Kill(ed) Bills: How Agenda Control Affects Roll Call Ideal Point Estimates

Samuel Frederick* Diana Da In Lee†

Columbia University Columbia University

This version: September 15, 2023‡

Abstract

Statistical models that estimate spatial locations of individual decision-makers are a critical component of the scientific study of legislative politics. While ideal point estimates based on roll call voting have contributed significantly to the study of Congress, considering only those bills that reach the floor may generate results influenced by the majority party's agenda choices and over-emphasize intra-party homogeneity and inter-party polarization. We employ natural language processing and predictive machine learning models to impute preferences on *all* bills introduced between the 103rd and 116th Congresses and estimate the ideal points of members of Congress with these imputed preferences. Using these new ideal point estimates, we uncover various aspects of legislative behavior that were masked by the conventional roll-call-based ideal point estimation, including the multi-dimensionality of the policy space as well as intra-party conflict. We also show that roll call estimates of ideal points display more partisan polarization than our estimates using pre-floor bills. In sum, we find that agenda setting may lead to overestimates of partisan polarization and underestimates of intra-party heterogeneity in floor-based ideal points.

legislative behavior | agenda-setting | ideal point | bill text | superlearner

*PhD Candidate in the Department of Political Science, Columbia University (sdf2128@columbia.edu).
†PhD Candidate in the Department of Political Science, Columbia University (dl2860@columbia.edu).
‡Preliminary and incomplete. Please do not circulate.

1 Introduction

The estimation of ideal points using roll call votes has been among the most important developments in legislative studies as well as in studies of executive and judicial institutions. The development of NOMINATE by Poole and Rosenthal (1985, 1997), a measure that estimates legislator's preferred policy in low-dimensional Euclidean space, has led to a flourishing literature on the measurement of latent preferences. For example, alternative estimators of ideal points have been introduced, including Bayesian item response models for ideal point estimation (Clinton, Jackman and Rivers, 2004) as well as a generalized supervised learning methodology using campaign contributions (Bonica, 2018). Beyond legislative institutions, ideal point estimation has also been applied to the Supreme Court (Martin and Quinn, 2002; Bailey, Kamoie and Maltzman, 2005; Bailey, 2013), U.S. presidents (McCarty and Poole, 1995), bureaucratic agencies (Clinton et al., 2012), state legislatures (Shor and McCarty, 2011), as well as legislatures outside of the U.S. (Londregan, 2000; Voeten, 2000; Hix, Noury and Roland, 2006).

However, conventional measures of ideal points using floor votes are based on non-random subsets of the policy space, strategically selected by those who control the agenda. In Congress, particularly in the House of Representatives, the majority party selects agenda items to bolster its party brand and to tarnish the opposing party's image (Cox and McCubbins, 2005; Lee, 2016). As a result, the policies on which members of Congress vote tend to (1) emphasize partisan conflict (i.e., polarization) and (2) downplay intra-party disagreements, especially for the majority party. Thus, ideal point estimates based on roll call votes are potentially influenced by the majority party's agenda setting decisions toward finding greater partisan polarization and intra-party agreement. Moreover, partisan agenda control implies that, functionally, the subset of the policy space selected based on floor agenda is likely to be structured by a single dimension—potentially accounting for Poole and Rosenthal's (1997) striking findings of unidimensionality in congressional ideal point estimates.

Recent theoretical and methodological innovations have made it possible for scholars to rigorously study agenda setting in Congress. Cox, Kousser and McCubbins (2010) exploit variation in agenda control within two state legislatures to show that majority parties are rolled less and minority parties are rolled slightly more when the majority party has power over the agenda.[1] Napolio and Grose (2021) study the effects of an exogenous change in majority control in the Senate caused by deaths of several members within one Congress. They do so by estimating ideal points both before and after the member deaths using Item Response Theory (IRT) and find large shifts in ideal point estimates due to changes in majority control. Finally, Ballard (2022) uses natural language processing on congressional bill texts to impute member preferences for bills which do not reach the floor and finds a high degree of agenda control in both the House and Senate: whereas floor bills rarely result in majority party rolls, pre-floor bills, had they reached the floor, would likely have ended in more rolls of the majority party.

Our paper contributes to the growing body of research on the dynamics of agenda setting in legislative politics and to the well-established literature on ideal point estimation. We examine how ideal point estimates which account for *all* introduced bills—even those which never reach the floor—compare to standard ideal point measures based on roll call votes alone. We estimate the ideal points of individual representatives once using only pre-floor bills and again, using only floor bills to gauge the effects of agenda-setting on roll call measures of ideology. By developing a new measure of ideal points that is directly comparable to conventional measures using roll call data, we uncover various characteristics of congressional policymaking that were traditionally hidden behind the majority party's agenda control, including the multi-dimensionality of the policy space as well as intra- and inter-party conflict. Specifically, our estimates show a greater degree of ideological overlap between the parties and a large amount of heterogeneity within parties than do traditional,

[1]Party rolls refer to the passage of legislation against the wishes of a majority of a party's members.

floor-based estimates, such as NOMINATE.

2 Data and Procedures

To study the effects of agenda-setting powers on ideal point estimation, we collect the texts of all bills introduced in the House between the 103rd and 116th Congress from the official website of Congress as well as from govinfo.gov, a public website of the Government Publishing Office. Our analysis centers on the House of Representatives because (1) much of the literature about agenda-setting pertains to the House (Cox and McCubbins, 2005; but see Ballard, 2022 and Napolio and Grose, 2021); and (2) our analysis relies upon predictive models of final-passage votes, and the scarcity of such votes in the Senate would require extrapolating from limited samples.

Furthermore, whether bills make it from the Senate to the House is subject to many factors beyond the control of House party leadership, including individual senators, filibusters, and divided government (Binder, 1997). Therefore we focus our attention on bills originating in the House itself to more directly capture agenda-setting powers. Consequently, our sample includes a total of 135,422 bills, joint resolutions, concurrent resolutions, and simple resolutions introduced in the House over the nearly three decades.[2]

Our procedure is as follows. First, using the numerical representation of policy contexts in all bills, we train a prediction model to impute unobserved "votes" on pre-floor bills. We then fit unidimensional dyanmic IRT models to generate ideal point estimates separately for pre-floor and floor bills. Lastly, we compare the dimensionality, intra-party homogeneity, and partisan polarization between the pre-floor-only ideal point estimates against the floor-only estimates.

[2]Summary statistics are displayed in Supplementary Material Section A.

3 Estimating Voting Preferences

Ideal point estimation using all bills requires the estimation of members' votes on bills that never reached the floor. To do so, we first transform the text of each bill to a numeric vector. We then use these vectors as covariates to train predictive models of final passage votes and predict preferences on pre-floor bills using these models.

3.1 Numeric Representation of Bill Texts

Following Ballard (2022), we transform each bill text to a vector form using Doc2Vec (Mikolov et al., 2013; Le and Mikolov, 2014). Doc2Vec is a machine learning model that uses a shallow neural network to optimize predictions about words based on the context in which they appear and has been shown to out-perform predictions based on simple bag-of-words representations. There are two versions of Doc2Vec: distributed memory and distributed bag of words. The distributed memory algorithm trains word and document vectors simultaneously by sampling a document and a group of words within a document, and it attempts to predict the word following the group using the paragraph and word vectors (Le and Mikolov, 2014). On the other hand, distributed bag of words updates vectors to predict randomly sampled words from each document. As the original authors recommend combining the vectors produced by both versions (Ballard, 2022; Le and Mikolov, 2014), we train the two Doc2Vec algorithms using the text of all bills and combine the output.[3]

3.2 Predicting Votes on Pre-Floor Bills

As we do not have measures of the preferences of members of Congress on all bills, we must impute preferences for bills which do not reach the floor. To do so, we train an ensemble of models on floor votes using the document vectors from Doc2Vec along with other

[3]We use `gensim` to generate the combined vectors for each bill text in our dataset (Řehůřek and Sojka, 2010).

member- and bill-level characteristics. We then predict member preferences on pre-floor bills with our trained models.

The underlying assumption is that, regardless of bill status, members' voting decisions are primarily determined by the textual content of bills (as captured by the Doc2Vec weights) as well as individual and contextual factors at the time the bill was considered (as captured by the member- and bill-level covariates).

Scholars have previously utilized information embedded in bill texts to predict votes on those bills. For example, Ballard (2022) uses a logistic regression with Doc2Vec weights and a variety of individual-level covariates to predict votes on pre-floor bills. Similarly, de Marchi, Dorsey and Ensley (2021) use bill topics derived from Structural Topic Models and individual-level covariates in LASSO regression models to predict votes.

Our approach improves upon these prior prediction models by both accounting for variation in the voting behavior of the individual members of Congress and adopting multiple machine learning models to increase accuracy. First, the models fit in previous papers do not account for interactions with member-level covariates (i.e., they assume that bill characteristics have the same predictive coefficients for *all* members of Congress in each model within parties). Functionally, this means that the bill weights will be given large coefficients only if they predict that either a very large proportion or a very low proportion of the members as a whole or within each party voted for the bill. As a result, this method potentially misses individualized voting patterns. To allow for complicated interactions among variables and to capture individual voting patterns, we consider tree-based models as our prediction models, which explore complex interactions by construction. Furthermore, to mitigate the issue of overfitting while maintaining high prediction accuracy, we consider different methods fitting ensembles of trees. For example, random forests bootstrap sample from the full data and randomly select variables at each potential split, selecting split points which minimize error in predictions (Breiman, 2001). Thus, our use of random forests allows us to maintain

high predictive accuracy through essentially non-parametric modeling of the data while also minimizing the risk of overfitting by randomly sampling the data and variables.

Second, rather than relying on a single prediction model, we build an ensemble method called *super-learning* (Van der Laan, Polley and Hubbard, 2007*b*; Phillips et al., 2022) to generate predictions based on multiple machine learning algorithms. Super-learner uses a cross-validated measure of prediction performance to weight the contribution of methods to the final estimate. The super-learning algorithm is guaranteed, asymptotically, to perform at least as well as any of the component learners, if not better (van der Laan, Polley and Hubbard, 2007*a*). Therefore, rather than selecting a single model based on performance, we preserve as much information as possible by allowing the super-learner to weight and combine predictions from multiple models.

3.2.1 Prediction Model

Our super-learner model usees bill- and member-level covariates. The bill-level predictors include (1) the Doc2Vec vector of neural weights (generated with both distributed memory and distributed back of words); (2) the number of co-sponsors; and (3) the referred committee for each bill; and (4) bill type[4]. The member-level predictors include (1) individual member indicator variables; (2) the member's margin of victory in the most recent general election; (3) party affiliation; (4) an indicator for whether the member belongs to the House majority party; (5) an indicator for whether the member belongs to the same party as the Senate's majority; (7) an indicator for whether the member is the bill's sponsor; and (8) an indicator for whether the member is the bill's co-sponsor.[5]

Our super-learner algorithm considers three different tree-based prediction models—Random Forest, Gradient Boosting Machines (GBM), and Extreme Gradient Boosting (XGBoost)—

[4] i.e., Bill, Joint Resolution, Concurrent Resolution, and Simple Resolution.

