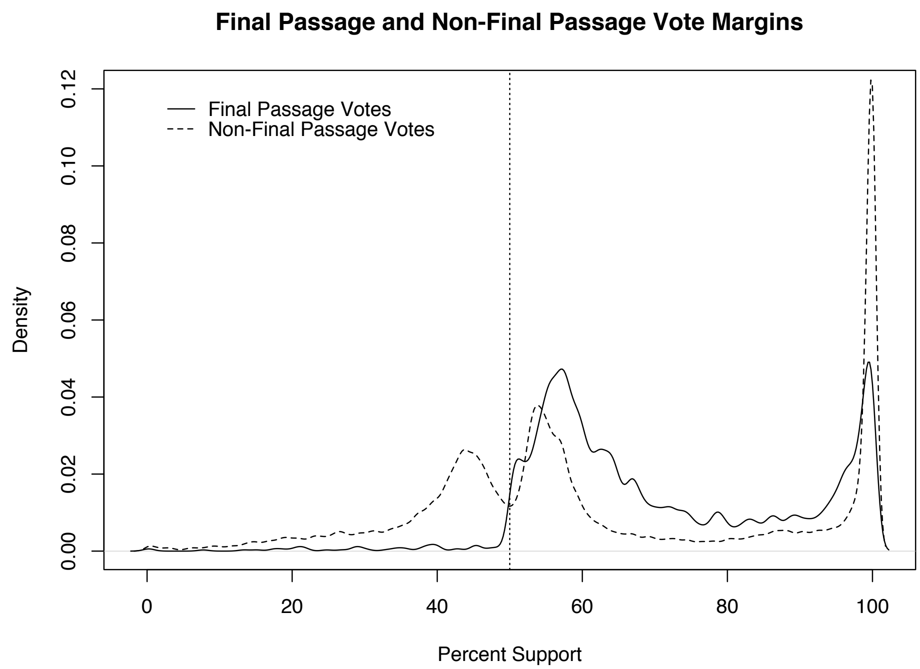


Figure 1: Final Passage and Non-Final Passage Vote Margins

questions that cover a range of topics and difficulties. For ideal points, the metaphor holds up in terms of policy areas (something that has been looked into already in Roberts *et al.* (2016)) and in terms of overall vote margins.

Figure 1 looks at the latter part directly by comparing the percent support for all votes across the range of outcomes. What we see is that non-final passage votes are tri-modal: proposals that barely lose, proposals that barely win, or unanimous votes. This will give us really strong information around these margins—presumably right around the majority-minority party lines—but rather poor spacing of information outside of that. For final passage votes, we find more coverage along the winning space, but virtually no defeats (because final passage votes rarely fail); this should give more information away from the marginal members and give more reasonable coverage along the upper end of the winning spectrum, but may not provide as much clarity in separating the parties. Suffice it to say, there is ample expectations that the results of comparing these results will be different at least, and there should be systemic differences between the two sets of ideal points.

What Characteristics do Latent Trait Models Need to be True?

Item response theory methods are a class of latent trait models developed in education testing to capture test taker aptitude along with an assessment of a questionnaire (i.e. test or exam) van der Linden & Hambleton (2013). Item response models provide the basis for most modern standardized tests and represent improvements made to the education testing literature from classical test theory over the last seven decades. The broad adoption of IRT models in psychometrics and education testing is primarily due to two important features of this class of models: they are highly flexible when it comes to specifications, and they model the difficulty and utility of any given test question which allows educators and employers to choose test questions that capture the latent aptitude they want to measure. This means that, unlike in classical test theory, tests constructed using an IRT framework have an adaptive approach to question weighting and forces all questions to be evaluated on their ability to discriminate between test takers based on their latent aptitude or skill (De Boeck & Wilson 2004a,b; Fox 2010; van der Linden & Hambleton 2013).

In education testing and psychometrics, IRT models are utilized primarily for the ease with which they fit what is thought of as the "skill" parameter for test takers. This framework creates some problems given how it is used in social science and legal research, because the models are not designed to scale infinitely (see Tahk 2018, for a non-parametric solution to this problem). While studying the consistency of ideal point models, Londregan (1999) linked the psychometrics testing literature with the spatial theory of legislative voting in order to derive important statistical insights. In particular, Londregan (1999) demonstrates that when the preferential choices are nominal (as in voting 'yes', 'no', 'partially yes' or 'partially no') consistency of ideal points in its usual statistical sense is not possible. With nominal choices, maximum likelihood estimators that attempt simultaneously to recover legislators' ideal points and roll call or item parameters must inherit the granularity of the choice data and so cannot capture the underlying continuous parameter space. Londregan (1999) solves this identification problem in IRT models by using constraints on underlying ideal points of certain legislators, which induces additional variability to all spatial IRT models. This added constraint provides additional subjectivity to the analysis of IRT models: namely, that the estimates generated by even the best possible model will ultimately be contingent on certain arbitrary elements that make specific point estimations inconsistent.

Ho & Quinn (2010) argue that the ordering that comes from these ideal point measures, by virtue of being consistent, should be the primary metric used by practitioners of ideal points, rather than the starting point. The specific location along any ideological line is compromised by the specifics of the subjective aspects of any modeling choices made. Ho & Quinn (2010) go on to argue that "the cardinal scale of the latent dimension is not identified from the data. In the standardized testing analogy, we have no sense of whether the 100-point difference between a score of 2100 and 2000 should be the same as the difference between 2300 and 2400. While prior assumptions about cutpoints affect such cardinal scaling, they generally do not affect the relative ranks. (pg. 847)" It is clear, then that we cannot and should not believe that we can generate ratio-level output consistently and without greater concerns about bias from a model that has nominal-value inputs.

Traditional IRT models simultaneously model a test-taker "skill parameter" θ and one or more item parameters. Most basic IRT models include only a single dimension for the item parameters (α) and are referred to as either 1-PL IRT's or Rasch models. Most commonly in political science, work with IRTs utilizes some form of a modified 2-PL model, which differs from the 1 PL model by having one parameter for each respondent and two for each item. The standard 2-PL model is specified as:

$$P(\theta) = \frac{1}{1 + e^{-\alpha(\theta - \beta)}}$$

with α and β being the parameters for item-discrimination and item-difficulty respectively, and θ being the parameter for latent trait level (ability or in the case of ideal point models, the ideal point of a given MC). It is also generally fit with a probit or logit link.

Given the basic IRT framework, it is no wonder use of IRTs as a method of modeling is very popular within the ideal points estimation literature in political science. If you have a way of effectively capturing distributions of the parameters, like an MCMC, then this model provides a very flexible framework that has many attractive qualities to scholars of latent trait modeling. It is not surprising that most of the attention paid to these models is focused on the θ dimension and distributional effects of the θ are studied intensely. One overlooked aspect, though, is the behavior of the item parameters.

