doi:xx.xxxx/xxxxx

Disagreement and Dimensionality: A Varying Dimensions Approach to Roll Call Scaling in the U.S. Congress

KEVIN MCALISTER *University of Michigan*

The dimensionality of ideal points is an aspect of roll call scaling which has received significant attention due to its impact on both substantive and spatial interpretations of estimates. In this article, I find that previous evidence for unidimensional ideal points is a product of the Scree testing procedure. I propose a new varying dimensions model of legislative voting and a corresponding Bayesian nonparametric estimation procedure (BPIRT) that allows for probabilistic inference on the number of dimensions. Using this approach, I show that there is strong evidence for multidimensional ideal points in the U.S. Congress and that using only a single dimension misses much of the disagreement that occurs within parties. Using BPIRT, I reexamine theories of U.S. legislative voting and find that empirical evidence for these models is conditional on unidimensionality. This article provides a framework for new examinations into the role of multidimensionality in studies of legislative behavior.

Word Count: 10,478

Kevin McAlister is a PhD Candidate, Department of Political Science, University of Michigan, 48103 MI. (kamcal@umich.edu)

I am grateful to many colleagues for their helpful comments. I thank Walter Mebane, Kevin Quinn, Christopher Fariss, Yuki Shiraito, Patrick Wu, Diogo Ferrari, the participants of the University of Michigan Interdisciplinary Workshop on American Politics, and the participants of the University of Michigan Statistical Learning Workshop for their helpful suggestions.

This is a manuscript submitted for review.

INTRODUCTION

Studies of legislative behavior focus upon the relationship between legislative preferences, institutional structure, and legislative outcomes. A common method used to better understand these relationships utilizes scaling models that uncover the ideal points of legislators. While there are many approaches to uncovering the ideal points of legislators, by far the most common approach uses the outcomes from the various roll call votes that are cast by members of Congress. Roll call scaling techniques such as NOMINATE (Poole and Rosenthal 1997) and its Bayesian analogue (Clinton et al. 2004) seek to project roll call data into a low-dimensional policy space that captures the complexities of how members of Congress make vote decisions. The ideal points can then be used to make comparisons of various behaviors between different members of the legislatures, such as the role of parties (Aldrich and Rohde 2000; Cox and Poole 2002; Cox and McCubbins 2005), influences within and between branches of the U.S. government (Binder 1999; Krehbiel 1998), and other features of the legislative institution.[1]

Ideal point models require assumptions that have implications for the interpretation of the estimated quantities. One such assumption is the dimensionality of the latent space. Assuming a unidimensional ideal point, legislators behave predictably and rational choice models can provide simple explanations of how legislators make policy proposals and vote choices under the rules of the institution (Krehbiel 1998; Cox and McCubbins 2005). On the other hand, multidimensional ideal points create an environment where legislators behave in a more nuanced manner - legislative behavior is conditional on the context of the vote and there are few guarantees of predictable outcomes (McKelvey 1976; Schofield 1978; Shepsle 1978). The choice of dimensionality also has strong substantive implications (Harbridge 2015; Lee 2009).

In most recent studies that leverage unidimensional estimates of ideal points, empirical justification for this assumption is given by referencing the "Unidimensional Congress" arguments of Poole and Rosenthal (2011) - across issue areas within an analyzed period of the U.S. Congress, little improvement to the overall fit of the roll call scaling model can be made by including more than one dimension. However, many works show evidence for multidimensionality in U.S. Congressional roll call voting; there is strong evidence that certain bundles of votes map to different dimensions and less aggregated analysis of roll call votes reveals this heterogeneity (Heckman and Snyder Jr 1996; Roberts et al. 2016; Smith 2007; Hurwitz 2001; Crespin and Rohde 2010; Norton 1999; Bateman et al. 2017).

If U.S. Congressional voting appears to be multidimensional and linked to specific vote topics in less aggregated studies of roll call voting, why does aggregate roll call scaling show strong evidence for only a single liberal-conservative dimension? In this paper, I seek to solve this puzzle. I present a method of roll call scaling that allows for aggregate-level summaries of legislative decision making while also allowing for examinations of multidimensionality at the bill-episode level. Leveraging work related to from Aldrich et al. (2014) and Roberts et al. (2016), I contend that evidence for the low-dimensional conjecture is due to the statistical tests used to assess inclusion of new dimensions. To address this problem, I present a new spatial model in which a voter makes vote decisions using both the positions of alternatives within the policy space and a vote-specific bundle of dimensions in which those policy positions exist. The corresponding empirical model allows for rigorous statistical inference related to the overall dimensionality of the ideal points and identification of the dimensions of the policy

[1]There has been significant work in the area of roll call scaling and ideal point estimation for the U.S. Congress beyond these two models. Lauderdale and Herzog (2016), Tausanovitch and Warshaw (2017), Bonica (2014), Ramey (2016), Jessee and Malhotra (2010), and Tahk (2018) are just a few of the models proposed in recent literature.

space associated with each vote. Unlike previous approaches, this method accurately estimates the dimensionality of the ideal point space even under high levels of party bloc voting. Similarly, the ideal points and vote-level estimates of dimensionality allow for new tests related to theories of legislative behavior that properly take vote-level dimensionality into account. Given that the vast majority of empirical tests related to legislative decision making use unidimensional NOMINATE scores, the estimates achieved from this new model allow for a finer examination of the role of dimensionality in many important theories of U.S. legislative behavior.

Overall, I make several important contributions to the literature in this article. Methodologically, I present a new spatial model of voting that has an explicit empirical analogue under assumptions about utility structures. This model uses novel advancements in the field of Bayesian nonparametrics related to estimating the infinite latent feature model (Paisley and Carin 2009; Knowles and Ghahramani 2011) to address the question of dimensionality in aggregate sets of roll call votes by simultaneously estimating ideal points and dimensionality. Substantively, I analyze the entire history of the U.S. House and U.S. Senate ($1^{st} - 115^{th}$ sessions) and show that there is strong evidence of multidimensional voting through the history of the U.S. Congress. In line with many of the conclusions by Heckman and Snyder Jr (1996), I find that votes tend to bundle based on topic and these votes share similar multidimensional vote patterns. In turn, this allows for identification of key issues that split members of Congress, both within and across parties. This work produces a new set of ideal points across U.S. Congressional history that should provide a starting point for further work related to topic-level voting in the U.S. legislative body. Finally, I apply the estimates from the new roll call scaling method to two specific theories related to U.S. Congressional voting: the pivotal voter model (Krehbiel 1998) and the party cartel model (Cox and McCubbins 2005). I show that much of the empirical evidence that exists for these models changes when dimensionality is properly accounted for in empirical tests of these theories. This analysis is just the starting point for potentially reassessing many other predictions made by models of U.S. legislative voting under conditions of multidimensionality.

A SPATIAL MODEL OF ROLL CALL VOTING

For a legislature, assume there are N voting members that cast P votes over the course of time analyzed. For any given vote $j \in (1, P)$, legislator $i \in (1, N)$ must choose between two alternatives: to approve the proposed alternative ($\boldsymbol{A}_j$) or to reject the proposed alternative ($\boldsymbol{S}_j$). Behavior in this legislature is assumed to be describable in a K-dimensional policy space - all votes that are made by legislator i can be described by the K-dimensional point locations of $\boldsymbol{A}_j$ and $\boldsymbol{S}_j$ within the space and a K-dimensional ideal point, ω_i, which encapsulates the policy preferences of legislator i.

A legislator must choose whether to vote for $\boldsymbol{A}_j$ or $\boldsymbol{S}_j$. Using a utility maximization model that assumes quadratic loss in distance from her ideal point, assume that she chooses the alternative which grants the highest utility:

$$\begin{aligned} U_i(\boldsymbol{A}_j) &= -\|\omega_i - \boldsymbol{A}_j\|^2 + \eta_{i,j} \\ U_i(\boldsymbol{S}_j) &= -\|\omega_i - \boldsymbol{S}_j\|^2 + \nu_{i,j} \end{aligned} \tag{1}$$

where $\eta_{i,j}$ and $\nu_{i,j}$ are stochastic elements of the utility functions. This model is completely specified if a known structure is placed on $\eta_{i,j}$ and $\nu_{i,j}$ (Heckman and Snyder Jr 1996; Poole and Rosenthal 1997; Clinton et al. 2004).

Let $\boldsymbol{Y}$ be a matrix of roll call votes and $y_{i,j}$ be the vote choice that legislator i makes on proposal j : $y_{i,j} = 1$ if legislator i votes "Yea" on vote j and $y_{i,j} = 0$ if she casts a "Nay" vote. Given the model construction, the probability that legislator i votes for $\boldsymbol{A}_j$ can represented as:

$$P(y_{i,j} = 1) = \Xi(\boldsymbol{\lambda}_j' \boldsymbol{\omega}_i - \alpha_j) \tag{2}$$

where $\Xi(\cdot)$ is the CDF associated with the chosen error structure, $\alpha_j = \frac{\boldsymbol{A}_j'\boldsymbol{A}_j - \boldsymbol{S}_j'\boldsymbol{S}_j}{\sigma_j^2}$, and $\boldsymbol{\lambda}_j = \frac{2(\boldsymbol{A}_j - \boldsymbol{S}_j)}{\sigma_j^2}$.