[5] Due to the large size of data, we fit our model separately for Congresses between 103$^{\text{rd}}$ and 109$^{\text{th}}$ and Congresses between 110$^{\text{th}}$ and 116$^{\text{th}}$.

and uses Random Forest as a meta-learner to assign weights to each of these individual models.[6] Furthermore, in taking advantage of the super-learner model, we also include two generalized models employed in previous studies—logistic and LASSO—and allow the model to determine the weighted combination of the candidate prediction models.[7]

3.2.2 Validating the Prediction Model

Our ideal point estimation critically hinges on the assumption that predicted votes on pre-floor bills reflect how members *would have* voted had those bills reached the floor. To validate this assumption, we must demonstrate that our prediction model is highly predictive of the members voting decisions regardless of the bill status, and that the function predicting each member's vote using bill- and member-level covariates is the same across floor and pre-floor bills.

***Prediction Accuracy*:** We first assess the performance of our prediction method by calculating the overall error rate, which represents the proportion of floor votes the model incorrectly classifies.[8] To avoid overfitting, we calculate the error rates using 5-fold cross-validation. By randomly leaving out parts of the sample during model fitting and predicting outcomes for those observations, we approximate our ultimate task of out-of-sample prediction. The out-of-sample predictive accuracy of our model is shown in Figure 11. The overall classification error across all Congresses and members' majority status is 0.026. This error rate is significantly lower than other text-based models that predict member's voting behavior on final passage votes in the House of Representatives—e.g., 0.052

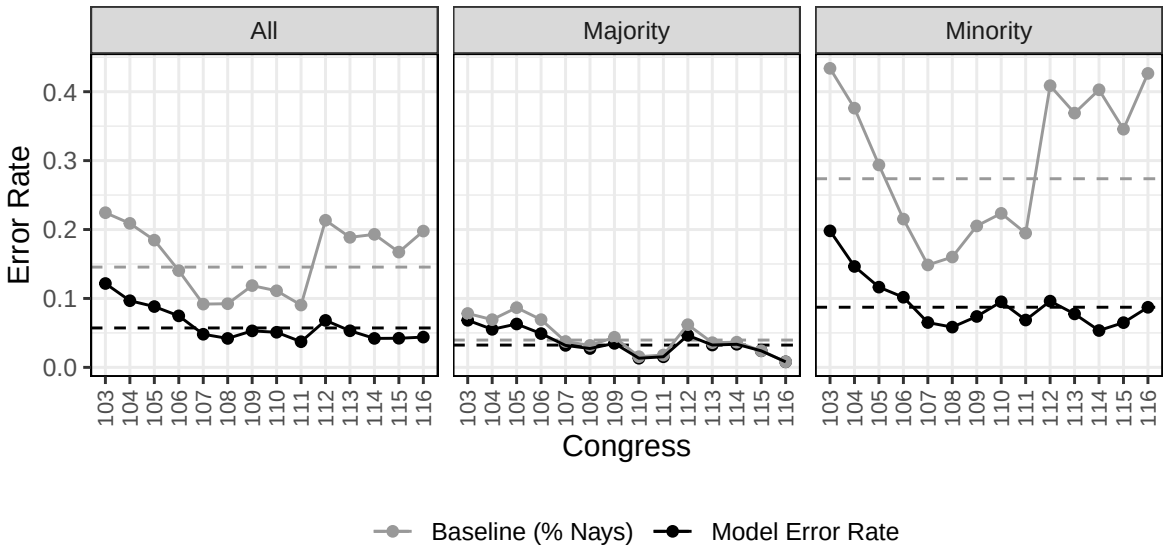
[6]For random forest, we grow 800 trees, each time randomly selecting half of the covariates available. For GBM, we grow 50 trees with a 20% learning rate. For XGBoost, 30 trees and a 30% learning rate were used. We use a 5-fold cross-validation scheme for all learners. According to the practical guidelines by the original author of super-learner, a cross-validation scheme as small as 2 is acceptable for datasets larger than 10,000 observations (Phillips et al., 2022). Additionally, our results are substantively unchanged using GLM as a meta-learner

[7]As we note in the next section, our final super-learner model allocates a significant portion of the weights to the three tree-based models and assigns almost no weights to the two generalized models.

[8]We also compute the rates of false positives (Type I error) and false negatives (Type II error). See Appendix B.1.

by Ballard (2022) and 0.056 by de Marchi, Dorsey and Ensley (2021).[9]

Figure 1: Prediction Model Classification Error Rate



Note: Classification error rate (black) and the baseline error rate (gray)—i.e., share of nay votes—on floor bills by member majority status. Dotted lines represent the average across all Congresses. See Appendix B.1 for a full error analysis.

Figure 11 shows that the model performs better for members of the majority party (center plot) than for minority party members (right plot). This is partly because the rate at which the minority party members vote yea is significantly lower than the members of the majority party members (as illustrated in gray lines). Specifically, the share of yea votes in a given Congress ranges between 56.6% and 85.1% among minority party members and between 91.3% and 99.2% among majority party members. Given that our prediction model predicts yea votes more frequently than nay votes, any votes predicted as yea are likely to result in a correct classification among the majority party members whereas the predicted yea votes are likely to generate false positives among the minority party members.

In each Congress, the model increases prediction accuracy over a baseline error rate achieved from predicting all yea votes in the House (roughly 0.15 across all Congresses and

[9]See Ballard (2022) Table 2 for reference.

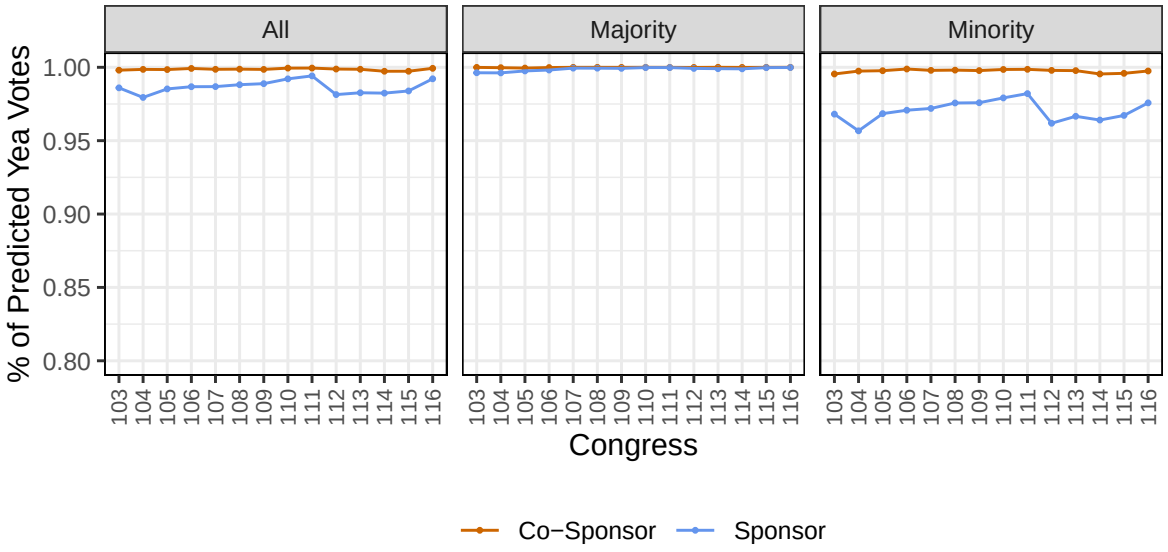
members). However, we observe slight variation in the error rate across different Congresses with the earlier Congresses generating relatively higher error rates. The misclassifications are largely driven by appropriations bills, particularly on those that were vetoed by President Bill Clinton, which the minority party members tended to vote in favor of, generating high false negatives.[10]

***Validating against Pre-floor Bill Sponsorship*:** For a second set of validation checks, we use members' sponsorship decisions on pre-floor bills. While the absence of sponsorship does not imply opposition to the bill, presumably, the positive decision to sponsor a bill implies support for the bill. Indeed, members of Congress vote yea on bills that they sponsor over 99.6% of the time and on bills that they co-sponsor over 99.2% of the time. Thus, we assume that the members would have voted in favor of the pre-floor bills that they (co-)sponsored; doing so allows us to compare predicted votes on pre-floor bills to an existing measure of member preferences on pre-floor bills.

Figure 2 shows the proportions of predicted yea votes separately for co-sponsors and sponsors of pre-floor bills. The total shares of predicted yea votes are 98.7% among all co-sponsors and 97.7% among all sponsors. Assuming that all (co-)sponsors would vote in favor of pre-floor bills they (co-)sponsor, our model predicts the preferences of these members fairly well. However, examining the totals separately by the majority status shows that, while almost all (co-)sponsors' votes from the majority party are correctly predicted, the minority party members' votes are not. This is largely because the minority party members rarely (co-)sponsor bills as compared to members of the majority party. For example, in the 116$^{\text{th}}$ Congress, only 4% of minority members co-sponsored a bill as compared to over 20% of the majority members. Consequently, a small number of errors for the minority party can result in larger error rates.

[10]The Republican party had the majority in both the House and the Senate between the 104$^{\text{th}}$ and 106$^{\text{th}}$ Congresses.

Figure 2: Share of Predicted Yea Votes Among Co-Sponsors and Sponsors of Pre-floor Bills

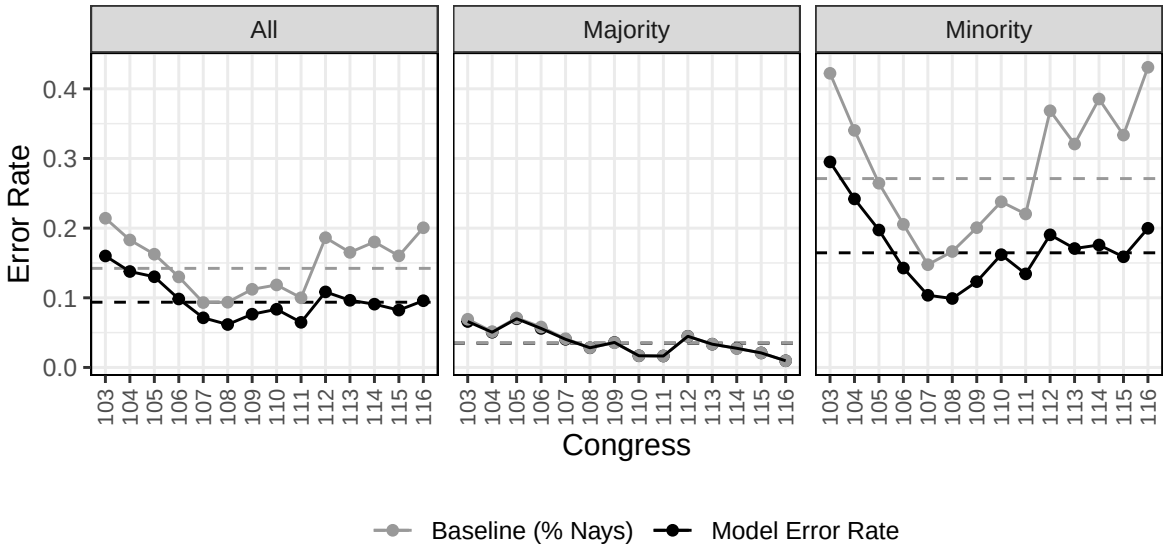


Note: Proportion of predicted yea votes from the prediction model among co-sponsors (red) and sponsors (blue) of pre-floor bills by member majority status.