The item difficulty parameter, denoted as β above, is defined as the point on the scale of ability (θ) at which the probability of correct response to a given item is $P(\theta) = 0.5$.[6] More simply, the difficulty parameter (β) indicates the level of ability (θ) a test-taker needs in order to a has a 50% chance of answering the question correctly. The discrimination parameter, α, is a scalar multiple of the slope of the curve (in the item characteristic curve or ICC) where $\theta = \beta$ and $\mathrm{P}(\theta) = 0.5$. Items with discrimination parameters that have high values are better at differentiating among individuals as those with lower ability are more likely to answer incorrectly than those with higher ability.

In general, political scientists have mostly focused using α and β instrumentally to calculate the latent trait θ, as this parameter is what most directly applies to measurement models of ideal points. Most work on IRT models in political science disregards the α and the β parameters. Further discussion of this point can be found in Lerner *et al.* (2021). But what we should take away from this is that the behavior of these parameters has direct implications for how internally consistent our models are, and how well specified we think the underlying data generating process is. If we know there are serious discrepancies in how the α values behave, then we can suspect that driving this deviation is a problem with the items themselves—clear evidence of a violation of the IID assumptions discussed above.

[6] Theoretically, this parameter can range from $-\infty$ to ∞. However, this parameter typically ranges from -3 to 3 (Baker and Kim 2004, pg. 18). Also, latent ability is set to have a mean of zero and a standard deviation equal to one.

Theoretically, splitting the votes by vote type could alleviate that concern, if the different underlying assumptions of the preference structure are driving the α inconsistencies.

Data and Methods

For this project, I use roll calls from the 103^{rd} Congress through the 115^{th}. I downloaded the roll calls from VoteView.com and separated them if they were final passage votes or not. To engage in estimation of ideal points, I used the standard Bayesian IRT model built on Clinton *et al.* (2004) with simple properties and with priors set as member scores from the previous cycle (save for the 103^{rd} Congress or new members, where I use a flat prior by party).[7]

A Bayesian IRT built off of IDEAL was chosen because "[i]t stands alone in not having otherwise chronic artificial extremism or polarization bias; it strictly dominates other measures in terms of mean squared errors and mean absolute deviations; in addition (not heretofore discussed), it can also be estimated alongside other covariates and flexibly incorporate information on the legislative agenda." (Krehbiel & Peskowitz 2015, page 694) A Bayesian IRT should be biased against finding stronger party effects than using NOMINATE or another method. Given the diagnostics we can use with IRTs as well (that probably exist for NOMINATE as well, but are not as well defined), it seems clear why we would choose the Bayesian IRT to build our measure from.

One of the key assumption of IRT models is that the item parameter α is monotonic: all the α values must be in the same direction. Without a unidirectional α, making comparisons between items is impossible in that it cannot "discriminate" between the various skill levels. Now, it is common for various IRTs to have different levels of α values–indeed, in another paper, my coauthors and I look at this exact problem and propose a solution for dealing with this in judicial ideal points (Lerner *et al.* 2021). But for this paper, it is sufficient to

[7]Note that this choice is largely irrelevant. Bridging between observations should bias the results to move less over time, but the correlation between runs with the flat party priors and the dynamic priors was consistently around 98%-99%

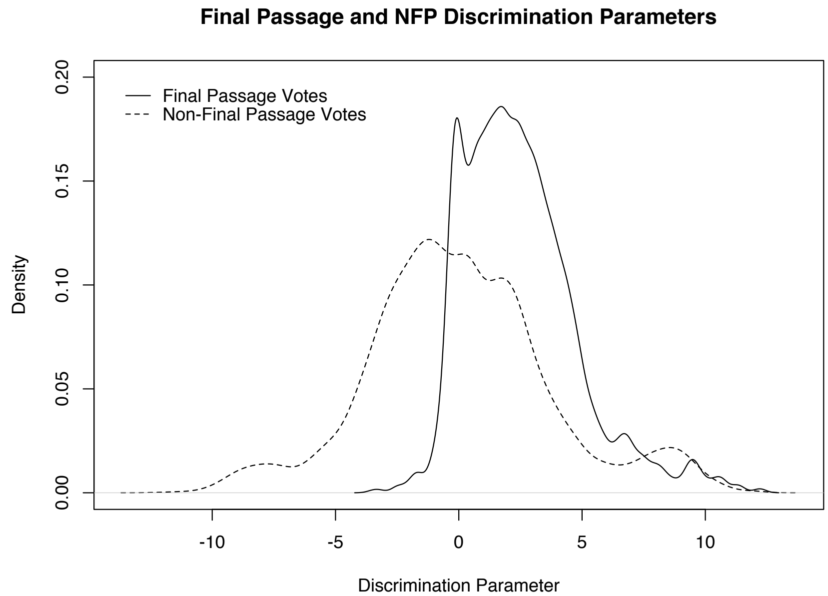


Figure 2: Final Passage and Non-Final Passage Discrimination Parameter Distributions

say that having too many αs in multiple directions is only problematic for identification and consistency within an IRT model.

Figure 2 shows the distribution of the discrimination parameters for each congress by final passage and non-final passage votes (final passage votes are solid lines while non-final passage votes are in dotted lines). What we see is a pretty consistent picture: for non-final passage votes, there is almost always a nearly even distribution in direction of the α values, making consistent estimation harder. But for final passage votes, we see something else entirely: a consistency within Congress of the directionality. This suggests that the singular dimension we are extracting from the roll call data is the correct one more often over final passage votes than over non-final passage votes. So based on the simplest of tests of latent trait models, we find that the final passage based model is more consistent with a unidimensional measure.[8] Given what was laid out in the previous section on IRTs and the agenda consistency of final passage votes, it is unsurprising then the IRT fit on final passage

[8]It is clear that both models before rather well as predictors and as spatial estimates of voting preferences. The main driver of this additional clarity from the item parameters should indicate that these votes have a greater inherent consistency as measures, but it does not disqualify the other set.

votes would be more internally consistent than the one built on non-final passage votes.

Expectations

Given that the α values are much more consistent with final passage based ideal points and that the final passage votes present a more homogeneous agenda and would apply a more uniform utility structure to the MCs voting on them, we should expect that ideal points based on final passage votes and non-final passage votes will differ along several key dimensions. If this is true, we will always have issues using ideal points that are generated with both sets of votes included because it will, by definition, violate assumptions about item and group invariance.

Given that the changes in the structure of voting for members on final passage votes will be greatest for those in the majority, I expect that the scores will differ the most for majority party members and not very much for minority party members. If there are no major differences by majority status, or if it seems as if the only differences are random, then there is little evidence for invariance problems and the use of measures that use both sets of roll calls are totally fine. Where this will manifest most directly is iin the rankings. Because of the logic of Ho & Quinn (2010), we would expect that even if there were substantial differences in the exact cardinal location of given members, the rankings should be consistent. This is a much bigger problem for the use of existing ideal points if the rank ordering differs substantially, because of the number of our theories that rely on pivotal members being well identified; inconsistencies along this dimension could complicate large portions of our literature.

My expectations are that ideal points estimated with final passage votes and non-final passage votes will differ significantly within the majority party, that this will permeate into rankings as well as raw spatial location, and that the non-final passage vote ideal points will end up looking more like DW-NOMINATE and other methods that use all votes than the final passage vote based measure. This all should give pause in how we have been thinking of spatial models of voting in light of these concerns.