This construction admits a corresponding statistical model that allows for estimation of the *structural parameters* $\boldsymbol{\alpha}$ and $\boldsymbol{\Lambda}$ and the latent variables, $\boldsymbol{\Omega}$. Assuming the errors are independent and identically distributed, a likelihood function can be derived:

$$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\Lambda}, \boldsymbol{\Omega} | \boldsymbol{Y}) = \prod_{i=1}^{N} \prod_{j=1}^{P} \Xi(\boldsymbol{\lambda}_j' \boldsymbol{\omega}_i - \alpha_j)^{y_{i,j}} \times \left(1 - \Xi(\boldsymbol{\lambda}_j' \boldsymbol{\omega}_i - \alpha_j)\right)^{1 - y_{i,j}} \tag{3}$$

Bayesian implementations of this model place priors on all of the structural parameters and estimation proceeds using Markov Chain Monte Carlo methods (Clinton et al. 2004). With other minor theoretical changes, this model is equivalent to the NOMINATE procedure (Poole and Rosenthal 1997).

Under this specification, a choice of the number of dimensions, K, leads to fully tractable model that can be estimated. Currently, this choice is made by performing a Scree test (Cattell 1966). The spirit of the Scree test revolves around estimating the latent variable model under a number of different assumptions for the number of dimensions and plotting the fit metric to find the "elbow" in the plot. Once adding a new dimension no longer adds "enough" value to the fit metric, then no new dimensions are added. Typically, the choice of fit metric revolves around the proportion of votes correctly classified under a model: aggregate proportion reduction in error (Poole and Rosenthal 2011) or marginal proportion reduction in error (Roberts et al. 2016).

Regardless of the choice of metric, I contend that the Scree test presents many problems for inference. By its very nature, the Scree test is inherently subjective; the choice of how much reduction in error is enough to include a new dimension is subjective and can lead to biases in the number of dimensions chosen. For example, the Scree test cannot detect small improvements in model fit that are due to adding dimensions that only contribute to a few votes. Rather, the Scree test says that the model improvement over the aggregated set of roll call votes is small and the dimension should not be included. This feature of the Scree test is not ideal as there is an entire body of the literature which shows that dimensions in roll call voting appear at the vote-topic level and important dimensions can appear infrequently (Roberts et al. 2016).

Along the same lines, Aldrich et al. (2014) and Roberts et al. (2016) point to the high frequency of votes that occur along party lines as a problem for current dimensionality testing procedures. When many votes are explained by party lines, the perceived influence of party can be much higher than what is actually present within the data. Given that there is often correlation between vote choices in specific policy domains and party membership, party bloc voting can appear to account for all of the explainable variation within roll call voting data sets when many dimensions are truly influencing decisions. This result is corroborated by the finding that scaling within parties reveals many dimensions even when using Scree tests as the decision making criterion (Aldrich et al. 2014). Similarly, multidimensionality is highly apparent when dimensionality is tested within and across topically similar bill-episodes (Roberts et al. 2016).

These findings point to a couple of features that a roll call scaling method that accurately uncovers dimensionality should have:

1. Dimensionality should be tested under distributional assumptions. In turn, probabilistic tests of whether or not a dimension provides a non-zero improvement to the model under an assumption about what constitutes random noise can be performed.
2. Dimensionality should be tested at the vote level. Each vote should be allowed to draw on a different set of dimensions, if necessary. The aggregated set of vote-level dimensionalities then dictates the dimensionality of the ideal point. However, each set of vote-level dimensions should be subject to overfitting penalties in order to estimate substantively useful parameters that account for both vote-level and aggregate behavior of legislators over the course of time analyzed.

A roll call scaling method that meets these conditions should provide an accurate representation of the dimensionality of the data while also reducing the dimensionality of the data to something useful for further examination of aggregate legislative voting behavior.

A ROLL CALL SCALING MODEL WITH VARYING DIMENSIONS

A Spatial Model of Voting with Varying Dimensions

To address the above conditions for accurately estimating dimensionality, I propose a new roll call scaling model with varying dimensions. As before, a legislator must choose to vote for $\boldsymbol{A}_j$ or $\boldsymbol{S}_j$. She chooses to cast a vote for the alternative that maximizes her utility under a quadratic loss function such that:

$$\begin{aligned} U_i(\boldsymbol{A}_j) &= -\|\boldsymbol{r}_j(\omega_i - \boldsymbol{A}_j)\|^2 + \eta_{i,j} \\ U_i(\boldsymbol{S}_j) &= -\|\boldsymbol{r}_j(\omega_i - \boldsymbol{S}_j)\|^2 + \nu_{i,j} \end{aligned} \tag{4}$$

Under this specification, the new addition is the binary vector $\boldsymbol{r}_j$. $\boldsymbol{r}_j$ is a vector of length K where $r_{j,k} = 1$ if she considers dimension $k \in (1, ..., K)$ in vote j. On the other hand, $r_{j,k} = 0$ if she does not utilize her ideal point on dimension k when making a decision for vote j. Note that $\boldsymbol{r}_j$ is assumed to be globally known to all legislators.

One advantage of this approach is that the length of $\boldsymbol{r}_j$ does not need to be explicitly set before specifying the model. For example, if only dimensions one and three are needed to dictate the utility function associated with a vote (i.e. $r_{j,1} = 1$, $r_{j,2} = 0$, $r_{j,3} = 1$), then this vector is equivalent to one where $r_{j,4} = 0$, $r_{j,5} = 0$, and all subsequent elements of the vector are set to zero. Thus, the vector of length three and the corresponding vector of infinite length are equivalent. This characteristic of the binary vector is key to addressing the shortcomings of the standard roll call scaling model.

Placing all of the vote level binary vectors, $\boldsymbol{r}_j$, into a matrix with the number of rows equal to the number of votes and the number of columns equal to the number of dimensions creates a binary matrix, $\boldsymbol{R}$, that dictates the mapping of individual votes to ideal point dimensions. $\boldsymbol{R}$ captures the dimensionality of the underlying ideal point space across all votes analyzed. Recall that each $\boldsymbol{r}_j$ can be of infinite size, but only the elements equal to one matter for the underlying utility model. Thus, the dimensionality of the overall space can be modeled as the number of columns in $\boldsymbol{R}$ which have at least one non-zero element if $\boldsymbol{R}$ has a known probability measure. Similarly, the structure of $\boldsymbol{r}_j$ is allowed to vary across votes - each vote can call on a different set of dimensions to construct the parameters of the assumed utility calculations that lead to vote decisions.

As with the standard model, the construction admits a corresponding statistical model. With similar rearrangement, a likelihood function is determined:

$$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\Lambda}, \boldsymbol{\Omega}, \boldsymbol{R}|\boldsymbol{Y}) = \prod_{i=1}^{N}\prod_{j=1}^{P} \Xi\left((\boldsymbol{r}_j \odot \boldsymbol{\lambda}_j)'\omega_i) - \alpha_j\right)^{y_{i,j}} \times \left(1 - \Xi\left((\boldsymbol{r}_j \odot \boldsymbol{\lambda}_j)'\boldsymbol{\omega}_i - \alpha_j\right)\right)^{1-y_{i,j}} \tag{5}$$

where $\odot$ is the Hadamard product of two vectors.[2] As with the likelihood function for the standard roll call scaling model, the likelihood is comprised of the structural parameters $\boldsymbol{\Lambda}$ and $\boldsymbol{\alpha}$ and the ideal points, $\boldsymbol{\Omega}$. $\boldsymbol{R}$ is assumed to modify the loadings, $\boldsymbol{\Lambda}$. Like the standard model, the parameters of the estimated statistical model explicitly link back to the formal theoretic foundation where $\alpha_j = \frac{\boldsymbol{r_j}(\boldsymbol{A}_j'\boldsymbol{A}_j - \boldsymbol{S}_j'\boldsymbol{S}_j)}{\sigma_j^2}$ and $\boldsymbol{\lambda}_j = \frac{2\boldsymbol{r}_j(\boldsymbol{A}_j - \boldsymbol{S}_j)}{\sigma_j^2}$.[3]

The varying dimensions model of voting explicitly adds the two conditions listed previously. First, $\boldsymbol{R}$ constitutes a new quantity that can be estimated. With distributional assumptions, $\boldsymbol{R}$ can be estimated with the other structural parameters of the ideal point model to determine the number of dimensions needed to effectively model the ideal point space. If this choice penalizes against adding many dimensions, $\boldsymbol{R}$ can dictate a sparse set of dimensions that directly model the aggregate set of roll call votes. Second, each binary vector, $\boldsymbol{r}_j$, contains a mapping of each vote to some subset of the ideal point space. This allows each vote to modeled in a potentially different set of dimensions. Again, under proper distributional assumptions about $\boldsymbol{R}$, this allows each vote to be modeled as a subset of aggregate dimensions. However, if only one dimension is needed for the collection of votes, $\boldsymbol{R}$ can be reduced to estimate only one dimension. All in all, this model of voting allows for the dimensionality of the latent space to be estimated simultaneous to the structural parameters of the ideal point model.