***Validating against Earlier Versions*:** Finally, to determine whether we can reliably predict votes using pre-floor versions of bills, we predict votes for floor bills using early versions of the bills and compare the error rates with those from our main models. Our analysis requires the assumption that the voting models are the same for both floor- and pre-floor bills (i.e., that votes on pre-floor bills, had they reached the floor, would be similarly predicted by member- and bill- characteristics as those for floor bills). The earlier, unamended versions of the floor bills are likely to be more similar to those bills which never reached the floor. Therefore, if our prediction model performs well for the earlier versions of the floor bills, it lends credibility to our model's ability to perform well on the pre-floor bills.

Figure 3 shows the overall classification error rate among earlier versions of the final bills that reached the floor. The overall error rate across all Congresses and member majority is 0.11 (black dotted line in the left plot). While our model predicts vote preferences on the older versions of the bills fairly well, it does not perform as well as it does on the final versions of the floor bills. Again, the predictions of the majority party members' votes tend

Figure 3: Classification Error Rates on Early Versions



Note: Classification error rate (black) and the baseline error rate (gray) by member majority status using early versions of floor bills. Dotted lines represent the average across all Congresses. The final voting decision made on the final version of the bills are used as a true outcome. See Appendix B.5 for a full table.

to be more accurate than the predictions for the minority party members.

4 Estimating Ideal Points and Dimensionality

After generating and validating preferences on all bills introduced in the House of Representatives, we are able to examine the effects of agenda-setting on ideal point estimates and the conclusions drawn from those estimates. Following Clinton, Jackman and Rivers (2004), we fit unidimensional item response models to test our hypotheses that majority-party agenda control increases estimates of intra-party homogeneity and partisan polarization. Specifically, we fit dynamic IRT models of the form:

$$y_{it} = \beta_{bt}^{T}\theta_{it} - \alpha_{bt}$$

where β_{bt}^T is a discrimination paramater for bill b in Congress t, α_{bt} is a difficulty parameter for bill b in Congress t, and θ_{it} is the ideal point for legislator i in Congress t.

Moreover, per Martin and Quinn (2002), this model places the following priors on individual and bill parameters:

$$\theta_{it} \sim \mathcal{N}(\theta_{i,t-1}, \omega_\theta^2)$$
$$\left[\begin{smallmatrix}\alpha_{bt}\\ \beta_{bt}\end{smallmatrix}\right] \sim \mathcal{N}_2(\mathbf{b_0}, \mathbf{B_0})$$

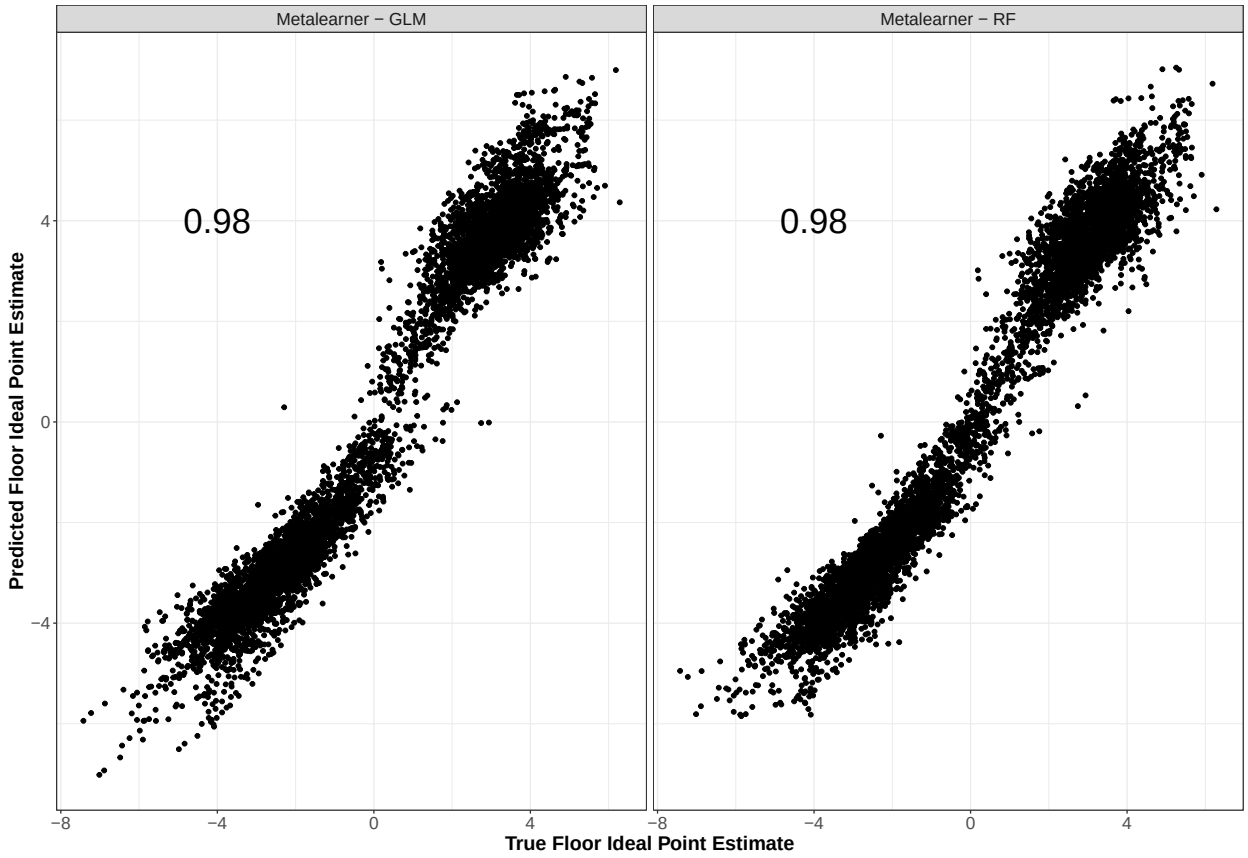
Thus, individual ideal points depend only on ideal points in the previous Congress and an evolution variance parameter, ω_θ^2, and bill parameters are modeled according to multivariate normal priors. Due to the large size of our dataset (more than 1300 representatives and more than 100,000 bills), we calculate IRT estimates using variational expectation maximization which has been shown to be much faster than but nonetheless comparable to fully Bayesian IRT (Imai, Lo and Olmsted, 2016). To avoid ad hoc selection of "liberal" and "conservative" members for priors, we set set strong priors on party leaders' ideal points in each Congress, with means of -1 for Democratic leaders and 1 for Republican leaders and standard deviations of 0.1. Following Poole and Rosenthal (1997), we exclude bills with (predicted) passage votes greater than 97.5% because the lack of variation in outcomes can make estimation of bill parameters untenable and because these bills contribute very little to distinguishing among members.

4.1 Validating the IRT Model

One potential concern with our approach is that the errors generated from our prediction model may result in some systematic errors in ideal point estimation. If this is true, any substantively meaningful patterns observed in our ideal point estimates might simply be

the product of systematic misclassifications from the prediction model. We check whether this is the case by re-estimating the ideal points using floor bills. Specifically, we split the floor bills into a training set (50%) and a test set (50%) and train our prediction model only on the training set. We then predict members' votes on the floor bills in the test set and fit separate IRT models using predicted and actual votes on floor bills. Finally, we compare the two sets of estimates and examine how well the ideal points based on predicted votes recover the ideal points based on observed outcomes.

Figure 4: Correlations between True and Predicted Ideal Point Estimates



Note: Correlation between ideal points generated with predicted (y-axis) and true (x-axis) votes on floor bills. The overall correlation is 0.98 for both metalearners. See Appendix D.2 for the correlation analyses by Congress.

Figure 12 plots the correlation between ideal point estimates generated with predicted ideal point estimates and those generated from true votes on floor bills. The overall corre-

lation of the two estimates are 0.98. Correlations within Congresses are all at least 0.97. In sum, the high correlation between the two ideal points indicates that the ideal points based on predicted votes are able to recover the patterns observed with the true voting behavior, and that the prediction errors are unlikely to drive the patterns we observe in the ideal point estimates.

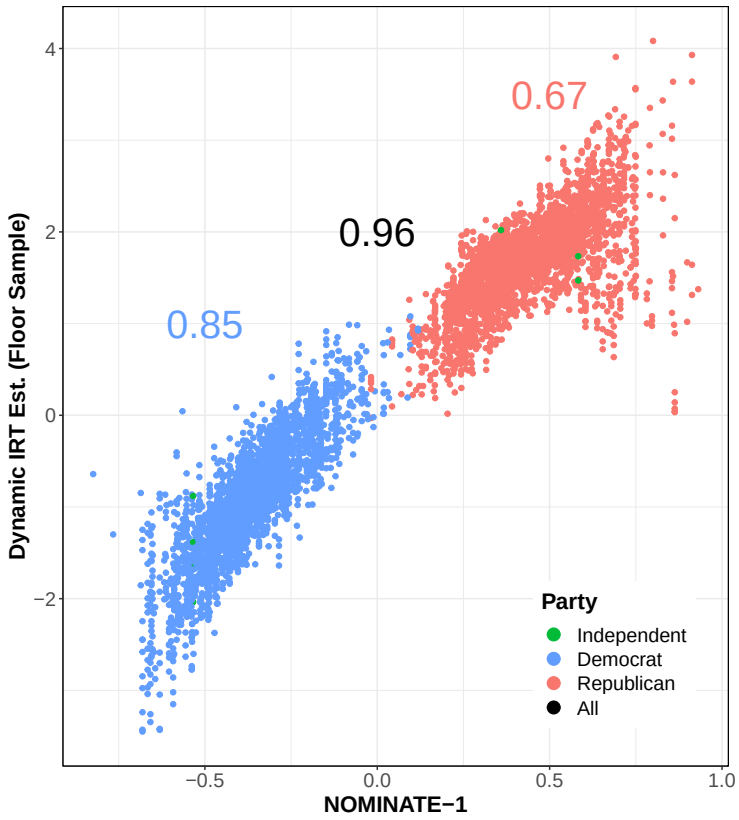
Another way our model might differ from traditional ideal point measures is that we utilize only votes on bills rather than all votes (i.e., our approach necessarily excludes votes on amendments and procedural questions). To ensure this is not the case, we train an IRT model on only the final passage votes in our prediction model's training set and compare our ideal point estimates with DW-NOMINATE scores. The results, shown in Figure 5, indicate that our exclusion of amendments and procedural matters should not substantially alter our findings: our floor-bill IRT estimates are highly correlated with first dimension NOMINATE scores, obtaining a correlation of 0.96 overall and 0.85 (0.67) for Democrats (Republicans).

4.2 An Overview of IRT Estimates

Having validated our prediction and IRT models, we turn now to the analysis of our IRT results. We begin with a descriptive exploration of our new measure and a comparison of our measure with floor-based measures before proceeding to a more rigorous statistical testing of our hypotheses. Though we found a high correlation between NOMINATE scores and IRT estimates based on final passage votes alone (see Figure 5), below, we plot pre-floor IRT estimates against floor IRT estimates from final passage roll calls for comparability.