Results

We now begin comparisons between the final passage, non-final passage ideal points, and DW-NOMINATE. For this section, each results are shown Congress by Congress to facilitate easy visual comparisons between when each party is in the majority party. For the Congresses covered here, the Democrats are in the majority for the 103$^{\text{rd}}$, 110$^{\text{th}}$, and 111$^{\text{th}}$. The Republicans are in the majority for the 104$^{\text{th}}$-109$^{\text{th}}$ and the 113$^{\text{th}}$-115$^{\text{th}}$. We begin with raw comparisons of ideal points. We then move onto comparisons of the rankings, where the discrepancies between the various methods are more visually striking.

Figure 3 shows differences between ideal points estimated only on final passage votes and ideal points estimated on all other votes. What is abundantly clear from the outset is two fold; first, the overall orientation is rather similar. This is not surprising, as the same underlying dimensionality should hold true regardless. Secondly, the similarity within parties varies a decent bit by whether or not the party is in the majority. For the minority, the within-party correlation between the two sets of scores ranges from 0.7-.0.85, not very bad at all. For the majority party, though, the correlation is much weaker, getting as low as 0.05 for the 115$^{\text{th}}$ Republican party. And, for a number of reasons, it seems like this gap is getting stronger over time. For each Congress, it is sufficient to look at this chart and immediately know which party is in charge; the minority party will have a very linear shape to its distribution whereas the majority party will have something of a triangular shape indicating sizable differences along this dimension.[9]

This becomes even more stark when comparisons are made between the general rankings derived from the scores as the raw scores. Since comparability and identification are always an issue with ideal points (see Londregan 1999; Ho & Quinn 2010, for detailed discussions of

[9] Comparisons between final passage ideal points and DW-NOMINATE are available in the appendix—non-final passage ideal points and DW-NOMINATE end up looking nearly identical, as seen in Appendix Figure XXX.

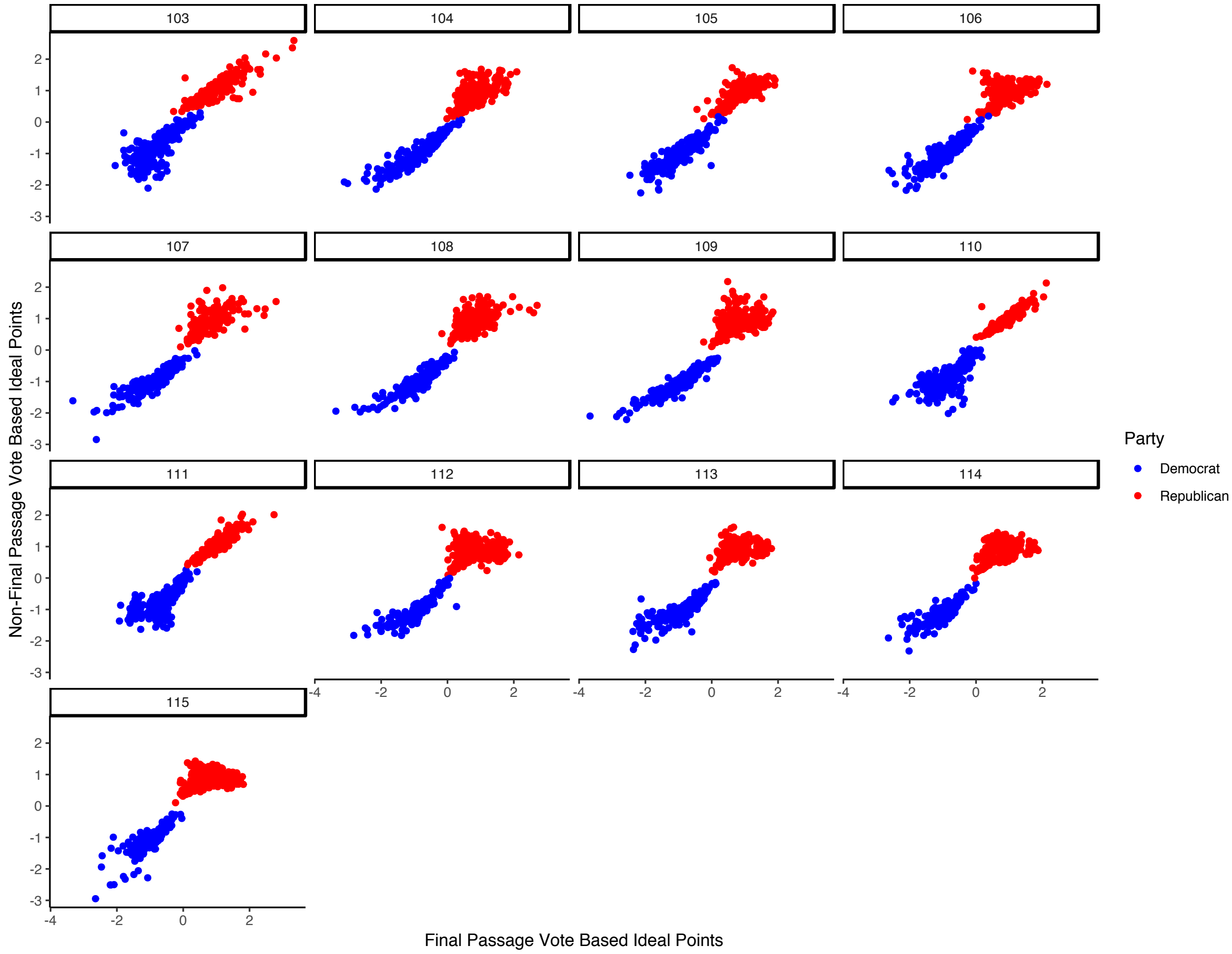


Figure 3: Final Passage and Non-Final Passage Ideal Points Over Time

these problems), we should not only base our conclusions about various distributive qualities of the ideal points from just their raw scores, but their rankings as well. Ho & Quinn (2010) go as far as to argue that only rank order consistency is possible consistently without controversy and that it is the essence of a stable measure.

Figure 4 provides a direct representation of this by comparing the distribution of the absolute value of the difference in Final Passage and Non-Final Passage Ideal Point rankings by majority party status. Note that the median difference in rank goes up over time–consistent with increasing party homogeneity and, therefore conditions of conditional party government (CPG) (Aldrich & Rohde 2000)–so that by the 115$^{\text{th}}$ Congress, the median difference equals 65 position difference in rank *for the majority party alone.* Given that the