Estimating the Roll Call Scaling Model with Varying Dimensions

Using equation (5) as a starting point, a Bayesian scaling procedure for binary roll call votes with varying dimensions can be defined.[4] Let $\boldsymbol{X}$ be a continuous latent mapping of the observed roll call votes $\boldsymbol{Y}$ such that:

$$x_{i,j} \sim \begin{cases} \mathcal{TN}_{-\infty,0}((\boldsymbol{r}_j \odot \boldsymbol{\lambda}_j)\boldsymbol{\omega_i} - \alpha_j, 1) \text{ if } y_{i,j} = 0 \\ \mathcal{TN}_{0,\infty}((\boldsymbol{r}_j \odot \boldsymbol{\lambda}_j)\omega_i - \alpha_j, 1) \text{ if } y_{i,j} = 1 \\ \mathcal{N}((\boldsymbol{r}_j \odot \boldsymbol{\lambda}_j)\omega_i - \alpha_j, 1) \text{ if } y_{i,j} \text{ is missing} \end{cases} \tag{6}$$

where $\mathcal{TN}_{l,u}(\mu, \sigma^2)$ is the truncated normal distribution truncated between l and u. Assuming a probit structure on the errors and, without loss, an infinite dimension ideal point space, define the varying dimension roll call scaling model as:

$$P(y_{i,j} = 1|-) = \int_0^\infty \mathcal{N}\left(x_{i,j}; \sum_{k=1}^{\infty} r_{j,k}\lambda_{j,k}\omega_{i,k} - \alpha_j, 1\right) dx_{i,j} \tag{7}$$

[2]The standard roll call scaling models is a special case of the varying dimensions version - setting all elements of $\boldsymbol{R}$ equal to one and fixing K to a known finite value replicates the model of Clinton et al. (2004).

[3]Like the Bayesian IRT approach of Clinton et al. (2004) and the NOMINATE model, this model assumes that all voters vote sincerely based on their underlying ideal point. Using a model that ties abstention to strategic behavior, this assumption could be changed.

[4]The benefits of approaching the roll call scaling problem under the Bayesian paradigm are well documented. A thorough discussion of these benefits are presented by Clinton et al. (2004).

This voting model shares many parameters with the two-parameter item response model used in educational testing where $\boldsymbol{\lambda}_j$ is a vector of item discrimination parameters, α_j is the item difficulty parameter, and $\boldsymbol{\omega}_i$ is a vector of ideal points associated with the vote decisions made by a legislator (Poole and Rosenthal 1997; Londregan 1999).

Estimating the Binary Matrix

A number of approaches for selecting the appropriate number of dimensions in latent variable models have appeared in the statistics and social sciences literatures (Kim et al. 2018). However, these approaches often are developed for the purpose of finding a reduced dimensionality representation of the covariance matrix and estimation of the structural parameters and item-level differences in dimensionality are not addressed. I choose to use nonparametric priors on the number of dimensions and allow them to stochastically vary. These priors have strict probabilistic properties that make identification and estimation of structural parameters plausible (Bhattacharya et al. 2011).

One Bayesian nonparametric approach which allows probabilistic modeling of both the overall and vote-level dimensionalities uses the beta process (Paisley and Carin 2009). For a given element of $\boldsymbol{R}$, let the prior be:

$$\begin{aligned} P(r_{j,k}|\pi_{j,k}) &\sim \text{Bern}(r_{j,k};\pi_{j,k}) \\ P(\pi_{j,k}|\alpha_{j,k},\beta_{j,k}) &\sim \text{Beta}(\pi_{j,k};\alpha_{j,k},\beta_{j,k}) \end{aligned} \tag{8}$$

Letting $K \to \infty$ allows all possible dimensions to be potentially present in $\boldsymbol{R}$ and constitutes a beta process.

Without further constraint, the model will always find that the optimal number of features is equal to the number of items - each item is modeled by its own dimension. This outcome is akin to overfitting in regression and provides a solution that is not useful for summarizing high-dimensional data; the roll call scaling model needs a *sparse* estimate of $\boldsymbol{R}$. Along with other challenges related to fitting an infinite set of dimensions to a finite set of data, estimation under the beta process prior has proven a difficult task in the statistics literature.

A marginalized approach exists that allows for a relatively simple sampling scheme in the infinite limit that prevents against over fitting by allowing the number of dimensions found to scale with the number of observations and items (Paisley and Carin 2009). Here, the beta-Bernoulli process dictating the values of the infinite matrix, $\boldsymbol{R}$, has a prior such that:

$$\begin{aligned} P(\pi_k) &\sim \text{Beta}\left(\pi_k;\frac{a}{K},\frac{b(K-1)}{K}\right) \\ P(r_{j,k}|\pi_k) &\sim \text{Bern}(r_{j,k};\pi_k) \end{aligned} \tag{9}$$

where a and b are hyperparameters and K is arbitrarily large such that:

$$E[\pi_k|K] \approx 0 \tag{10}$$

which induces a sparse estimate of the binary matrix.

The beta process prior of this form constitutes a simple stochastic process. Marginalizing over π, this process is the *two parameter Indian Buffet Process* (IBP) (Ghahramani and Griffiths 2006). IBP has two notable properties for modeling sparse matrices. First, over the entire set of P votes, the number of dimensions sampled follows:

$$
\begin{aligned}
P(K^+ = h) &= \text{Pois}\left(h; \sum_{j=1}^{P} \frac{ab}{b+j-1}\right) \\
E[K^+] &\approx O(\ln(P))
\end{aligned}
\tag{11}
$$

where K^+ is the number of columns of $\boldsymbol{R}$ with at least one element equal to one. This shows that as P increases, the number of dimensions that can potentially appear in the latent space increases. However, the expected number of dimensions is small relative to P and a sparse solution is ensured. On the other hand, if P is small, then the number of dimensions that can potentially appear in the latent space is also small. This property ensures that the number of dimensions estimated is supported by the amount of data present during estimation.

Second, IBP exhibits a "rich get richer" property:

$$
P(r_{j,k} = 1) \propto \frac{-r_{j,k} + \sum\limits_{h=1}^{P} r_{h,k}}{P + r_{j,k} - \sum\limits_{h=1}^{P} r_{h,k} + 1} \tag{12}
$$

As a dimension becomes more popular, the probability that it is sampled in other votes increases. In turn, features that are not popular are rarely sampled and have a small chance of being included in the final model specification. Thus, IBP explores the feature space in accordance with P, but still promotes sparsity by allowing popular dimensions to dominate the feature space. This allows the IBP prior to prevent against overfitting. However, in the face of strong statistical evidence, IBP still allows an unpopular feature to emerge.

A Beta Process IRT Model

Under the specification in (7), $\boldsymbol{r}_j$ induces a *spike and slab* prior on the vector of discrimination parameters, $\boldsymbol{\lambda_j}$. Placing an Indian Buffet Process prior on $\boldsymbol{R}$, the induced prior on $\lambda_{j,k}$ is:

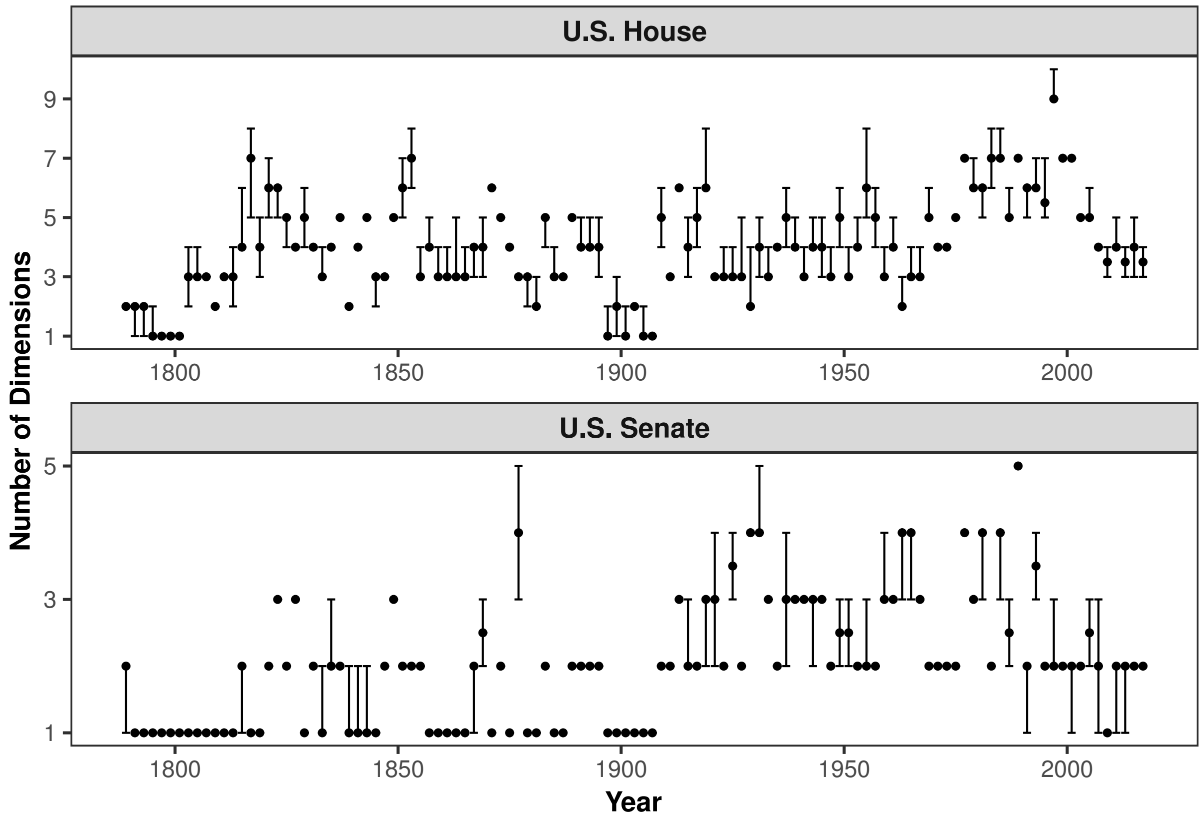
$$
P(\lambda_{j,k} | r_{j,k}) \sim r_{j,k}\ P(\lambda_{j,k}) + (1 - r_{j,k})\delta_0 \tag{13}
$$