In Figure 6, we plot pre-floor IRT estimates against floor IRT estimates based only on final passage floor votes. There appears to be a wide variation in the strength and shape of the relationships between floor and pre-floor IRT estimates. Only five out of 28 within-party correlations reach or surpass 0.7, suggesting that our pre-floor estimates are capturing a distinct facet of congressional behavior. Interestingly, within-party loess lines indicate

Figure 5: Correlation between Floor-Bill IRT Estimates and DW-NOMINATE

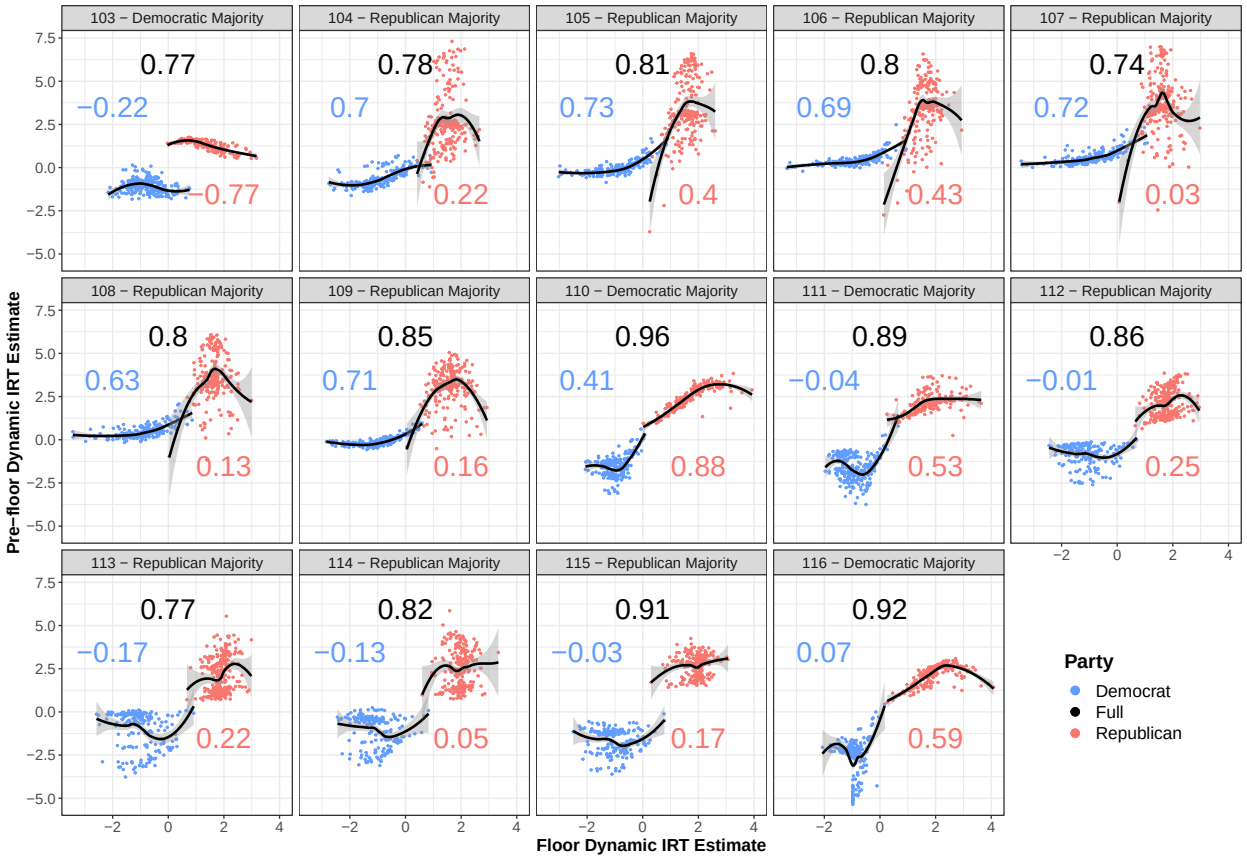


Note: Correlation between IRT ideal points estimated using only final passage votes on bills and NOMINATE's first dimension. These measures correlate at 0.96 overall, 0.85 for Democrats (blue), and 0.67 for Republicans (red).

curvilinear relationships between floor and pre-floor estimates for many Congresses. This curvilinearity is quite pronounced for the majority party in most Congresses, pointing to unique variation in majority party ideal points not displayed through the floor agenda alone.

Finally, Figure 7 displays the relationship between pre-floor and floor IRT estimates in the 116$^{\text{th}}$ Congress. Many members of Congress end up with similar relative locations (i.e., Reps. Katko (R-NY), Peterson (D-MN), and Cuellar (D-TX), as well as members of the "Squad", Reps. Ocasio-Cortez (D-NY), Tlaib (D-MI), Pressley (D-MA), and Omar (D-MN) are close to the center in both floor and pre-floor IRT estimates). However, for other members, relative positions are largely different for floor and pre-floor estimates. For

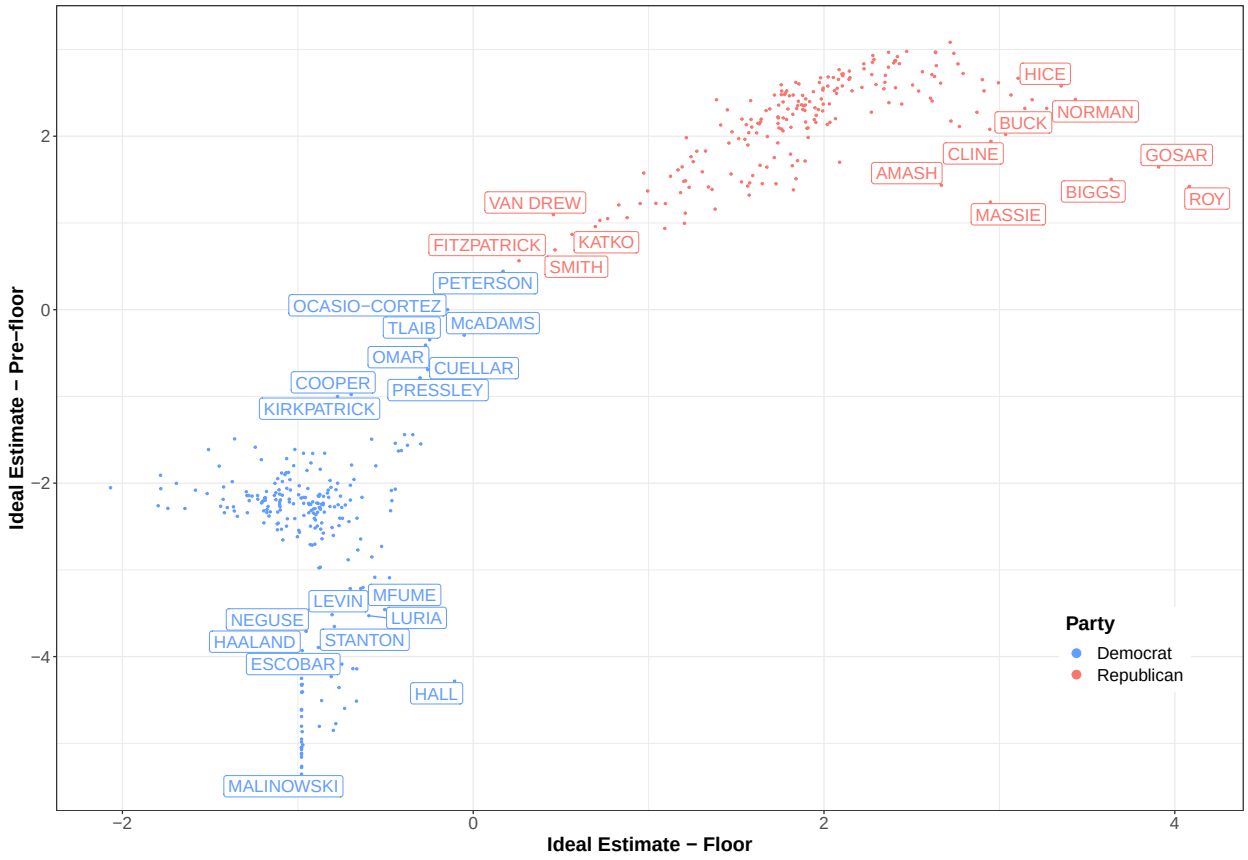
Figure 6: Pre-floor vs. Floor IRT Estimates



Note: IRT estimates from pre-floor predicted votes are plotted against IRT estimates from final passage roll call votes. Correlations are displayed alongside loess fits.

example, prominent members of the archconservative House Freedom Caucus, including Reps. Gosar (R-AZ), Biggs (R-AZ), Roy (R-TX), and Hice (R-GA), appear quite extreme on the floor distribution but are closer to the center of the pre-floor distribution. Among Democrats, some members who appear to be moderates based on their floor ideal points, are further from the center when examining pre-floor estimates (see e.g., Reps. Luria (D-VA) and Malinowski (D-NJ)). In short, then, our ideal point estimates derived from imputed pre-floor preferences, while related to floor ideal point estimates, are nonetheless distinct.

Figure 7: Pre-floor vs Floor IRT Estimates in the 116th Congress



Note: Pre-floor and floor IRT estimates are displayed for the 116th Congress.

5 IRT Results: The Role of Agenda Setting

Existing literature leads us to expect that majority agenda control will circumscribe the policy space on the floor. Specifically, Cox and McCubbins (2005) argue that the majority party uses its cartel-like control of the agenda to prevent bills that might result in rolls of the majority party from reaching the floor. Further, Lee (2016) shows that the Republican takeover of the House in 1994 ushered in a new era of partisan competition for congressional majority status, incentivizing party leaders to structure the agenda to bolster and distinguish their own party brands. These theories carry several implications in the context of our estimation of pre-floor ideal points. First, by preventing votes for which majority and

minority party members might join, majority party leadership should suppress intra-party heterogeneity, making it appear, based on floor estimates, that the majority party is more homogeneous than it is. Estimating ideal points using bills that do not make it to the floor should allow us to see greater intra-party heterogeneity than would be visible using floor votes alone. Second, relatedly, by censoring bills likely to roll the majority party, majority party leadership, by definition, suppresses bills generating agreement *between* the parties. Consequently, pre-floor estimates should display lower levels of partisan polarization if the majority party agenda setters systematically keep bills with inter-partisan agreement off of the floor. Third, the argument that the majority party uses the floor agenda to distinguish the party brands implies that the floor agenda will be largely structured by a single dimension: that of partisan conflict. Indeed, this is precisely what Poole and Rosenthal (1997) find, leading them to proclaim the existence of the "Unidimensional Congress." Below, we test each of these hypotheses in succession using our new estimates of pre-floor preferences.

5.1 Intra-Party Homogeneity

The work of Cox and McCubbins (2005) and Lee (2016) suggest that the majority party structures the floor agenda to generate distinctive party brands, attempting to clearly demarcate the majority and minority parties. By preventing majority party rolls, the majority party seeks, in particular, to present a unified image while remaining agnostic as to the implications of agenda control for perceived minority party homogeneity (Cox and McCubbins, 2005). Thus, we expect that our pre-floor IRT estimates will exhibit more heterogeneity than will floor IRT estimates—especially for the majority party.