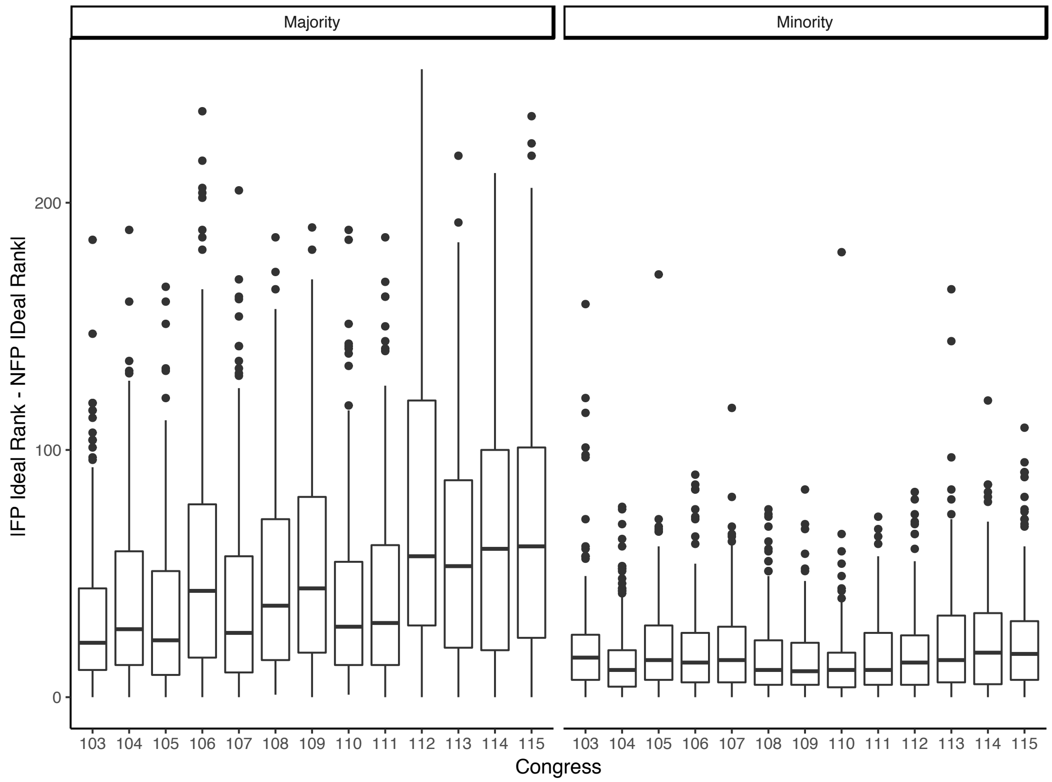


Figure 4: Absolute Value of Difference in Final Passage and Non-Final Passage Ideal Point Rankings By Majority Status

average size of the majority coalition over this time is 241.6 members, this means that the average difference is 26.9% of the total. Displacement amongst the majority party ranking over a quarter of the total range. For the minority party, the average rank difference is about 10. Since the average minority party coalition has 202 members, that means the average difference is only 4.9% of the total, a marked difference.

We then formalize this test with a series of regressions. Table 1 shows four different versions of this model. Columns 1 and 2 have the absolute value of the raw difference in the ideal point as the dependent variable, whereas columns 3-4 have the absolute value of the difference in ranks as the outcome of interest. All four models include Congress fixed effects. Columns 1 and 3 utilize the strictest test, with member fixed effects and clustered standard errors on the MC. This means that all the variance in the DV have to be from the only variable in the model: majority party status. Columns 2-4 relax the member

Table 1: Predicting Differences Between Different Scoring Methods

	Dependent variable:			
	|Ideal FP - Ideal NFP|		|Rank difference FP-NFP|	
	(1)	(2)	(3)	(4)
Majority	0.1035**	0.088**	27.985**	26.685**
	(0.007)	(0.007)	(1.088)	(1.162)
DWNom D1		0.251**		42.566**
		(0.042)		(6.355)
DWNom D2		−0.113**		−12.228**
		(0.016)		(2.389)
Republican		−0.211**		−26.689**
		(0.031)		(5.166)
Member FE's	Yes	No	Yes	No
Congress FE's	Yes	Yes	Yes	Yes
Member Clustered SEs	No	Yes	No	Yes
Observations	5,772	5,772	5,772	5,772
R^2	0.410	0.077	0.509	0.223
Note:				*p<0.5; **p<0.01

fixed effects assumption and instead just rely on clustering on members as the control for individual heterogeneity in scores. We also include 1st and 2nd Dimension DW-Nominate scores (to condition on middle of the distribution being less effected) and add in a dummy for Republican to account for possible partisan asymmetries. These are all additional variables added to try and dilute the possible majority party effect, accounting for the heterogeneity in member variability that would otherwise have been conditioned out in the fixed effects model.

There is a clear differences driven by majority party status that is stable even accounting for individual member differences. Consider the two fixed effects models, which both have majority status as the only variable of interest, yet explain 40% and 50% of the variation respectively (as seen by the R^2). The effect of being in the majority party drives 0.10 marginal scoring differential between the two methods for model 1, which translates to a difference of rank of 28 slots by model 3. This suggests that the entirety of the difference is driven by mis-estimation of the scores of majority party members in non-final passage voting situations. The good news from this set of results is that this agenda-bias will not show up much if you are interested in the minority party; there seems to be very little difference between pooling or separating scores by vote type for the minority. But for the majority, heterogeneities in the size and scope of the coalition are going to be missed if using non-final passage votes in estimating the ideal points. Further, the ways in which this impacts spatial models could be severe if it most directly effects members close to the majority party median (as shown in the previous figures). The consequences of this disruption will be explored in the following section, which applies these scores to an individualized spatial agenda setting model.

Testing Agenda Setting Theories with Final Passage Ideal Points

Given the problems with ideal points estimated on non-final passage votes as latent traits, what can be learned from final passage ideal points? Anything that relies on estimating the affect of agenda setting on outcomes will have to rely on ideal points estimated with some agenda setting model. If we are interested in how majority party blocking powers manifest at an individual level, spatial models of agenda control will need to be tested. Already, we can see that ideal points which account for the consistent agenda setting role in their estimation will get a more exact effect of spatial predictors of agenda control. To examine this more directly, I will address models of individual roll rates, a concept introduced in Cox & McCubbins (2005), but more fully explored in Carson *et al.* (2011); Cox & McCubbins (2014); Robinson (2015). I start with a discussion of what roll rates are, what the individual roll rate is, what the basic spatial models of rolls assumes, and why we should expect differences in how ideal points are measured to impact what the winning spatial coalitions look like. Using final passage ideal points, I unsurprisingly find strong evidence for the majority party's negative agenda setting powers. I also find that, for members of the majority party, being more extreme than the median of the majority does not predict a higher roll rate, which is indicative of the homogeneity of the majority's agenda. Because of this homogeneity amongst the majority party, ideal points estimated with final passage find stronger conditions for *positive* agenda setting powers than with non-final passage ideal points.