where δ_0 is a point mass PDF at zero and $P(\lambda_{j,k})$ is the marginal prior on $\lambda_{j,k}$. Thus, if $r_{j,k} = 1$, $\lambda_{j,k}$ is allowed to take a non-zero value. On the other hand, if $r_{j,k} = 0$, it is restricted to be equal to zero. In the context of roll call scaling, this is equivalent to estimating whether or not a vote draws on a specific dimension when estimating the parameters of the underlying utility model.

Using the IBP prior on $\boldsymbol{R}$ allows for a full model definition, which is outlined in Section A.2 of the Appendix. The full model, a beta process item-response theory model (BPIRT), is a close analogue to the *infinite latent feature model* developed by Knowles and Ghahramani (2011). BPIRT has many of the same properties as the standard Bayesian IRT model (Clinton et al. 2004).

Under this specification, estimation of the model using Markov Chain Monte Carlo methods proceeds in a relatively straightforward manner. Technical details related to estimation of the BPIRT model are included in the appendix. The full conditional distributions and sampling methods are outlined in Section A.3. Methods for determining good starting values and assessing convergence of the Markov chains are outlined in Section A.5 and Section A.6. The BPIRT model also uniquely identifies estimates for the structural parameters; this property is discussed in Section A.4. Finally, simulations which how accurately BPIRT uncovers the binary matrix under a known model are shown in Section A.7.

FIGURE 1. Dimensionality and Corresponding 95% HPD Intervals Estimated by BPIRT for the $1^{st} - 115^{th}$ Sessions of the U.S. House and the U.S. Senate.



MULTIDIMENSIONALITY IN THE U.S. CONGRESS OVER TIME

The varying dimensions model of voting is well suited to examining the question of multidimensionality in voting and BPIRT provides a rigorous statistical tool for creating estimates of dimensionality in roll call voting. The results related to multidimensionality from BPIRT promise to provide insight into this problem.[5] I begin by assessing how many dimensions are estimated in the $1^{st} - 115^{th}$ (1789 – 2017) sessions for each chamber of the U.S. Congress. Figure 1 shows the estimated number of dimensions and corresponding 95% highest posterior density intervals for each session within each chamber of Congress. This plot shows a generally multidimensional legislature. In the case of the U.S. House, sessions near the beginning of U.S. history and some sessions in the late 1800s and early 1900s have credible intervals that include a single dimension. However, the vast majority of sessions are estimated to need more than one dimension to best model roll call behavior.

A similar story is seen in the U.S. Senate. While there are more sessions which are estimated to require only one dimension, the majority require at least two dimensions to best model the roll call voting variation. One important caveat for these unidimensional sessions is that the IBP prior which drives BPIRT is limited in its ability to estimate multidimensionality when there are a small number of votes and/or voters. In early sessions of the U.S. House, there were less voting members and there were typically less votes than in more recent sessions of the U.S. House. As shown in simulations, BPIRT underestimates dimensionality in these settings, so there is reason to believe that this is the case

[5]Additional figures and discussion for this section can be found in Section B of the appendix.

for the U.S. House. Similarly, there are always around 100 voting U.S. Senators in a given session which inherently places a cap on the number of dimensions which can be modeled for a session of this chamber. This is not to say that these results should be discarded - rather, it is important to point this aspect of BPIRT out as a weakness for estimating dimensionality in smaller sets of roll call data.

In order to examine multidimensionality at the vote level, a measure of multidimensionality must be established (Roberts et al. 2016; Bateman et al. 2017; Smith 2007). One method of summarizing the dimensionality of a single vote is to simply use the posterior probability that the vote took on more than one dimension in the binary matrix (MD). However, this metric suffers from minor theoretical deficiencies; it does not explain how much a vote needs each dimension. Thus, a second supplementary measure of multidimensionality is used - the proportion of variance explained (PVE) by each dimension on a specific vote:

$$\text{PVE}_{j,k} = \frac{r_{j,k}\lambda_{j,k}^2}{\sum\limits_{h=1}^{K} r_{j,h}\lambda_{j,h}^2} \tag{14}$$

Using PVE to examine the influence of the first dimension on a vote, PVE takes a value of one when only the first dimensions is needed. As the influence of other dimensions increase, PVE for the first dimensions decreases. Therefore, PVE measures the overall influence of a given dimensions on vote outcomes. MD and PVE are highly correlated, but provide different views of each dimension's necessity in the individual case.

Using the proposed measures of multidimensionality, I examine the role of the first dimension and the set of dimensions beyond the first estimated by BPIRT in explaining variation within U.S. House roll call data sets.[6] One of the many advantages of the BPIRT approach is that these metrics can be examined for any subset of votes within the analyzed sets. An application of the property is examining the difference between the aggregate set of all roll calls and more important "key votes" (Smith 2007; Roberts et al. 2016).

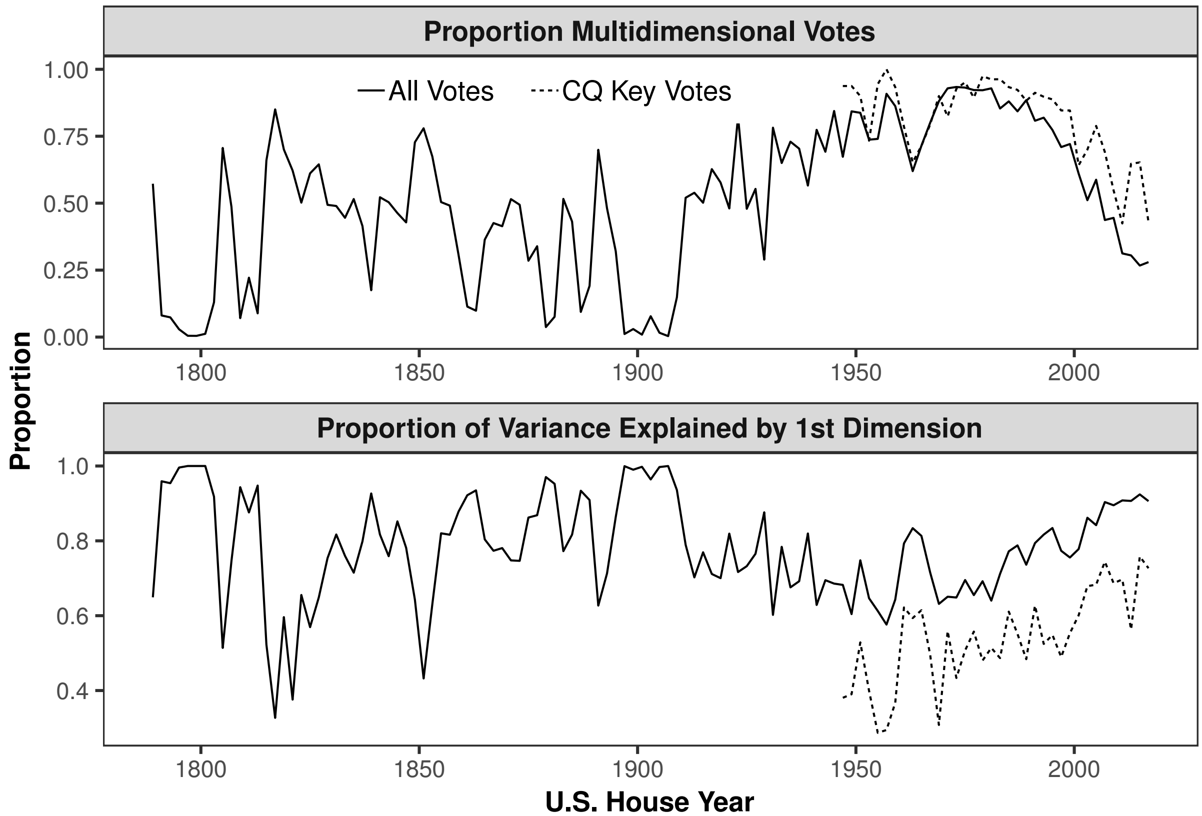
I examine MD and PVE for both the full aggregate sets of roll call votes for the $1^{st} - 115^{th}$ sessions of the U.S. House as well as the set of votes classified as "key votes" by Congressional Quarterly for the $80^{th} - 115^{th}$ sessions. Figure 2 shows these quantities over time for the U.S. House. Examining the proportion of multidimensional votes in each session shows highly multidimensional voting within the U.S. House, especially in the 20^{th} and 21^{st} centuries. Even in recent sessions of the U.S. House, which are considered to be extremely party driven and one-dimensional, a significant number of votes require more than one dimension to best explain variation in the roll call votes. On the other hand, the PVE for the first dimension is relatively high throughout time. While voting in recent sessions is certainly explained more heavily by the first dimension than in the mid and late 1900s, the reliance on the first dimension is equal to many non-unidimensional sessions during Reconstruction and the Great Depression.