To test this hypothesis, we first generate standardized measures of heterogeneity in each Congress for both pre-floor and floor IRT estimates and for both parties. Conventional standardized measures of heterogeneity take the ratios of measures of spread to measures of the central tendency. We employ three such measures of heterogeneity: the Coefficient

of Variation, the ratio of the mean absolute deviation to the median, and the ratio of the interquartile range to the median. We define each of these measures as follows:

$$
\begin{aligned}
CV(\mathbf{x}) &= \frac{\sigma_x}{|\bar{\mathbf{x}}|} \\
MAD\text{-}M(\mathbf{x}) &= \frac{med(|x_i - med(\mathbf{x})|)_{i\in(1,\ldots,N)}}{|med(\mathbf{x})|} \\
IQR\text{-}M(\mathbf{x}) &= \left|\frac{(Q_{75} - Q_{25})}{med(\mathbf{x})}\right|
\end{aligned}
$$

where $\mathbf{x}$ consists of the estimated ideal points for a given party in a given Congress for floor or pre-floor preferences. In each case, a higher statistic indicates that variation in the ideal points is greater relative to the central tendency. As they are all ratios, these statistics result in asymmetries, so to allow for symmetrical, linear comparisons we take the logarithm of each measure in our analyses. We expect that floor IRT estimates will display *less* variation than pre-floor estimates for the majority party. To test this hypothesis, we regress our measures of party heterogeneity on a dummy variable indicating whether this measure is based on floor IRT estimates or pre-floor estimates. Our results, shown in Table 1, demonstrate that, for the majority party, floor IRT estimates systematically display less heterogeneity than pre-floor IRT estimates. Results for the minority party, shown in Table 2, show no such relationship between party heterogeneity and floor or pre-floor IRT estimates.

5.2 Partisan Polarization

Not only should majority party leadership suppress intra-party heterogeneity, but it should also place an emphasis on areas of partisan conflict, according to Lee (2016). This emphasis on partisan conflict should help the majority party distinguish itself electorally from the minority party. In short, we should expect to find more polarization using floor ideal points than with pre-floor ideal points.

Table 1: Comparing Majority Party Heterogeneity for Floor and Pre-floor

	log(CV)		log(IQR-M)		log(MAD-M)	
Floor	-0.375**	-0.375+	-0.412*	-0.412	-0.414*	-0.414
	(0.123)	(0.175)	(0.173)	(0.244)	(0.177)	(0.251)
Constant	-0.891***	-0.494***	-0.717***	-0.116	-1.424***	-0.793***
	(0.064)	(0.087)	(0.098)	(0.122)	(0.112)	(0.125)
Congress Fixed Effects	-	✓	-	✓	-	✓
Num. Obs.	28	28	28	28	28	28
R^2	0.262	0.643	0.194	0.573	0.182	0.580
Adj. R^2	0.234	0.259	0.163	0.113	0.151	0.128

Note: Errors clustered by Congress. + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table 2: Comparing Minority Party Heterogeneity for Floor and Pre-floor

	log(CV)		log(IQR-M)		log(MAD-M)	
Floor	-0.278	-0.278	-0.229	-0.229	-0.042	-0.042
	(0.318)	(0.449)	(0.209)	(0.296)	(0.178)	(0.251)
Constant	-0.198	-0.968***	0.004	-0.647***	-0.862***	-1.457***
	(0.359)	(0.225)	(0.238)	(0.148)	(0.217)	(0.126)
Congress Fixed Effects	-	✓	-	✓	-	✓
Num. Obs.	28	28	28	28	28	28
R^2	0.022	0.639	0.030	0.684	0.001	0.726
Adj. R^2	-0.016	0.250	-0.007	0.343	-0.037	0.430

Note: Errors clustered by Congress. + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

To compare floor and pre-floor levels of partisan polarization, we first standardize the average difference between the parties' ideal point estimates. Our measure of partisan polarization is then the standardized mean difference, defined as follows:

$$SMD(\mathbf{x_{pft}}, \mathbf{x_{-p,ft}}) = \left| \frac{(\bar{\mathbf{x}}_{\mathbf{pft}} - \bar{\mathbf{x}}_{\mathbf{-p,ft}})}{\sqrt{\frac{s^2_{\mathbf{x_{pft}}} + s^2_{\mathbf{x_{-p,ft}}}}{2}}} \right|$$

where $\mathbf{x_{pft}}$ is the vector of ideal points for party p on the f floor or pre-floor in Congress t, and s^2 is the variance of ideal points for each party. The standardized mean difference allows us to compare difference-in-means estimates even when the original variables are not on the same scale. In the context of our hypothesis, we expect the standardized difference between the parties' average ideal points to be larger for floor votes than for pre-floor bills. To compare floor to pre-floor estimates of polarization, we run linear regressions of standardized mean differences on an indicator for floor estimates. In Table 3, we display the results of regressions using all data, only data with Democratic majorities, and only data with Republican majorities. With the exception of Democratic-majority Congresses (of which there were four in our sample), our results suggest that partisan polarization tends to be much higher on the floor than on bills which do not reach the floor.[11] In other words, majority party agenda control has the effect of increasing partisan polarization relative to the pre-floor policy space.

5.3 The Unidimensional Congress?

Lastly, given the findings that the majority party structures the floor agenda to emphasize partisan conflict and to de-emphasize intra-party disagreements, we expect the floor

[11] We leave the question of why this increased polarization does not appear under Democratic majorities to future work. However, looking at each Congress individually, it appears that the 103rd Congress with a Democratic majority is a large outlier in the direction opposite our expectations, pulling the Democratic Party's coefficient estimates downward. It is notable that this outlier occurs in the final Congress *before*, Lee (2016) argues, partisan competition became a driving force of congressional behavior.

Table 3: Comparing Partisan Polarization for Floor and Pre-floor IRT Estimates

	Standardized Mean Difference					
	All Majorities		Democratic Majorities		Republican Majorities	
Floor	0.456	0.456	-1.390+	-1.390	1.194**	1.194*
	(0.449)	(0.635)	(0.595)	(0.772)	(0.314)	(0.444)
Constant	4.221***	5.020***	5.959***	5.943*	3.526***	2.658***
	(0.425)	(0.318)	(0.359)	(1.279)	(0.387)	(0.222)
Congress Fixed Effects	-	✓	-	✓	-	✓
Num. Obs.	28	28	8	8	20	20
R^2	0.034	0.593	0.462	0.546	0.258	0.843
Adj. R^2	-0.004	0.154	0.372	-0.059	0.217	0.669

Note: Errors clustered by Congress. + p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

agenda to largely be structured by a single dimension—the dimension of partisan conflict. In fact, this is precisely what Poole and Rosenthal (1997) show using NOMINATE scores. However, that the floor agenda is structured by a single dimension does not imply that the American policy space writ large is unidimensional. To the contrary, we argue that the floor agenda masks a great deal of the policy space. To test our hypothesis, we use Multiple Correspondence Analysis (MCA) to estimate the underlying dimensionality of the policy space for both floor and pre-floor bills (see e.g., Le Roux and Rouanet, 2005).

One specific challenge of comparing the dimensionality of the floor and pre-floor policy spaces is that our pre-floor policy space is composed of many more bills than our floor policy space. As a result, the pre-floor policy space inherently has many more dimensions, and running MCA on the entire set of pre-floor bills would reflect this higher dimensionality. To ensure that our estimates of the dimensionality of the policy spaces are not influenced by the number of bills in the floor and pre-floor sets, we sample the pre-floor bills with

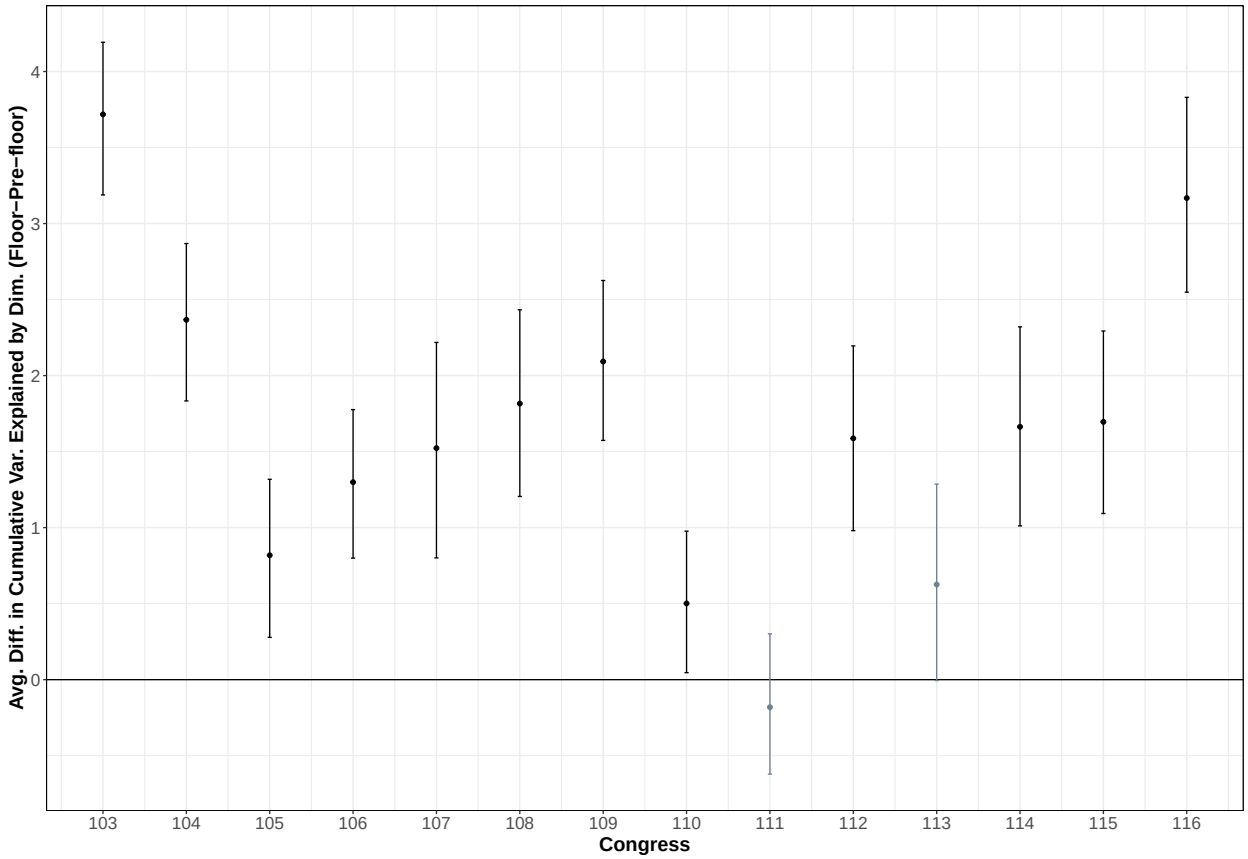
replacement, such that the number of sampled pre-floor bills, b_{pf}, is equal to the number of floor bills, B_f. Then, we run MCA separately on the imputed vote matrices for sampled pre-floor bills and on the entire roll-call matrix of floor bills. We repeat this process 1,000 times for each Congress in our sample. Our last step allows us to capture the dimensionality of the floor and pre-floor policy spaces: we store the cumulative variance explained by each additional dimension of our MCA fits. If the pre-floor policy space has more dimensions than the floor policy space, a higher amount of cumulative variance should be explained by fewer dimensions for the floor policy space. In other words, the average difference between the cumulative variance explained by each additional dimension for the floor and the pre-floor samples should be positive. Indeed, this is precisely what we find.