The Individual Roll Rate: Measuring Individual Agenda Powers

Final passage votes yield four distinct agenda setting outcomes: given a proposal passes, it is either a success if the agenda setting supports it or a roll if they oppose it. If it fails and the agenda setter supports it, it is a disappointment, and if they opposed it, it is a block (Cox & McCubbins 2005, 2014; Jenkins & Monroe 2016; Clarke *et al.* 2017). The advantage of grounding the results of these votes in terms of the agenda setter and the outcome is that it orients it as an expression of realized agenda control; success and blocks are examples of the majority party successfully exercising positive and negative agenda control respectively, whereas disappointments and rolls are examples of their respective failures.[10]

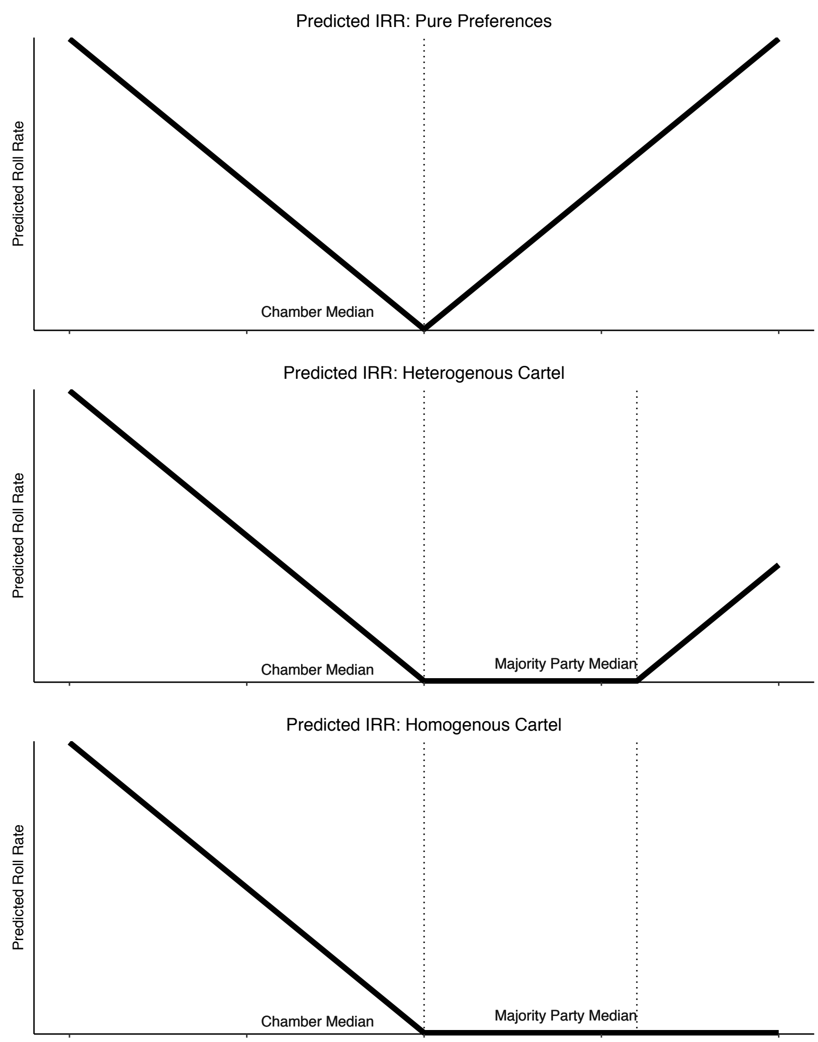
For this analysis, I'm going to restrict my analysis to rolls, primarily because it has both an aggregate form, most widely theorized in Cox & McCubbins (2005), as well as an individualized form. The aggregate performances of outcome models are unlikely to make a huge difference regardless of how the model is specified since there are so few successful rolls in Congress. Individualized rolls provide a more clear test of the spatial considerations of how measures of ideal points impact the testing of agenda setting theories. For this, we turn to the individual roll rate (Carson *et al.* 2011; Cox & McCubbins 2014; Robinson 2015).

The individual roll rate (IRR) measures the extent to which individual members are rolled: namely, the rate that individual members vote against bills that end up passing. The

[10] This is not the only way of thinking about who wins and who loses on roll calls in Congress, and there are several tests of theories of voting using roll calls that use different formulations (see Cox & Poole 2002; Lawrence *et al.* 2006; Hirsch 2011, among many others as examples). Agenda setting models are the most directly relevant test of the agenda-constant properties of the final passage ideal point; since most of these theories have spatial elements but are tested using measures of spatial preferences derived from votes with a variety of underlying data-generating processes, it cannot directly test the pure form of the ideal point underlying the theory. Since rolls and disappointments are rather rare events, studying the agenda powers of individual members of congress could be a fruitful area to extend this measurement.

nice thing about the IRR is that it has clear spatial predictions based on different theories of partisan competition. Figure 5 shows different spatial predictions made under different theories of agenda setting, adapted from Cox & McCubbins (2005); Carson *et al.* (2011); Cox & McCubbins (2014); Robinson (2015).

Figure 5: Predicted IRR for Pure Spatial Preferences, Cartel with Heterogeneous Preferences, and Cartel with Homogeneous Preferences



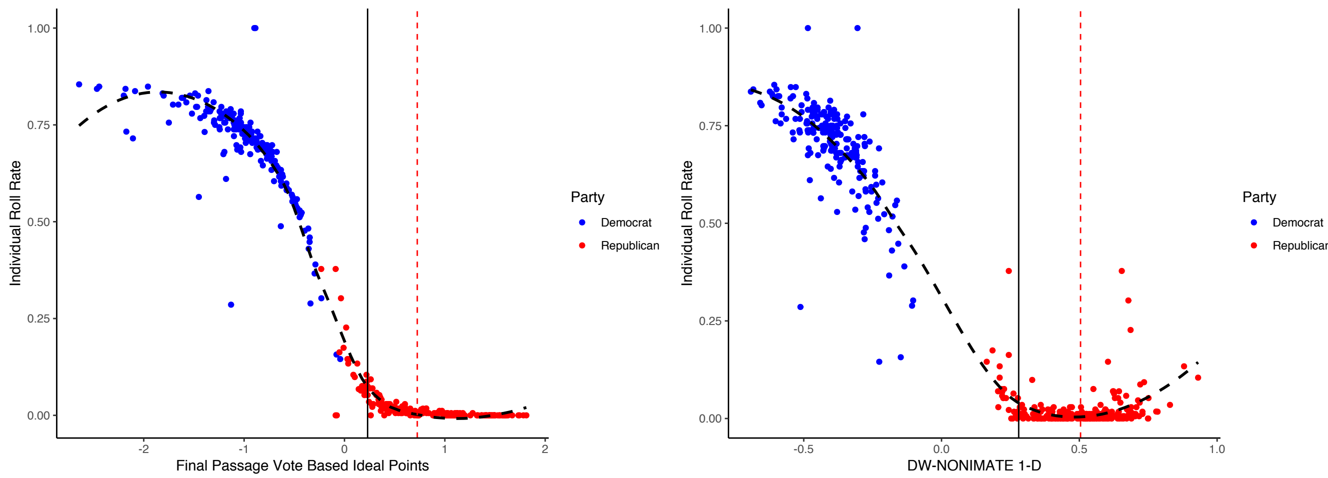
Following Krehbiel (1996) let us consider the pure preference–based theory of agenda control. The top plot from Figure 5 represents the policy space along the x-axis, where individual members would be positioned from most liberal (left) to most conservative (right), and individual's predicted roll rates along the y-axis. Assuming legislation is considered under an open rule, a pure preference-based model makes a straightforward prediction about individual roll rates over the policy space: the further an individual is from the floor median, the more often she should be rolled. This prediction is represented in the figure by the "v-shape" roll rate function. To see how this prediction is derived, consider the conditions under which an individual in the top figure would be rolled. Assume that individuals (with single-peaked symmetrical preferences) seek to maximize their own utility, which is defined by the distance between their preferred policy and the actual policy. For an individual to be rolled, she must prefer the status quo to the proposed bill. This implies that her ideal point must be farther from the floor median's ideal point than the status quo. The most obvious status quos that fit this description are those between the individual's ideal point and the floor median. As individuals get farther from the floor median, more status quos inhabit the space between their ideal points and that of the floor median. Accordingly, opportunities for these individuals to be rolled should increase. Thus, in this pure preference formulation, roll rates should increase the same on the right and left as the distance between legislators and the floor median increases (Krehbiel 1996; Carson *et al.* 2011; Robinson 2015).