Looking only at key votes provides support for the theory that important votes are multidimensional and require more than simply using the first dimension of ideal points.[7] While MD shows a modest

[6] For this and the proceeding examinations of multidimensionality in roll call voting, I choose to only present results for the U.S. House. The trends and inferences made from U.S. House data are similar to those that are made from U.S. Senate data. For the sake of brevity, I withhold figures and other summaries of the U.S. Senate data in this paper. Results from my analysis of the U.S. Senate can be seen in the replication files included with this paper.

[7] For each session where CQ key votes were examined, there were between 80 and 300 votes that were classified as important votes by Congressional Quarterly.

FIGURE 2. Proportion of Multidimensional Votes and Proportion of Variance Explained by the 1^{st} Dimension Estimated by BPIRT in the $1^{st}-115^{th}$ Sessions of the U.S. House



Note: Values reported are posterior means.

difference between the aggregate roll call sets and the set of key votes, PVE shows that a significantly lower amount of variance can be explained by the first dimension in key votes. On average, approximately 25% less variance is explained using the first dimension. This finding along with conflicting results for the aggregate roll call sets provides evidence for the aggregation hypotheses presented by Roberts et al. (2016) and should serve as a starting point for more fine-grained examinations of dimensionality in landmark legislation over time.

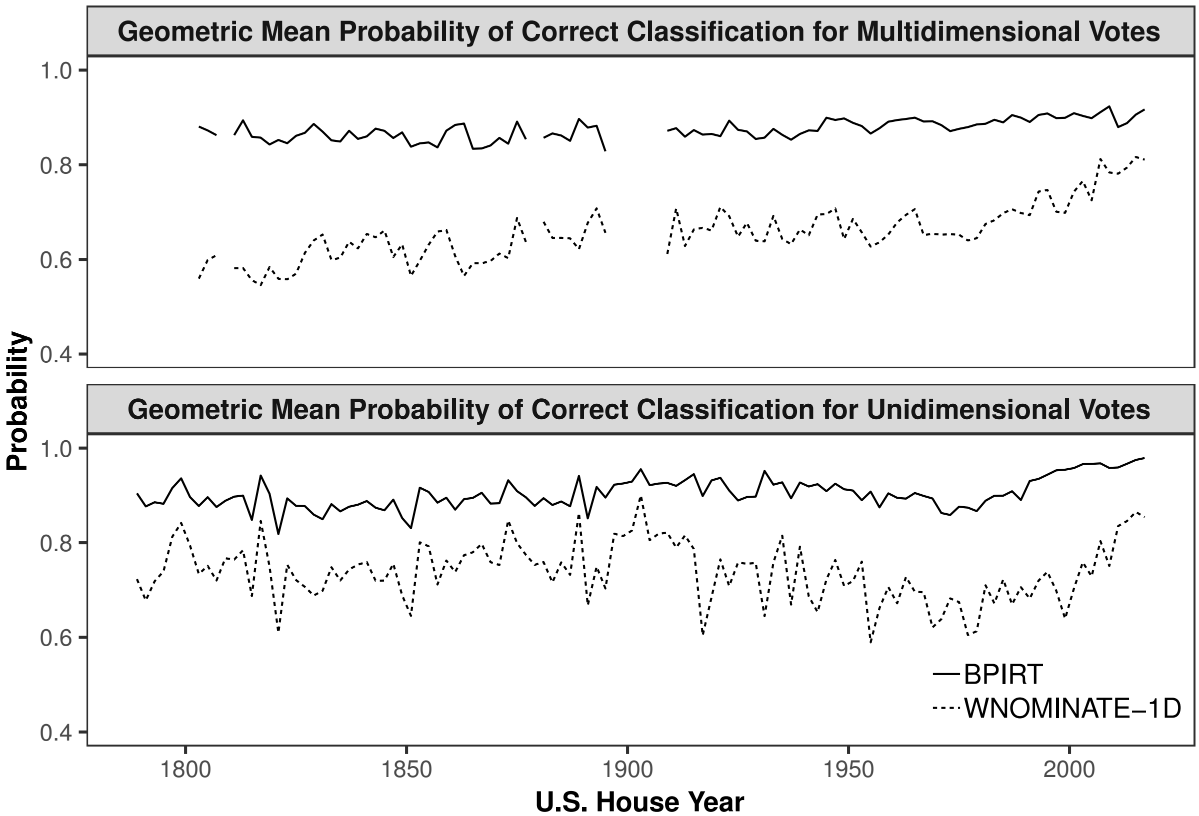
Examining vote level dimensionality is not the only way to demonstrate the necessity of dimensions past the first - BPIRT shows marked improvements over other roll call scaling techniques in terms of overall model fit. One method of comparison that rewards correct classification of model outcomes given the ideal points while also penalizing inefficient estimates is the geometric mean probability of correct classification (Carroll et al. 2009). For a given set of votes, the geometric mean probability of correct classification (GMP) is:

$$\text{GMP} = \left(\prod_{i=1}^{N} \prod_{j=1}^{P} P(\hat{y}_{i,j} = y_{i,j}) \right)^{\frac{1}{N*P}} \tag{15}$$

where $\hat{y}_{i,j}$ is the predicted vote for a legislator and $y_{i,j}$ is the observed vote.

Figure 3 shows the GMP metric for each session of the U.S. House broken out by the dimensionality

FIGURE 3. Geometric Mean Probability of Correct Classification for Unidimensional and Multidimensional Votes in the $1^{st} - 115^{th}$ Sessions of the U.S. House



Note: Values reported are posterior means.

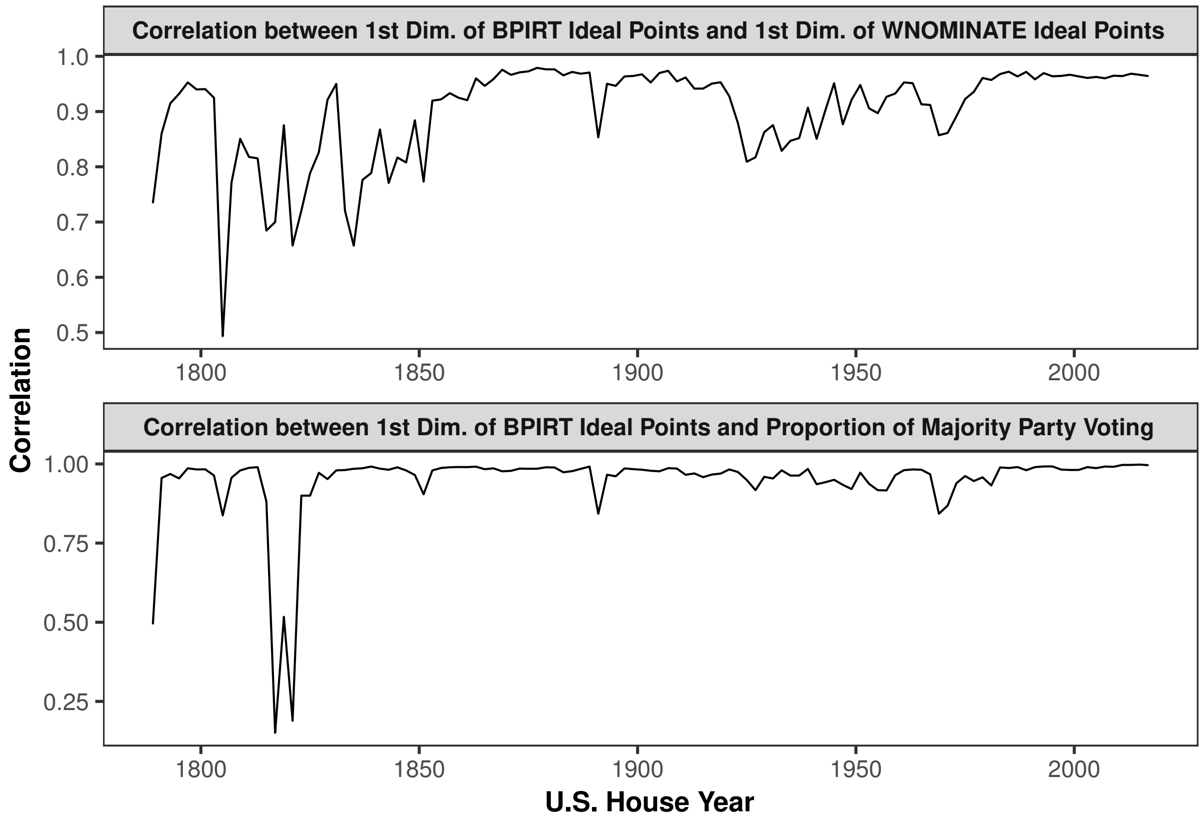
of the vote implied by BPIRT.[8] Unsurprisingly, there are significant gains made in model fit when examining multidimensional votes. The difference in model fit between WNOMINATE-1D and BPIRT on these votes is quite large. Combined with the knowledge that many votes within each session are multidimensional, this provides strong evidence that unidimensional models are missing out on a large portion of the variation which drives voting in the U.S. House. BPIRT also shows large gains over WNOMINATE-1D when analyzing unidimensional votes. This result is somewhat unintuitive as the underlying formal model for a unidimensional vote is essentially the same for BPIRT and WNOMINATE. However, this result can be attributed to proper placement of zeros in the binary matrix and, in turn, ensuring that each vote corresponds only to the correct subset of potential dimensions of the policy space.