Figure 8 displays our findings. We plot the average difference in cumulative variance explained by each additional dimension between floor bills and our pre-floor samples (averaged across each of our 1,000 pre-floor samples per Congress). In 13 of 14 Congresses, more cumulative variance on floor bills is explained by each additional dimension than is explained on pre-floor bills, and in 12 Congresses, average differences from at least 95% of our pre-floor samples are greater than 0. Therefore, we conclude that, holding the number of bills constant, the floor policy space is better represented in lower-dimensional space than is the pre-floor policy space. The policy space, when subject to majority party agenda control, appears to have fewer dimensions, yet the American policy space itself does not appear to be inherently unidimensional.

6 Discussion

For decades, scholars' main approach to studying Congress has relied on unidimensional models of roll call voting. Still, this work relies on the assumption that the floor of the House is representative of the entire policy space, which Cox and McCubbins (2005) suggest might

Figure 8: Diff. in Cumulative Variance Explained by MCA Dimensions (Floor - Pre-floor)



Note: Figure depicts the average difference between the cumulative variance explained by each additional MCA dimension for the floor and the pre-floor. Uncertainty intervals are formed by taking the 2.5 and 97.5 percentiles of MCA estimates from the 1000 pre-floor bill samples in each Congress. Higher values indicate that floor votes are better explained by fewer dimensions than are pre-floor preferences.

not hold. Our paper shows the extent to which roll call measures of ideal points are subject to the influence of the majority party's agenda control, which both increases estimates of partisan polarization and decreases estimates of intra-party heterogeneity. Further, while the floor agenda seems to be better explained by fewer dimensions, this unidimensionality does not well describe the pre-floor policy space. In demonstrating the limitations of ideal point estimates based on floor bills, we present a new framework for estimating ideal points independently of this agenda control.

We hope our estimates of pre-floor bill preferences and ideal points can prove useful to scholars of congressional behavior. In particular, future work can explore the substance of the dimensions of pre-floor preferences, including examining the policies captured by each dimension and what estimated dimensions tell us about members of Congress. Moreover, with new estimates of congressional ideal points, scholars can probe the relationship between these preferences and other congressional behavior and can re-examine old findings with our new measures. Finally, in a political environment that seems increasingly riven by partisan conflict, our results offer some reason for optimism and suggest where individuals interested in reducing partisan polarization might look: the pre-floor policy space. Other researchers can examine the policies that generate the observed inter-partisan agreement or internally divide the parties. Our pre-floor results suggest that the parties are not as polarized and internally unified as it might seem based on the House floor alone.

References

Bailey, Michael A. 2013. "Is today's court the most conservative in sixty years? Challenges and opportunities in measuring judicial preferences." *The Journal of Politics* 75(3):821–834.

Bailey, Michael A, Brian Kamoie and Forrest Maltzman. 2005. "Signals from the tenth justice: The political role of the solicitor general in Supreme Court decision making." *American Journal of Political Science* 49(1):72–85.

Ballard, Andrew O. 2022. "Bill Text and Agenda Control in the US Congress." *The Journal of Politics* 84(1):000–000.

Binder, Sarah A. 1997. *Minority rights, majority rule: Partisanship and the development of Congress*. Cambridge University Press.

Bonica, Adam. 2018. "Inferring roll-call scores from campaign contributions using supervised machine learning." *American Journal of Political Science* 62(4):830–848.

Breiman, Leo. 2001. "Random forests." *Machine learning* 45(1):5–32.

Clinton, Joshua D, Anthony Bertelli, Christian R Grose, David E Lewis and David C Nixon. 2012. "Separated powers in the United States: The ideology of agencies, presidents, and congress." *American Journal of Political Science* 56(2):341–354.

Clinton, Joshua, Simon Jackman and Douglas Rivers. 2004. "The statistical analysis of roll call data." *American Political Science Review* 98(2):355–370.

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

Cox, Gary W, Thad Kousser and Mathew D McCubbins. 2010. "Party power or preferences?

Quasi-experimental evidence from American state legislatures." *The Journal of Politics* 72(3):799–811.

de Marchi, Scott, Spencer Dorsey and Michael J Ensley. 2021. "Policy and the structure of roll call voting in the US house." *Journal of Public Policy* 41(2):384–408.

Hix, Simon, Abdul Noury and Gerard Roland. 2006. "Dimensions of politics in the European Parliament." *American Journal of Political Science* 50(2):494–520.

Imai, Kosuke, James Lo and Jonathan Olmsted. 2016. "Fast Estimation of Ideal Points with Massive Data." *American Political Science Review* 110(4):631–656.

Le, Quoc and Thomas Mikolov. 2014. Distributed Representations of Sentences and Documents. In *Proceedings of the 31st International Conference on Machine Learning*. Vol. 32 Beijing, China: PMLR pp. 1188–1196.

Le Roux, Brigitte and Henry Rouanet. 2005. *Geometric Data Analysis: From Correspondence Analysis to Structured Data Analysis*. Dordrecht: Kluwer Academic Publishers.

Lee, Frances E. 2016. *Insecure majorities: Congress and the perpetual campaign*. University of Chicago Press.

Londregan, John Benedict. 2000. *Legislative institutions and ideology in Chile*. Cambridge University Press.

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

McCarty, Nolan M and Keith T Poole. 1995. "Veto power and legislation: An empirical analysis of executive and legislative bargaining from 1961 to 1986." *JL Econ. & Org.* 11:282.

Mikolov, Tomas, Kai Chen, Greg Corrado and Jeffrey Dean. 2013. "Efficient estimation of word representations in vector space." *arXiv preprint arXiv:1301.3781* .

Napolio, Nicholas and Christian Grose. 2021. "Crossing Over: Majority Party Control Affects Legislator Behavior and the Agenda." *American Political Science Review* 116(1):359–366.

Phillips, Rachael V, Mark J van der Laan, Hana Lee and Susan Gruber. 2022. "Practical considerations for specifying a super learner." *arXiv preprint arXiv:2204.06139* .

Poole, Keith and Howard Rosenthal. 1985. "A Spatial Model for Legislative Roll Call Analysis." *American Journal of Political Science* 29(2):357–384.

Poole, Keith T and Howard Rosenthal. 1997. *Congress: A Political-economic History of Roll Call Voting*. Oxford University Press on Demand.

Řehůřek, Radim and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. Valletta, Malta: ELRA pp. 45–50. http://is.muni.cz/publication/884893/en.

Shor, Boris and Nolan McCarty. 2011. "The ideological mapping of American legislatures." *American Political Science Review* 105(3):530–551.

van der Laan, Mark, Eric Polley and Alan Hubbard. 2007*a*. "Super Learner." *Statistical Applications of Genetcs and Microbiology* 6(1).

Van der Laan, Mark J, Eric C Polley and Alan E Hubbard. 2007*b*. "Super learner." *Statistical applications in genetics and molecular biology* 6(1).

Voeten, Erik. 2000. "Clashes in the Assembly." *International organization* 54(2):185–215.

Appendix

Contents

A Descriptive Statistics of Congressional Bills

Table 4: Bill Actions by Congress

Version	Congress													
	103	**104**	**105**	**106**	**107**	**108**	**109**	**110**	**111**	**112**	**113**	**114**	**115**	**116**
ASH	0	0	0	0	0	1	0	0	0	0	0	0	0	0
ATH	104	97	107	103	126	58	16	16	2	0	0	0	0	0
CDH	7	6	0	2	1	0	1	0	1	0	1	0	0	0
CPH	38	19	11	13	21	5	19	16	0	2	0	0	0	1
EAH	23	9	10	14	12	2	4	16	15	4	11	12	12	8
EH	884	886	999	1,283	1,087	1,184	1,273	1,964	1,743	757	830	1,053	1,314	1,050
ENR	344	308	312	492	381	421	410	391	329	235	255	264	324	247
FPH	0	0	0	0	0	0	1	0	0	0	3	0	0	0
IH	6,307	4,943	5,620	6,528	6,664	6,627	7,786	9,013	8,452	7,593	6,707	7,533	8,610	10,373
IPH	0	0	0	0	0	0	1	0	0	0	0	0	0	0
LTH	0	8	1	0	0	4	8	9	12	2	2	0	5	5
PAP	0	0	0	0	0	0	0	0	0	0	0	1	2	0
PCH	0	1	0	0	0	0	0	0	0	0	0	0	0	0
PP	33	29	26	28	20	25	18	12	14	5	2	3	0	0
RCH	4	8	2	2	2	1	0	2	1	1	0	0	0	0
RFH	0	0	0	0	0	2	0	0	0	0	0	0	0	0
RH	672	727	726	876	695	708	651	867	653	617	643	831	1,045	663
RHUC	0	0	0	0	0	0	0	0	1	1	0	0	0	0
RIH	0	1	1	0	2	1	1	0	0	0	0	0	0	0
RTH	0	1	0	0	0	0	0	1	6	0	0	2	0	0
SC	1	2	0	0	0	0	0	0	0	0	0	0	0	0

Note: See https://www.govinfo.gov for the definitions of the bill versions.

Table 5: Bill Types by Congress

Congress	Type of Bill				**Total**
	H.Con.Res.	H.J.Res.	H.R.	H.Res.	
103	438	549	6,564	866	8,417
104	347	270	5,544	884	7,045
105	512	213	6,105	985	7,815
106	684	234	7,343	1,080	9,341
107	780	187	7,085	959	9,011
108	762	171	6,822	1,284	9,039
109	728	140	7,754	1,567	10,189
110	683	122	9,197	2,305	12,307
111	501	131	7,956	2,641	11,229
112	203	143	7,861	1,010	9,217
113	188	166	7,168	932	8,454
114	261	121	8,187	1,130	9,699
115	196	187	9,519	1,410	11,312
116	164	134	10,600	1,449	12,347
Total	6,447	2,768	107,705	18,502	135,422

Note: H.Con.Res = House Concurrent Resolution; H.J.Res = House Joint Resolution; H.R = House Bill; H.Res = House Simple Resolution.