Following the Cox & McCubbins (2005) "cartel model" we include one additional assumption: a majority party leader can veto any proposal he or she wishes. The party median, however, is assumed to veto bills on the same basis that she would vote against them. They would make no special consideration as a party leader and maximizes her utility like any other individual in the policy space. As in the pure preference model, a bill is proposed at the floor median's ideal point. The majority party median elects to veto the bill or let it come to a vote, thus exercising negative agenda control. If a vote occurs, everyone votes sincerely, as in the pure preference model. From Carson *et al.* (2011); Robinson (2015), we would expect that there is no status quo that can produce a roll of members between the floor median and the majority party median. This is not true for members outside that interval. The middle plot from Figure 5 represents the individuals' predicted roll rates under the cartel model. Like the pure preference model prediction, the more extreme members get, the higher their predicted roll rates. However, under the cartel model, roll rates do not increase on either side of the floor median, but instead increase on either side of the interval anchored by the floor median and the majority party median. All members within that interval have a predicted roll rate of zero. Carson *et al.* (2011) hypothesize that "individual roll rates should be flat with respect to ideological distance within the interval bounded by the majority and floor medians, roll rates should increase as individual legislators are further, in both directions, from this interval (page 22)."

To this, we add one wrinkle: the prediction from Carson *et al.* (2011); Cox & McCubbins (2014) and Robinson (2015) lead to the the middle plot from Figure 5, what has colloquially been referred to as the "swoosh". With that, it relies on their being clear spatial differences between members of the majority party as they get further from their party median in parallel to each other, as well as a clear distribution of status quos. If the actual distance between "extreme" members – those greater than the party median – and the rest of the party is less than the those who are less extreme than the "protected zone" members – those MCs between the chamber median and the majority party median – than we would see less a "swoosh" and more of a logistic form – the predictions from the lower plot from Figure 5. Evidence that there is little policy difference between the extreme members and the median members of the majority party would be a strong signal of the majority party's homogeneity, and would therefore imply the existence of positive agenda powers (Rohde 1991; Aldrich & Rohde 2000). Now, it is critical to note, that in this formulation we would still have the cartel derived from Cox & McCubbins (2005). We uncover the necessary majority party homogeneity required by conditional party government: finding the implications and residue of positive agenda power in a test of negative agenda power.

Consider Figure 6. The three panels show the IRR for the 115th Congress with different ideal points used as the x-axis: top left is final passage ideal points, top right is non-final

Figure 6: IRR for 115th Congress by Ideal Point



Solid line represents the median of the chamber. The dotted red line represents the median of the majority party. The black fitted line shown is the LOESS fit.

passage ideal points, and the bottom is DW-NOMINATE D1. We see in non-final passage ideal points and in DW-NOMINATE something that looks like the classic cartel described by Carson *et al.* (2011); Cox & McCubbins (2014) and Robinson (2015). The final passage ideal points, however, show a much flatter range from beyond the majority party median, and indicate a more homogeneous majority coalition, consistent with the bottom of Figure 5. To get a better grasp on this, we need a formal test of this model. Consistent with Carson *et al.* (2011); Cox & McCubbins (2014) and Robinson (2015), we now move onto modeling this relationship directly.

We now begin with the parametric tests from these models. Table 2 looks at predictors of individual member roll rates. Column 1 includes only a dummy for majority party and has no spatial component. Columns 2-4 and 5-7 have spatial components to the model; the only difference is the ideal point used for that inference. Columns 2-4 use the ideal points derived from final passage votes alone; columns 5-7 use all non-final passage votes.[11] Columns 2 and 5 include only the distance to the chamber median—the test of the pure spatial preferences model. Columns 3 and 6 interact the majority dummy with distance from the party median—

[11] An alternative version of this table with DW-NOMINATE scores is included in Appendix XXX.

where party median should be negative and the interactive term should be positive to be consistent with a cartel model. And columns 4-7 also interact the majority party dummy with distance to the chamber median—allowing for a differential response to the "swoosh". Both 3-4 and 6-7 would be consistent with either form of the cartel model. For each model, we fit a Bayesian Extended Beta Binomial model with random intercepts for both individual members and sessions of Congress.[12]

Table 2: Individual Roll Rates: Pure Preferences vs. Partisan Agenda Setting

	Dependent variable: Individual Roll Rate						
	Ideal Final Passage				Ideal Non-Final Passage		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Majority	-3.24^{**}		-3.39^{**}	-1.53^{**}		-3.81^{**}	-1.30^{**}
	(0.025)		(0.034)	(0.034)		(0.035)	(0.061)
Dist.C.Med		1.37^{**}		1.02^{**}	3.12^{**}		1.61^{**}
		(0.034)		(0.022)	(0.031)		(0.039)
Dist.P.Med			-0.333^{**}			-0.455^{**}	
			(0.037)			(0.051)	
Maj*Dist.P.Med			0.396^{**}			1.85^{**}	
			(0.068)			(0.082)	
Maj*Dist.C.Med				-1.92^{**}			-1.10^{**}
				(0.047)			(0.061)
Observations	5,772	5,772	5,772	5,772	5,772	5,772	5,772
Note:							*p<0.05 **p<0.01

Bayesian Extended Beta Binomial Regression with Congress and Member Fixed Effects

What we see in Table 2 is consistent with Robinson (2015); evidence for both a pure preferences model as well as a cartel model, though comparatively stronger results in favor of a cartel model. Indeed, the differences in significance and the magnitude of the effects are not such that we can definitively tell, though the final passage ideal models potentially hint at the presence of a more homogeneous majority party cartel than the non-final passage ideal

[12]Extended Beta Binomial is the preferred model to fit for roll rates (Cox & McCubbins 2005). Instead of fitting each model separately or comparing only Republican vs. Democratic controlled sessions like in Carson *et al.* (2011), we pool over all the sessions in our data. To deal with unobserved heterogeneities, I fit random intercepts for each member and Congress. Additional information about the model, as well as robustness checks using OLS with member and Congress fixed effects, are available in the Appendix XXX.

points. Indeed, final passage votes generally give larger coefficients to the party interacted spatial indicators than the non-final passage models, while estimating a smaller coefficient on the pure spatial effect (Models 2 and 5).