Interpretation of Ideal Points

BPIRT paints a picture of a legislature that behaves in a multidimensional manner; while not all votes require multiple dimensions to explain voting patterns, many votes need something beyond a single dimensional ideal point to best explain voting. A natural question that follows pertains to the meaning

[8]Gaps in the multidimensional plot occur when BPIRT estimates that a session has only one dimension or less than 10 votes required more than one dimension to best model its vote outcomes.

FIGURE 4. Correlation between the 1st Dimension of Ideal Points Estimated by BPIRT, the Ideal Points from WNOMINATE-1D, and the Proportion of Majority Party Voting



Note: Values reported are posterior means.

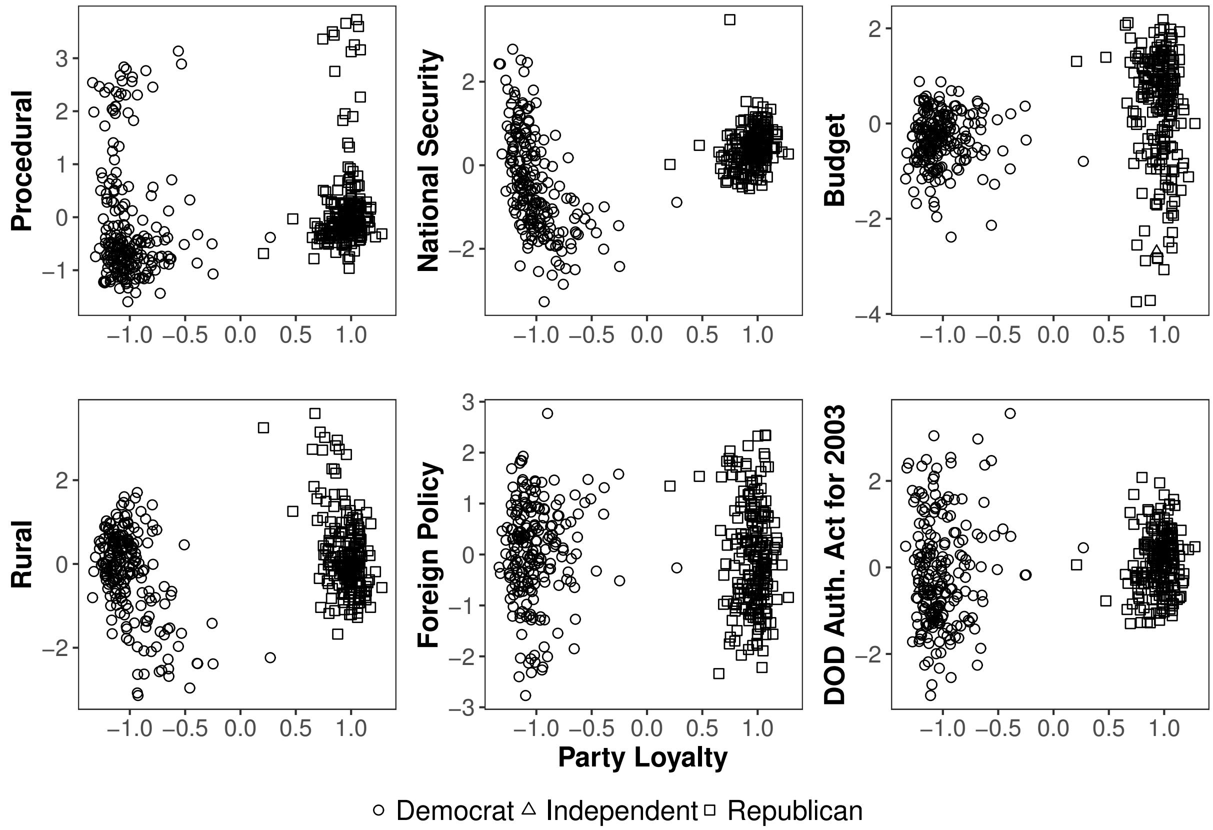
of the dimensions - what is represented by the first dimension and what concepts are represented by dimensions beyond the first?

First, I examine the meaning of the first dimension, over time. In particular, I examine whether or not the first dimension is simply providing a measure of the individual frequency of party bloc voting (Aldrich et al. 2014; Lee 2009; Harbridge 2015). In order to test this hypothesis, I measure how frequently a voting member of the legislature votes with that session's majority party.[9] Given that more than 99% of votes have a non-zero contribution from the first dimension, over time, understanding the meaning of this dimension is key.

Figure 4 shows the correlation between the ideal points from the first dimension for each voting member compared to the proportion of votes for which they cast the same vote as the majority party preference. The relationship between majority party voting and the uncovered ideal points strongly supports the theory that the first dimension of BPIRT ideal points is simply modeling party teamsmanship. Figure 4 also shows that BPIRT and WNOMINATE are highly correlated over time. This, in turn, implies that the first dimension of WNOMINATE is largely estimating the same construct with the first dimension.

[9]Specifically, I determine the majority party vote on a give roll call vote to be the most frequently made choice by the members of the majority party in a given session.

FIGURE 5. Ideal Points and Dimensions Estimated by BPIRT for the 107th Session of the U.S. House (2001 - 2003)



Note: The reported ideal points are from the iteration of the MCMC procedure with the highest complete-data likelihood.

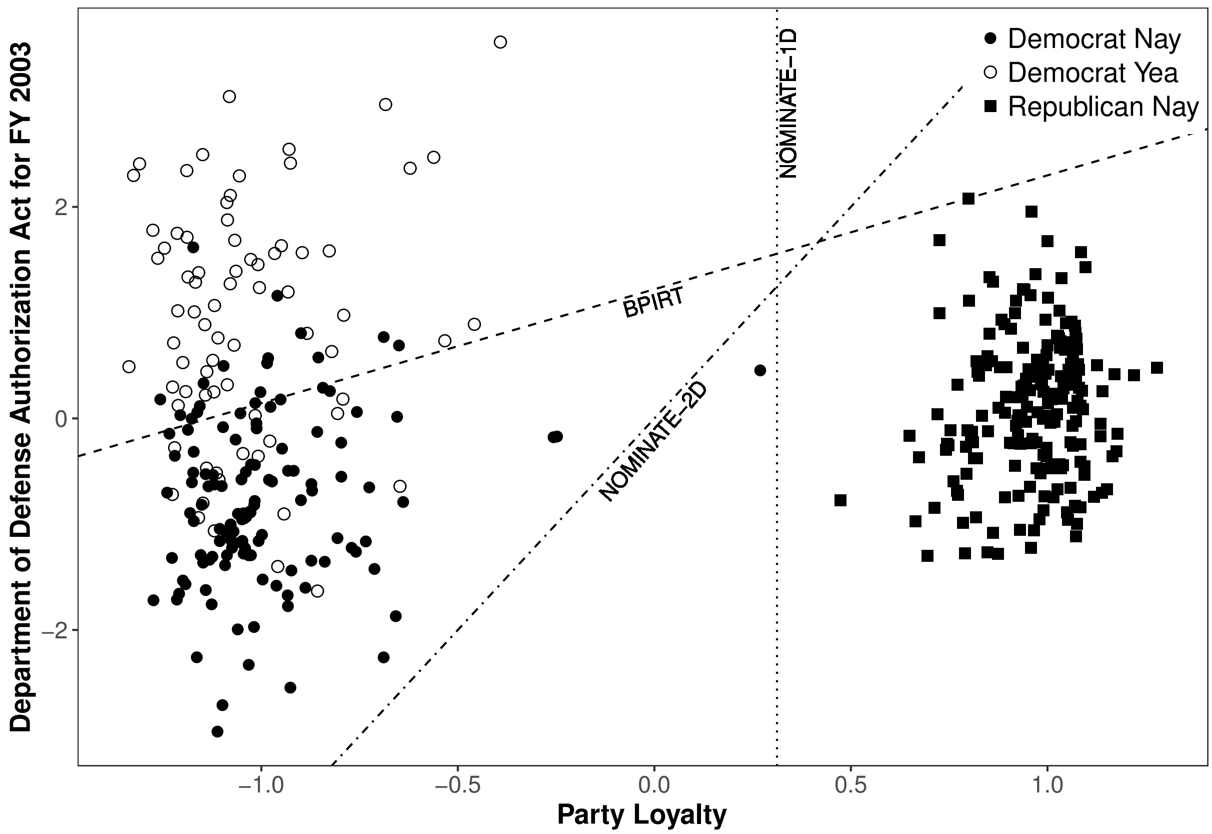
In order to demonstrate the importance of dimensions beyond the first in explaining legislative behavior that exists outside of party loyalty, I examine one particular session.[10] The 107^{th} session of the U.S. House took place between 2001 and 2003 and contained the September 11^{th} attacks and the ensuing scramble from the U.S. government attempting to respond to domestic and foreign security concerns. These issues created strong divisions within parties and led to a number of outcomes that appeared to favor the pro-war members of the U.S. Congress. It is reasonable to expect that a significant portion of roll call votes in this session require dimensions beyond party loyalty when explaining variation and estimating ideal points.