B Prediction Model Validation Checks

B.1 Prediction Accuracy

Table 6: Classification Error

Congress	Overall Error			False Negatives			False Positives		
	All	Majority	Minority	All	Majority	Minority	All	Majority	Minority
103	0.012	0.007	0.019	0.007	0.001	0.021	0.030	0.083	0.016
104	0.030	0.028	0.033	0.009	0.002	0.021	0.109	0.368	0.051
105	0.013	0.007	0.019	0.006	0.001	0.013	0.045	0.076	0.034
106	0.065	0.047	0.083	0.028	0.008	0.054	0.290	0.581	0.192
107	0.006	0.004	0.008	0.001	0.001	0.001	0.054	0.088	0.044
108	0.006	0.002	0.011	0.002	0.0003	0.005	0.047	0.069	0.042
109	0.004	0.002	0.006	0.001	0.0002	0.002	0.025	0.043	0.020
110	0.003	0.0003	0.006	0.001	0.00003	0.002	0.017	0.019	0.017
111	0.002	0.001	0.004	0.001	0.0001	0.003	0.013	0.044	0.009
112	0.028	0.042	0.009	0.004	0.002	0.007	0.116	0.648	0.012
113	0.001	0.001	0.002	0.001	0.0001	0.002	0.003	0.020	0.001
114	0.001	0.001	0.001	0.0005	0.0001	0.001	0.004	0.029	0.001
115	0.004	0.001	0.008	0.003	0.0002	0.007	0.011	0.044	0.008
116	0.001	0.001	0.002	0.001	0.00002	0.003	0.003	0.066	0.002
All	0.012	0.009	0.015	0.004	0.001	0.010	0.055	0.204	0.029

Note: The overall classification rate as well as false negative (Type I error) and false positive (Type II error) rates across Congress and majority status. The total sample size is 2.2 million.

B.2 Distribution of Error Rates

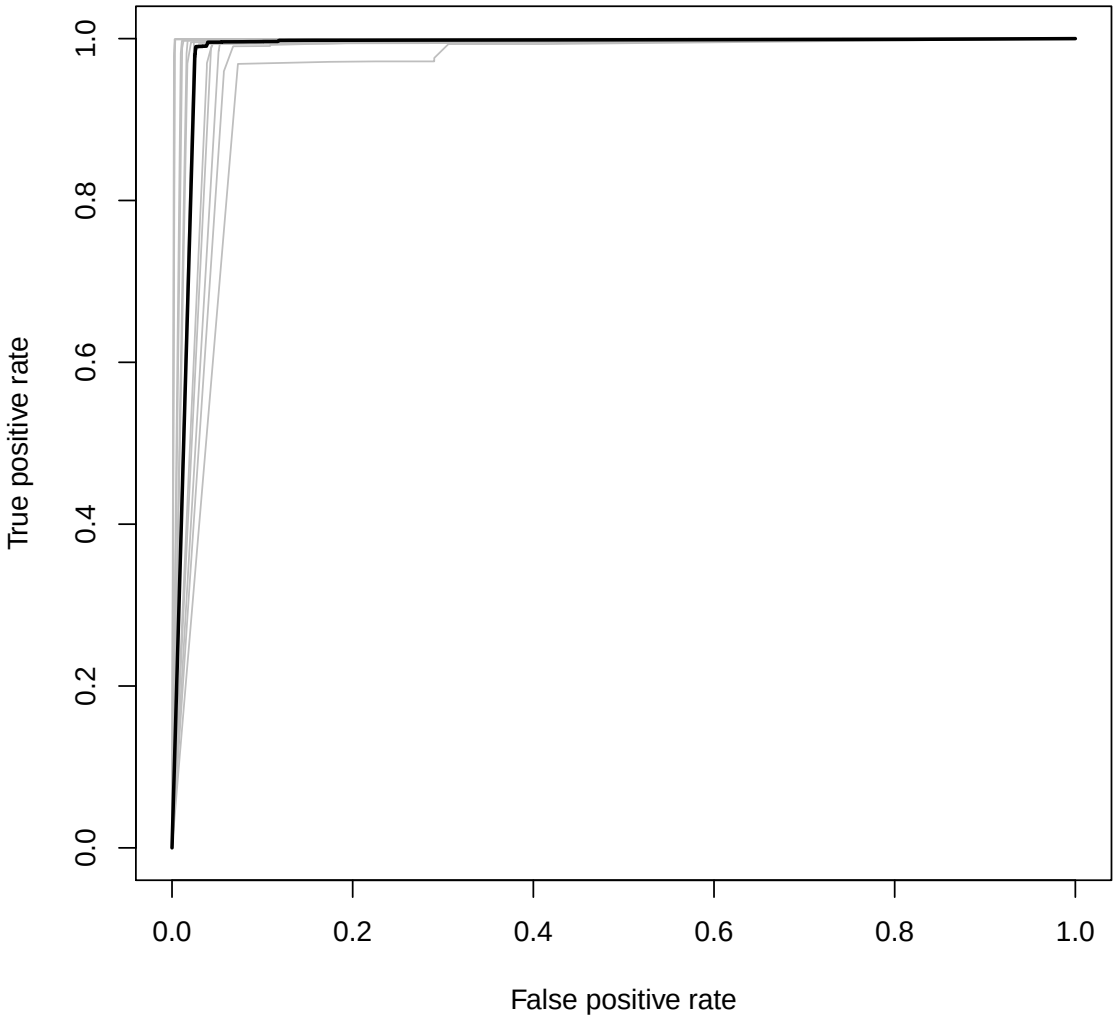
Table 7: Distribution of Error Rates by Member of Congress

Majority	Congress	Distribution						
Status		Mean	SD	Min	25th	Median	75th	Max
Majority	103	0.988	0.014	0.939	0.983	0.993	1.000	1.000
	104	0.969	0.020	0.884	0.958	0.972	0.986	1.000
	105	0.988	0.012	0.935	0.980	0.990	0.995	1.000
	106	0.947	0.055	0.547	0.937	0.964	0.977	1.000
	107	0.993	0.007	0.964	0.991	0.995	1.000	1.000
	108	0.996	0.005	0.973	0.992	0.996	1.000	1.000
	109	0.996	0.006	0.972	0.996	1.000	1.000	1.000
	110	0.999	0.001	0.994	1.000	1.000	1.000	1.000
	111	0.998	0.003	0.987	0.997	1.000	1.000	1.000
	112	0.951	0.057	0.567	0.937	0.970	0.985	1.000
	113	0.999	0.003	0.988	1.000	1.000	1.000	1.000
	114	0.998	0.004	0.980	0.995	1.000	1.000	1.000
	115	0.997	0.009	0.889	0.995	1.000	1.000	1.000
	116	0.999	0.003	0.985	1.000	1.000	1.000	1.000
Minority	103	0.971	0.017	0.924	0.96	0.973	0.983	1.000
	104	0.950	0.015	0.907	0.940	0.948	0.960	1.000
	105	0.967	0.011	0.941	0.960	0.967	0.975	1.000
	106	0.894	0.024	0.813	0.881	0.898	0.913	0.943
	107	0.986	0.009	0.954	0.981	0.986	0.991	1.000
	108	0.981	0.006	0.965	0.976	0.980	0.984	1.000
	109	0.989	0.007	0.967	0.984	0.989	0.994	1.000
	110	0.980	0.024	0.667	0.978	0.982	0.985	1.000
	111	0.993	0.004	0.908	0.991	0.993	0.995	1.000
	112	0.983	0.007	0.963	0.980	0.982	0.988	1.000
	113	0.997	0.004	0.982	0.994	1.000	1.000	1.000
	114	0.998	0.003	0.988	0.995	1.000	1.000	1.000
	115	0.984	0.006	0.967	0.981	0.983	0.987	1.000
	116	0.995	0.006	0.950	0.993	0.993	1.000	1.000

Note: Each row represents a descriptive statistics of the proportion of correct classifications for all members of congress in a given Congress and majority status. For example, among the majority party members in the 103rd Congress, each member's votes are correctly classified 98.8% of the time on average.

B.3 Receiver Operating Characteristic (ROC) Curve

Figure 9: ROC Curve

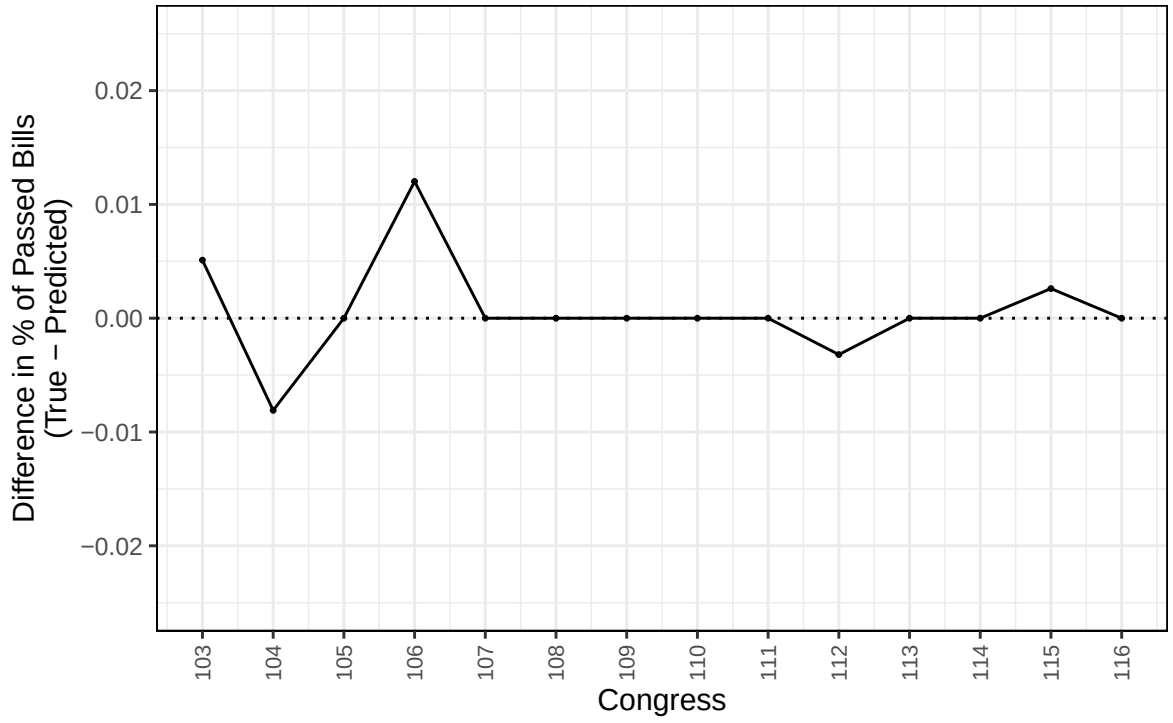


Note: ROC curves showing the performance of a classification model at all possible classification thresholds. Gray lines represent ROC curves based on each Congress-member majority data. Black line represent ROC curves based on the entire data.

B.4 Bill Outcomes

Using a threshold of a simple majority of yea votes in the House, we compute total number of passed bills for each Congress based on the predicted votes, and compare against the actual proportion of bills that passed in the House. The result is shown in Figure 10. Figure 10 shows the difference in the proportion of bills identified as passed based on actual and predicted votes. The proportion of bills predicted to have passed by the House majority from the prediction model closely tracks the actual share of bills that received the majority of House votes. The maximum difference is 1.2 percentage point in the 106^{th} Congress, which is consistent with the finding in Figure 1 in the main text.

Figure 10: Prediction Accuracy on Bill Outcome



Note: The difference in the proportion of bills that exceeds a simple majority in the House. Positive values indicate that the prediction under-estimates the true share of bills that achieved a simple majority. Negative values indicate that the prediction over-estimates the true share.