Since the inference from the previous model is unclear (evidence for both types of Cartel models), strictest test would be to limit ourselves to the majority party only. Some predictions follow:

- If MCs in the protected zone have an IRR $<$ MCs in the extreme zone than we have evidence of a heterogeneous cartel
- If MCs in the protected zone have an IRR $\geq$ members in the extreme zone than we have a homogeneous cartel
- If pure preferences model, distance from chamber median should be positive even if we only look at the majority party

We extend the same modeling strategy from Table 2 in Table 3, but focusing only on the majority party. It is clear from Carson *et al.* (2011); Cox & McCubbins (2014) and Robinson (2015) that the majority party is where we'd expect the biggest differences in comparing pure preferences to cartel models, and we also see from Table 1 that the differences in ideal points are greatest for the majority party. So this expectation seems like it would be a more reasonable direct test of this proposition. For each model in Table 3, we fit a Bayesian Extended Beta Binomial model with random intercepts for both individual members and sessions of Congress.

We can see with Table 3 that the spatial predictions hold much more strongly with final passage based ideal points than with non-final passage based ideal points. Both sets of ideal points allow for you to reject the pure preferences model (just comparing distance to the chamber median), but only the final passage ideal gives consistent evidence for a homogeneous cartel. Non-final passage ideal points only give weak evidence for the classic heterogeneous cartel. When looking only at the majority party, final passage ideal points actually find a negative relationship between roll rate and distance to the chamber median

Table 3: Individual Roll Rates: Majority Party Only

	Dependent variable: Individual Roll Rate			
	Ideal Point Measured with:			
	Final Passage Ideal		Non-Final Passage Ideal	
	(1)	(2)	(3)	(4)
Dist from Ch.Med	-0.789^{**}		-0.021	
	(0.042)		(0.058)	
Center		0.913^{**}		0.962^{**}
		(0.039)		(0.052)
Extreme Zone		-0.543^{**}		0.158^{**}
		(0.028)		(0.042)
Observations	3,141	3,141	3,141	3,141
Note:			*p<0.05 **p<0.01	

Bayesian Extended Beta Binomial Regression with Congress and Member Fixed Effects. Center is a dummy for an MC's ideal point less than the chamber median; Extreme is a dummy for if an MC's ideal point is greater than the party median. Reference Group is the Protected Zone.

(non-final passage ideal points finds a non-significant but negative effect instead). We are able to, with Modes 2 & 4 to test the "outside" in–the reference group for the three zones (center, protected, and extreme) is the protected zone. By classic cartel theory, we should see positive and statistically significant coefficients on both the center and the extreme categories, for pure preferences you would expect to see a negative coefficient on the center category and a positive on the extreme category. For a homogeneous cartel, however, you would expect the protected members to get rolled at the same rate or less than the extreme, so you would expect a negative coefficient on the Extreme category. What we see is that for final passage ideal points, essentially the homogeneous cartel laid bare: strongly positive coefficient on the center and strongly negative coefficient on the extreme. But for the non-final passage ideal points, we find a different story: consistent with the classic cartel, the coefficients on the center and the extreme are both positive and statistically significant.

We find weak to no evidence for the pure preferences model and we find reasonably strong evidence consistent with different types of legislative cartels, depending on how one measures spatial preferences. This does not get us around the counterfactual argument Robinson

(2015) makes, but it does get closer to a clearer test. It seems increasingly unlikely that the pure preferences model is driving these outcomes, especially when looking at ideal points estimated using only final passage votes.

Potentially, a way to think about the differences between the models is that the final passage ideal points are designed around estimating spatial preferences on votes that are not only consequential, but ideologically demonstrative, in a way that non-final passage votes may not be. Using only demonstrative votes to estimate spatial preferences, we find that individual roll rates are largely flat amongst those around the majority party median and above, consistent with the preferences of a largely homogeneous legislative cartel. This would be a downstream consequence of their being strong positive agenda setting powers because the legislative agenda they are pursuing *does not divide their caucus.* There is something telling that extreme members of the majority party in the most recent Congresses do not get rolled much at all.

Conclusion

In this paper, we have thought through how ideal points are estimated in Congress. Through a discussion of the assumptions of the spatial model of voting and the item response theory methods underlying estimation of ideal points, we've shown that ideal points estimated on final passage votes avoid adding in extraneous bad information, have consistent data-generating processes, and produce more consistent ideal points. After generating these scores, we then showed that methods that rely on non-final passage votes differ substantially with final passage ideal points when looking at the majority party only. This discrepancy persists regardless of who is in the majority, and that it drives both absolute differences and rank order differences. We then apply this measure to an existing set of tests about agenda setting in Congress, focusing on the individual roll rate of Carson *et al.* (2011); Cox & McCubbins (2014) and Robinson (2015). With different spatial orderings, we are able to use these to test theories of lawmaking in Congress. We find that substantial evidence for the negative agenda setting legislative cartel of Cox & McCubbins (2005). We also find indirect evidence of homogeneity amongst the majority party members that indicates the possibility of positive agenda setting, consistent with the conditions of conditional party government (Aldrich & Rohde 2000).

There are clear consistency improvements in using only final passage votes to estimate ideal points. This suggests that ideal point estimation is a process that may be improved by utilizing less data; sometimes more is not better. Since final passage votes are the most public facing votes, since they are consistent in how they come to be, and since they provide the most direct test of most lawmaking models, it is not unreasonable to at least privilege information learned from them over other votes. Potentially a case could be made to utilize both final passage and non-final passage votes in an informed way: this is something that could be achieved using structured priors over the votes and weighting the votes introduced in the model. Another approach would be to introduce forms of regularization/penalties to down-weight or remove uninformative votes (based on a combination of the outcome of the vote as well as information about the vote; potentially, not just vote type but policy area as well) and to up-weight informative votes. This builds on the insights of this paper, but instead of making the selection mechanistically by splitting the votes (as done here), vote weighting would be done algorithmically.

References

Aldrich, John H, & Rohde, David W. 2000. The consequences of party organization in the House: The role of the majority and minority parties in conditional party government. *Pages 31–72 of: Polarized politics: Congress and the president in a partisan era.*

Aldrich, John H, Montgomery, Jacob M, & Sparks, David B. 2014. Polarization and ideology:

Partisan sources of low dimensionality in scaled roll call analyses. *Political Analysis*, 435–456.

Baker, Frank B. 2001. *The basics of item response theory*. ERIC.

Baker, Frank B, & Kim, Seock-Ho. 2004. *Item response theory: Parameter estimation techniques*. CRC Press.

Carroll, Royce, Lewis, Jeffrey B, Lo, James, Poole, Keith T, & Rosenthal, Howard. 2013. The structure of utility in spatial models of voting. *American Journal of Political Science*, **57**(4), 1008–1028.

Carrubba, Clifford, Gabel, Matthew, & Hug, Simon. 2008. Legislative voting behavior, seen and unseen: A theory of roll-call vote selection. *Legislative Studies Quarterly*, **33**(4), 543–572.

Carson, Jamie L, Monroe, Nathan W, & Robinson, Gregory. 2011. Unpacking agenda control in congress: individual roll rates and the republican revolution. *Political Research Quarterly*, **64**(1), 17–30.

Chiou, Fang-Yi, & Rothenberg, Lawrence S. 2003. When pivotal politics meets partisan politics. *American Journal of Political Science*, **47**(3), 503–522.