Figure 5 shows the seven dimensions of ideal points uncovered by BPIRT for the 107^{th} session of the U.S. House.[11] First, and foremost, the party loyalty dimension is highly apparent and shows a split between Republican and Democratic voting (PVE = .77). Other dimensions are important and explain the other 23% of explainable variance. Some of the dimensions relate closely to specific policy topic areas such as rural/infrastructure issues (PVE = .03) and the government budget (PVE = .05). Another

[10] Roberts et al. (2016) and Bateman et al. (2017) show that dimensions beyond the first are colored by the salient issues of their time and the preferences of the agenda setters. This make a general analysis of dimensions beyond the first difficult.

[11] Additional material about how names for each dimension were determined and the structure of dimensions beyond the first are included in Section B.3 of the appendix.

FIGURE 6. Votes and Cutlines for Department of Defense Authorization Act for Fiscal Year 2003 Vote in the 107^{th} Session of the U.S. House (Roll Call No. 655)



Note: The reported ideal points and cutlines are from the iteration of the MCMC procedure with the highest complete-data likelihood.

dimension that emerges relates to purely procedural votes, such as ceremonial motions and approving the chamber's journal (PVE = .02). However, the set of dimensions that are most important to this session, beyond party loyalty, relate to the September 11^{th} attacks and the relating security measures. These dimensions include national security (PVE = .07), foreign policy (PVE = .04), and a dimension that relates to funding the war in Afghanistan (PVE = .02). Given the sets of issues that were salient in the 107^{th} session of the U.S. House, this set of dimensions beyond party voting makes sense.

Though all votes do require party loyalty to explain some of the variance, most votes require additional explanations from other sources. One example that is particularly relevant to the 107^{th} U.S. House relates to funding the war in Afghanistan. Votes related to funding military action arose after the September 11^{th} attacks. While the Republican party unanimously agreed to motions to increase funding to the Department of Defense for these actions, Democrats were split in these votes. Though the Republicans held the House majority, many Democrats used these votes to signify support for or against the war to their constituents and this created splits in the voting.

The BPIRT estimation procedure selects 14 of these votes and estimates that these votes require a common dimension in addition to the party loyalty dimension. While this dimension only accounts for around 2% of the total variance explained in this session, it models an important heterogeneity in Democrat voting. Figure 6 shows the vote outcome by party for one of these votes, which pertained to an amendment to the Department of Defense Authorization Act for 2003 proposed by Loretta Sanchez

(D-CA). This figure shows the BPIRT ideal points of the voters in two dimensions: the party loyalty dimension and the DOD dimension. Additionally, I illustrate three separate cutlines which show the line on which a voter would be undecided between a "Yea" or "Nay" vote. When this vote is scaled using WNOMINATE with only one dimension, the cutline indicates perfect within party agreement. This is not the case and is indicative of a second dimension at play. However, WNOMINATE in two-dimensions misses the important cut needed for this vote. On the other hand, BPIRT creates a cutline that splits the Democrats into those that support higher funding and those that oppose spending increases. This example perhaps best demonstrates the differences between BPIRT and other roll call scaling procedures; BPIRT estimates dimensions as a function of clusters of votes that share similar voting patterns and finds dimensions that are necessary for modeling their outcomes. This consistency in topic is a feature unique to BPIRT and provides a tool that can create in-depth inference of the topics that drive legislative voting throughout U.S. history.

U.S. LEGISLATIVE VOTING AND MULTIDIMENSIONALITY

BPIRT provides a tool for analyzing roll call votes and understanding the dimensionality of votes as well as the issue sets that drive variation in voting within the U.S. legislative chambers. While BPIRT shows marked improvements over previous approaches to roll call scaling, its benefit can be seen as bridging the aggregate roll call scaling approaches of Poole and Rosenthal (1997) and the issue-specific approaches of Roberts et al. (2016) and Bateman et al. (2017). This gives rise to measures of multidimensionality that are comparable within and across sessions and provides a unique measure that can assess the impacts of multidimensionality on theories of legislative behavior.

I leverage the ideal points estimated by BPIRT and the corresponding measures of dimensionality to explore the relationship between the dimensionality of a vote and the outcomes that are predicted by models of U.S. legislative voting. While there are numerous examples of models that appeal to unidimensionality and test theories utilizing unidimensional NOMINATE scores, I turn my attention to two specific models: the pivotal voter model presented by Krehbiel (1998) and the party cartel model presented by Cox and McCubbins (2005). These two models are widely cited in studies of U.S. legislative behavior and seek to explain the ways in which the organization of legislative voting and parties influence voting outcomes. These two models differ in their explanations of how voting decisions are made, but strongly leverage a unidimensional policy space in the theoretical and empirical examinations of their theories.

Specifically, I seek to understand how robust these theories are to the assumption of unidimensionality. Theoretical outcomes under multidimensionality are well established, but there are few empirical studies of the impact of multidimensionality on voting in the literature. Unidimensionality can be best described as a stabilizing assumption - when the underlying policy space is unidimensional, the outcome is predictable given assumptions about how legislators behave. In contrast, multidimensional votes are theoretically characterized by outcomes that can take any form. Thus, the goal is to measure the stability of outcomes implied by the pivotal voter and party cartel models when properly taking dimensionality into account.

As discussed previously, multidimensionality comes in many different shapes and sizes. For example, a vote can be multidimensional, but rely very heavily on one single dimension while there is only a small amount of variance explained by another set of dimensions. For this reason, I explore three separate ways in which multidimensionality may relate to vote instability:

1. **No Effect**: As the multidimensionality of a vote increases, there is no discernible change in the stability of voting outcomes.
2. **Continuous Effect**: Vote outcomes become more and more unstable as the multidimensionality of the vote increases; low levels of multidimensionality show more stable outcomes than votes with higher levels of multidimensionality.
3. **Threshold Effect**: Vote outcomes are stable and predictable up to a small amount of multidimensionality. Once this threshold is crossed, vote outcomes fundamentally change (McKelvey 1976; Schofield 1978). Even when the amount of multidimensionality is small, there is a marked difference between unidimensional and multidimensional outcomes.

Each of these mechanisms provide a different view of how theories of U.S. legislative voting might be influenced by the assumption of unidimensionality.

Evidence that multidimensionality influence vote outcomes has significant implications for the study of U.S. legislative voting. First, existing theories related to legislative voting must be examined for conditional relationships - if the vote is multidimensional, does the prediction from the theory change? Multidimensionality points to different factors that are necessary for contextualizing the conditions under which a vote are made. Second, the usage of unidimensional ideal points under evidence for multidimensionality leads to potential biases in further results. Given the interpretation of the first dimension examined previously, usage of unidimensional ideal points when multiple dimensions are needed essentially summarizes the level of majority party voting of a member while treating other sources of predictable roll call behavior as noise. Particularly when being used as proxies of preferences to test theories of party control, this can lead to acceptance of theories as a product of an endogenous measure. Finally, under the common assumption of rational voters, evidence that multidimensionality leads to more unpredictable outcomes points to ways in which rational proposers can skew proposals to their advantage (Riker 1980; Shepsle 1978; Shepsle and Weingast 1981; Baron and Ferejohn 1989). If multidimensional models are appropriate models of legislative voting, then the idea that strategic proposers can utilize multidimensionality to achieve better outcomes must be accounted for within theories related to the legislative process.