B.5 Error Rate on Older Versions of Floor Bills

Table 8: Classification Error on Older Versions of Floor Bills

Congress	Overall Error			False Negatives			False Positives		
	All	Majority	Minority	All	Majority	Minority	All	Majority	Minority
103	0.113	0.059	0.191	0.034	0.002	0.109	0.402	0.824	0.303
104	0.111	0.053	0.181	0.029	0.005	0.071	0.477	0.937	0.394
105	0.118	0.062	0.180	0.015	0.002	0.033	0.650	0.850	0.590
106	0.098	0.053	0.145	0.018	0.004	0.035	0.632	0.848	0.569
107	0.072	0.037	0.109	0.010	0.006	0.014	0.679	0.759	0.656
108	0.066	0.025	0.111	0.010	0.0004	0.022	0.609	0.877	0.558
109	0.074	0.036	0.118	0.012	0.005	0.023	0.565	0.894	0.498
110	0.076	0.017	0.145	0.013	0.003	0.030	0.542	0.853	0.517
111	0.059	0.015	0.122	0.014	0.0001	0.039	0.457	0.901	0.411
112	0.097	0.045	0.164	0.025	0.006	0.060	0.414	0.870	0.343
113	0.093	0.032	0.166	0.021	0.003	0.050	0.462	0.866	0.412
114	0.076	0.022	0.149	0.021	0.001	0.063	0.332	0.797	0.288
115	0.093	0.018	0.187	0.015	0.001	0.041	0.504	0.797	0.481
116	0.088	0.008	0.185	0.028	0.0001	0.087	0.328	0.871	0.313
All	0.085	0.033	0.149	0.017	0.003	0.041	0.496	0.853	0.440

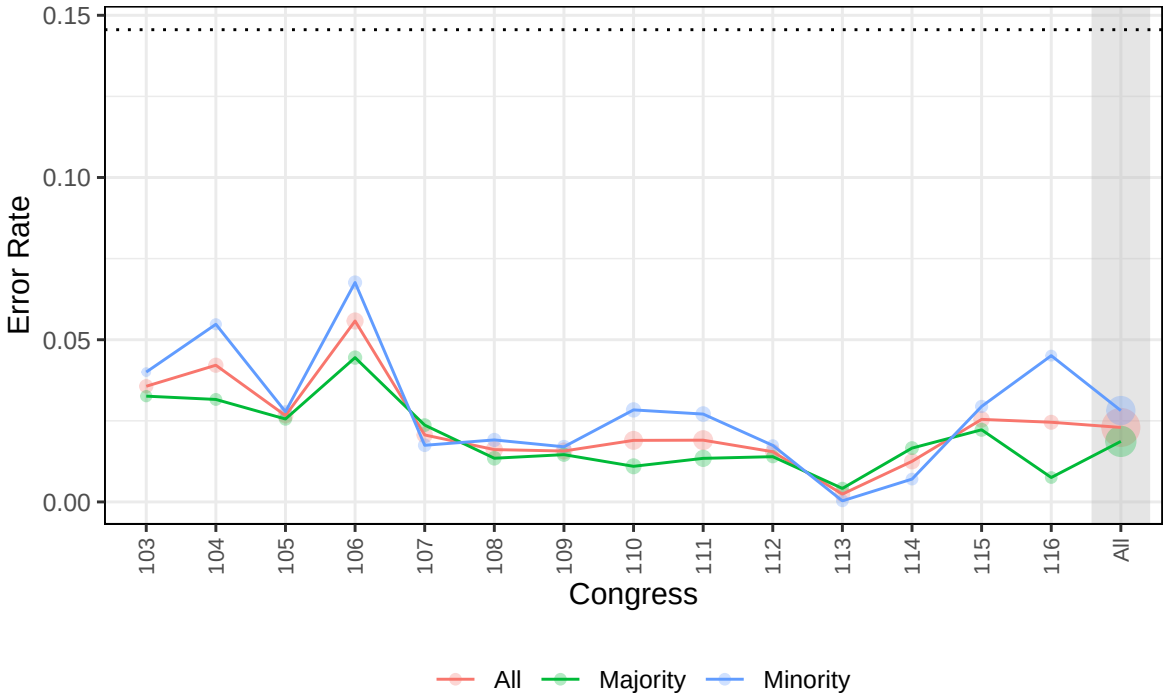
Note: The overall classification rate as well as false negative (Type I error) and false positive (Type II error) rates across Congress and majority status. The total sample size is 3.9 million.

C Alternative Prediction Model

We fit an alternative prediction model, which is an ensemble of logistic, LASSO, and two random forests. The first random forests grows 600 trees while randomly selecting $\sqrt{p}$ covariates and the second random forests grows 500 trees while randomly selecting $p/2$ covariates. Rather than running the model separately by Congress and the majority party status, we run it separately by year, while including the majority party status as an additional covariate.

The bill-level predictors include (1) the doc2vec vector of neural weights; (2) the number of co-sponsors; and (3) bill types (e.g., House Joint Resolutions, House Simple Resolutions, etc.). The member-level predictors include (1) the individual member indicator variables; (2) the member's margin of victory in the most recent general election; (3) an indicator for whether the member is the bill's sponsor; (4) an indicator for whether the member is the bill's co-sponsor; (5) an indicator for whether the sponsor is in the majority party; (6) indicator variables for each committee that members are assigned to; (7) state that the member represents; and (8) an indicator for whether the member is in the majority party.

Figure 11: Prediction Model Classification Error Rate



Note: Classification error rate on floor bills. The overall classification error rate across all Congresses is 0.02. The black dotted line represents the baseline error rate (i.e., share of nay votes). The size of points represent total number of floor bills introduced in each Congress.

D IRT Model Validation Checks

D.1 Data Generating Process of Vote Preferences

One conceptual issue in our ideal point estimation approach is that we assume two distinct data generating processes of members' votes: 1) a complex nonparametric model based on member- and bill-level variables; and 2) an IRT model based on spatial voting. To make the two models compatible, we conceptualize our prediction model as a general function of member- and bill-level characteristics while assuming that the members' ideal points represent an unknown function of such member- and bill-level parameters we include in our prediction model. Specifically, given the spatial voting model for a member i and bill b:

$$Pr(y_{ib} = 1) = \beta_b^T \theta_i - \alpha_b \tag{1}$$

as well as a prediction model with legislator-level covariates, $\mathbf{X_i}$, as well as bill-level covariates, $\mathbf{W_b}$:

$$Pr(y_{ib} = 1) = f(\mathbf{X_i}, \mathbf{W_b}) \tag{2}$$

the ideal point parameters θ_i can be thought as unknown functions g of $\mathbf{X_i}$ and $\mathbf{W_b}$:

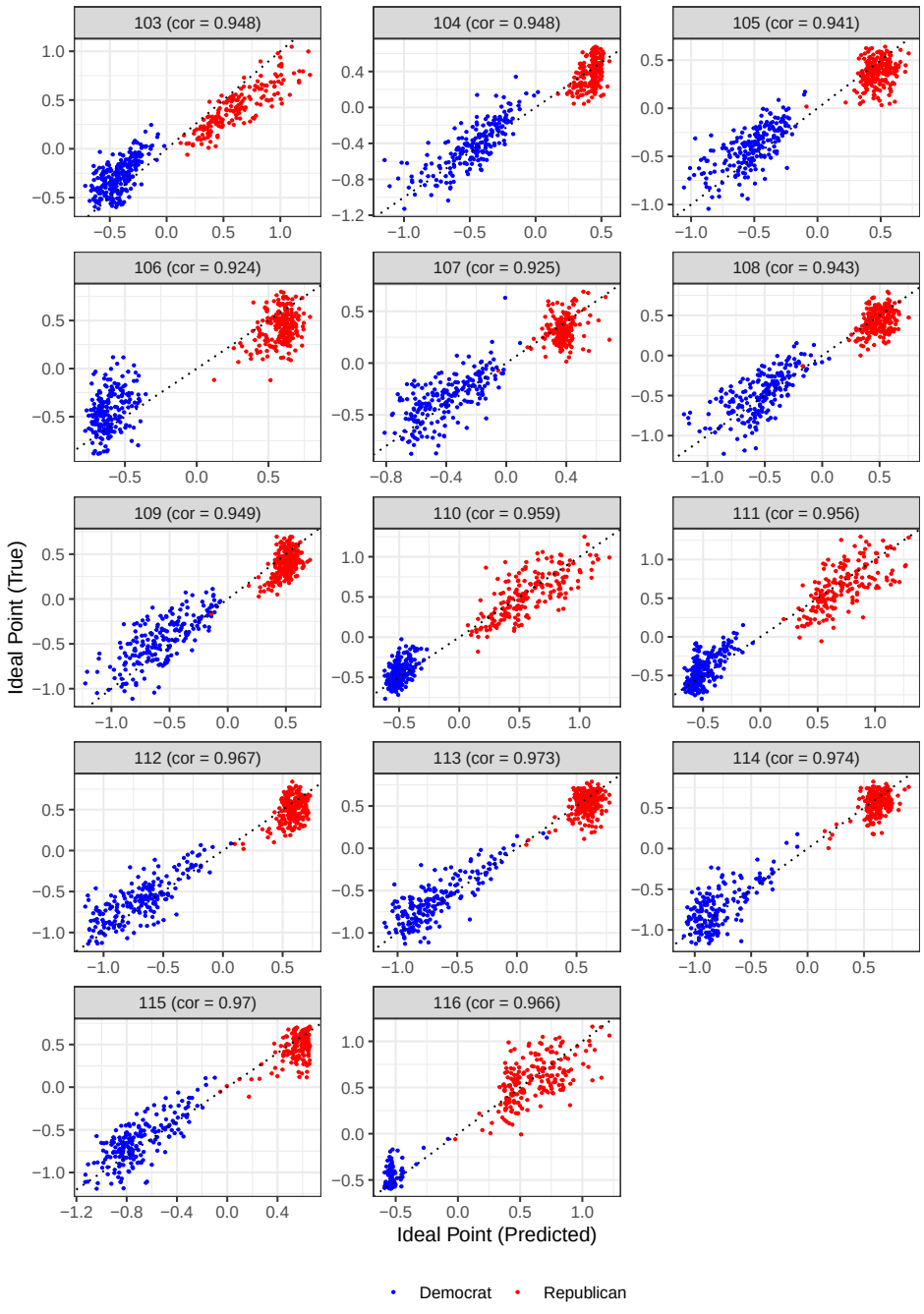
$$Pr(y_{ib} = 1) = f(\mathbf{X_i}, \mathbf{W_b}) \tag{3}$$

$$= \underbrace{g_1(\mathbf{W_b})^T g_2(\mathbf{X_i})}_{=\beta_b^T} - \underbrace{g_3(\mathbf{W_b})}_{=\alpha_b} \tag{4}$$

Moreover, in our view, ideal point models are simplifications of the complexities inherent in congressional voting and do not necessarily represent the "true" decision process underlying congressional voting. Instead, we recognize that these models are useful in summarizing congressional behavior.

D.2 Correlation Analyses of Ideal Point Estimations

Figure 12: Correlations between True and Predicted Ideal Point Estimates



Note: Correlation between ideal points generated with predicted (x-axis) and true (y-axis) votes on floor bills. The overall correlation is 0.954.