Clarke, Andrew J, Jenkins, Jeffery A, & Monroe, Nathan W. 2017. From Rolls to Disappointments: Examining the Other Source of Majority Party Failure in Congress. *Political Research Quarterly*, **70**(1), 82–97.

Clinton, Joshua, Jackman, Simon, & Rivers, Douglas. 2004. The statistical analysis of roll call data. *American Political Science Review*, 355–370.

Clinton, Joshua D. 2007. Lawmaking and roll calls. *The Journal of Politics*, **69**(2), 457–469.

Clinton, Joshua D. 2012. Using roll call estimates to test models of politics. *Annual Review of Political Science*, **15**, 79–99.

Clinton, Joshua D, & Jackman, Simon. 2009. To simulate or NOMINATE? *Legislative Studies Quarterly*, **34**(4), 593–621.

Cox, Gary W, & McCubbins, Mathew D. 1993. *Legislative Leviathan: Party Government in the House*. Vol. 23. Univ of California Press.

Cox, Gary W, & McCubbins, Mathew D. 2005. *Setting the agenda: Responsible party government in the US House of Representatives*. Cambridge University Press.

Cox, Gary W, & McCubbins, Mathew D. 2014. *Individual Roll Rates*. University of California, San Diego.

Cox, Gary W, & Poole, Keith T. 2002. On measuring partisanship in roll-call voting: The US House of Representatives, 1877-1999. *American Journal of Political Science*, 477–489.

Crespin, Michael H, & Rohde, David W. 2010. Dimensions, issues, and bills: Appropriations voting on the House floor. *The Journal of Politics*, **72**(4), 976–989.

Crisp, Brian F, & Driscoll, Amanda. 2012. The strategic use of legislative voting procedures. *Legislative Studies Quarterly*, **37**(1), 67–97.

De Boeck, Paul, & Wilson, Mark. 2004a. *Explanatory item response models: A generalized linear and nonlinear approach*. Springer Science & Business Media.

De Boeck, Paul, & Wilson, Mark. 2004b. A framework for item response models. *Explanatory item response models*, 3–41.

Denzau, Arthur T, & Mackay, Robert J. 1983. Gatekeeping and monopoly power of committees: An analysis of sincere and sophisticated behavior. *American Journal of Political Science*, 740–761.

Dougherty, Keith L, Lynch, Michael S, & Madonna, Anthony J. 2014. Partisan agenda control and the dimensionality of congress. *American Politics Research*, **42**(4), 600–627.

Fox, Jean-Paul. 2010. *Bayesian item response modeling: Theory and applications*. Springer Science & Business Media.

Hirsch, Alexander V. 2011. Theory driven bias in ideal point estimates—A Monte Carlo study. *Political Analysis*, 87–102.

Ho, Daniel E, & Quinn, Kevin M. 2010. How not to lie with judicial votes: Misconceptions, measurement, and models. *California Law Review*, **98**(3), 813–876.

Jackman, Simon. 2001. Multidimensional analysis of roll call data via Bayesian simulation: Identification, estimation, inference, and model checking. *Political Analysis*, **9**(3), 227–241.

Jenkins, Jeffery A, & Monroe, Nathan W. 2016. On Measuring Legislative Agenda-Setting Power. *American Journal of Political Science*, **60**(1), 158–174.

Jessee, Stephen A, & Theriault, Sean M. 2014. The two faces of congressional roll-call voting. *Party Politics*, **20**(6), 836–848.

Krehbiel, Keith. 1996. Institutional and partisan sources of gridlock: A theory of divided and unified government. *Journal of Theoretical Politics*, **8**(1), 7–40.

Krehbiel, Keith. 1998. *Pivotal politics: A theory of US lawmaking*. University of Chicago Press.

Krehbiel, Keith, & Peskowitz, Zachary. 2015. Legislative organization and ideal-point bias. *Journal of Theoretical Politics*, **27**(4), 673–703.

Lawrence, Eric D, Maltzman, Forrest, & Smith, Steven S. 2006. Who wins? Party effects in legislative voting. *Legislative Studies Quarterly*, **31**(1), 33–69.

Lerner, Joshua Y, McCubbins, Mathew D, & Renberg, Kristen. 2021. The Efficacy of Measuring Judicial Ideal Points: The Mis-Analogy of IRTs. *International Review of Law and Economics*.

Londregan, John. 1999. Estimating legislators' preferred points. *Political Analysis*, **8**(1), 35–56.

Lynch, Michael S, & Madonna, Anthony J. 2013. Viva voce: Implications from the disappearing voice vote, 1865–1996. *Social Science Quarterly*, **94**(2), 530–550.

Lynch, Michael S, Madonna, Anthony J, & Roberts, Jason M. 2016. The Cost of Majority-Party Bias: Amending Activity under Structured Rules. *Legislative Studies Quarterly*, **41**(3), 633–655.

Martin, Andrew D, & Quinn, Kevin M. 2002. Dynamic ideal point estimation via Markov chain Monte Carlo for the US Supreme Court, 1953–1999. *Political analysis*, **10**(2), 134–153.

Poole, Keith T. 2005. *Spatial models of parliamentary voting*. Cambridge University Press.

Poole, Keith T, & Rosenthal, Howard. 1985. A spatial model for legislative roll call analysis. *American Journal of Political Science*, 357–384.

Poole, Keith T, & Rosenthal, Howard. 2000. *Congress: A political-economic history of roll call voting*. Oxford University Press on Demand.

Poole, Keith T, & Rosenthal, Howard L. 2011. *Ideology and congress*. Vol. 1. Transaction Publishers.

Roberts, Jason M. 2007. The Statistical Analysis of Roll-Call Data: A Cautionary Tale. *Legislative Studies Quarterly*, **32**(3), 341–360.

Roberts, Jason M, Smith, Steven S, & Haptonstahl, Stephen R. 2016. The dimensionality of congressional voting reconsidered. *American Politics Research*, **44**(5), 794–815.

Robinson, Gregory. 2015. Competing agendas in theories of Congress: Assessing agenda control using counterfactual data. *The Journal of Politics*, **77**(3), 749–761.

Rohde, David W. 1991. *Parties and Leaders in the Postreform House.* University of Chicago Press.

Romer, Thomas, & Rosenthal, Howard. 1978. Political resource allocation, controlled agendas, and the status quo. *Public choice*, **33**(4), 27–43.

Stenner, AJ, Burdick, DS, & Stone, MH. 2008. Formative and reflective models: Can a Rasch analysis tell the difference. *Rasch measurement transactions*, **22**(1), 1152–1153.

Tahk, Alexander. 2010. Properties of ideal-point estimators. *In: 27th Annual Summer Meeting of the Society for Political Methodology, at Iowa City.*

Tahk, Alexander. 2018. Nonparametric ideal-point estimation and inference. *Political Analysis*, **26**(2), 131.

Theriault, Sean M. 2008. *Party polarization in congress.* Cambridge University Press.

van der Linden, Wim J, & Hambleton, Ronald K. 2013. *Handbook of modern item response theory.* Springer Science & Business Media.