A Theory of Pivotal Voters

Perhaps one of the most well known theories of U.S. legislative behavior, Krehbiel (1998) outlines a theory of pivotal voters in legislative voting. Under this model, a proposal, the status quo, and voters are mapped to a unidimensional, commonly-known policy space. Under the rules of the legislative body, Krehbiel (1998) contends that policies must be proposed outside of the gridlock zone in order to pass the chamber. The gridlock zone is defined by the median voter, the presidential veto pivot, and (when appropriate) a Senate filibuster pivot. These members of the legislature effectively control the proposals which pass and, in turn, rational proposers craft legislation with these constraints in mind. This model of legislative voting is simple and effective, leading to many insights about periods of low and high gridlock within the U.S. legislature.

To examine how multidimensionality influences empirical support for the theory of pivotal politics, I recreate the empirical analysis from Krehbiel (1998, Chapter 5) that examines the behavior of filibuster pivots in successive cloture votes in the U.S. Senate. When a vote to invoke cloture goes before the U.S. Senate, Krehbiel (1998) contends that the filibuster pivot controls whether or not debate on a vote is stopped. An interesting test of this theory arises when a cloture vote occurs multiple times in the same bill-episode. The theory of pivotal voters claims that changes in individual votes from vote to vote are most likely to occur for members close to the filibuster pivot location. Using unidimensional NOMINATE scores to find voters close to the theoretical filibuster pivot, Krehbiel (1998) finds evidence for the pivotal voter model.

I contend that the empirical evidence shown by Krehbiel (1998) is driven by the unidimensional assumption made when utilizing unidimensional NOMINATE scores; I expect that evidence for the pivotal voter model disappears in pairs of votes which are multidimensional. This conditional view of the pivotal politics model fits well within the original construction - if voters are able to collapse the preference space to a single dimension, then rational proposers can target changes in bills. However, under the multidimensional model, no such targeting can be made.

To examine these competing theories, I recreate this analysis on new data. Using the set of all cloture votes that took place in the the $89^{th} - 115^{th}$ sessions of the U.S. Senate, I examine all instances of votes to invoke cloture that occurred at least twice within the same bill-episode. I then grouped these into sequential vote pairs, mimicking the data set of Krehbiel (1998). This led to 471 cloture vote pairs over the course of time analyzed. For each vote pair, I then recorded whether each U.S. Senator switched their vote. This led to $44,710$ vote observations and $3,363$ vote switches. For each individual U.S. Senator, the ideal point associated with the first dimension of BPIRT scores was coded into quartiles: the filibuster quartile (FQ), the filibuster-adjacent moderates quartile (AM), the filibuster-adjacent extremists quartile (AE), and the non-adjacent extremists quartile (NE). Similarly, the controls outlined by Krehbiel (1998, Chapter 5) were recorded: President-side vetoes, voting under a unified government, and whether or not the voter was a Democrat.[12]

To measure the dimensionality of a pair of cloture votes, I used the PVE for dimensions beyond the party-loyalty dimension in both votes. If both votes were estimated by BPIRT to be unidimensional, then dimensionality was coded as zero. Otherwise, dimensionality was set to be equal to the posterior mean value of PVE. The effect of multidimensionality is measured by testing three models with different underlying mechanisms. The theory of no effect was tested by not including a control for the dimensionality of the vote. Under the theory of continuous effect, multidimensionality was coded as the PVE attributed to dimensions beyond party loyalty. Finally, the threshold effect was tested by including a dichotomous variable for dimensionality coded to be multidimensional if the PVE for other dimensions in the vote pair was greater than .001. 43% of vote pairs were classified as multidimensional under the threshold model.

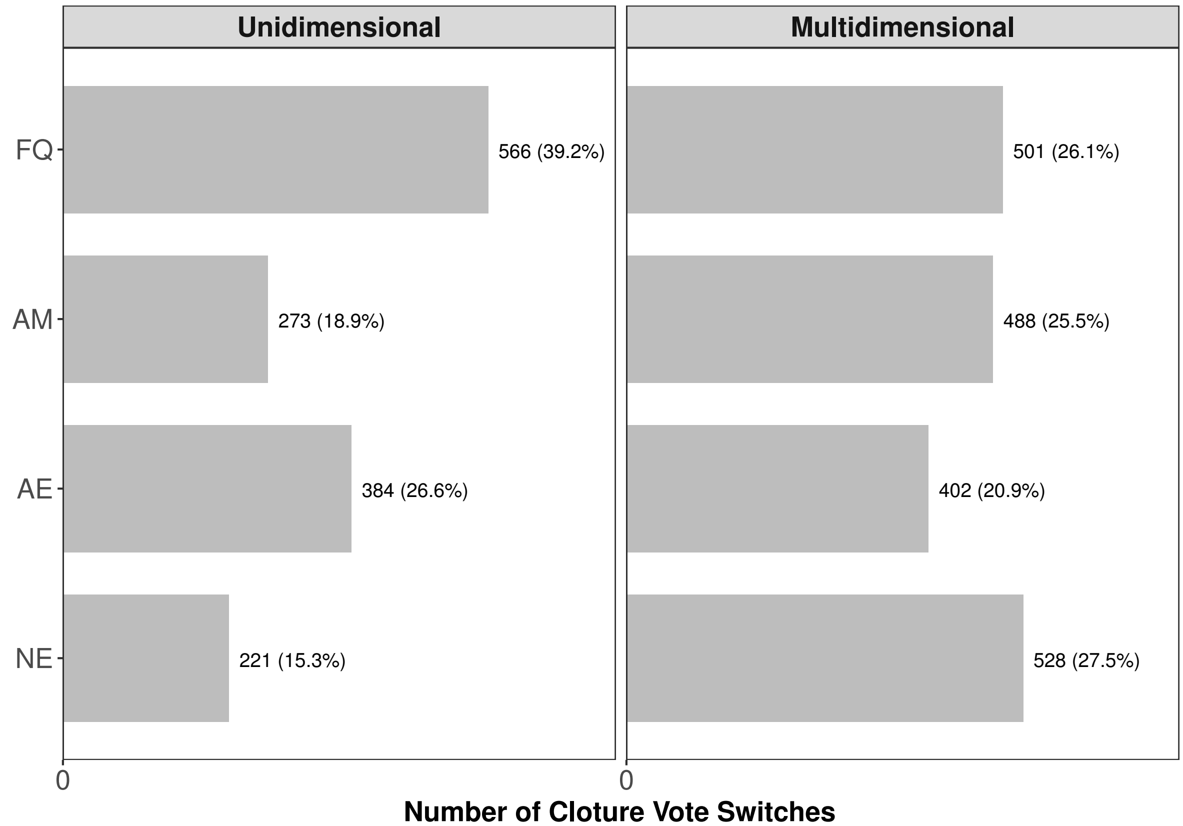
Figure 7 shows the frequency of vote switches for members of each quartile in both unidimensional and multidimensional votes as defined by the threshold model. This plot shows overwhelming support for the theory that dimensionality matters when assessing the appropriateness of the pivotal voter model in explaining vote switches. The switches that are associated with unidimensional votes are much more likely to arise from members of the f-quartile than in other quartiles. On the other hand, vote switches are generally much more likely in the multidimensional case with a more uniform distribution of switches over the quartiles.

Following the empirical exercise by Krehbiel (1998), I estimated the probability of a vote switch given the ideal point quartiles, dimensionality of the vote, and the controls.[13] Table 1 shows the

[12]Corresponding switches to the quartile coding were made when the vote met the conditions of a "president-side veto". For a further explanation of this control and the corresponding recoding, see Krehbiel (1998, Chapter 5, p. 106).

[13]While a case could be made that session-level random effects are needed, I appeal to the response of Krehbiel (1998) that such controls would be "utterly atheoretic" as there are no session specific elements of the pivotal voter theory.

FIGURE 7. Number of Vote Switches across Cloture Vote Pairs by Dimensionality and Ideal Point Quartile



results of this regression under the three different theories of how dimensionality might influence vote outcomes. Similarly, Figure 8 shows the probabilities of vote switching as a function of ideal point quartile and the dimensionality of the vote as defined by the threshold model. The results from the regressions show a number of interesting relationships. First, under the model of no effect, the results from Krehbiel (1998, Chapter 5) are largely replicated in the new data set - assuming that dimensionality has no effect on vote switching, voters in the f-quartile are most likely to switch votes between cloture vote pairs. This relationship is also seen in the other two models when the vote pair is unidimensional, providing strong support that the pivotal voter findings are supported in the strict unidimensional case. However, this relationship disappears in the multidimensional case. Results from the regression echo the vote switching distribution for multidimensional votes shown in Figure 7. When the vote pair is multidimensional, there is no statistical difference between any quartile in the 95% HPD intervals.

Comparing the marginal likelihood across the models provides an assessment of the degree to which controlling for multidimensionality improves model fit. First and foremost, it is clear that the marginal likelihood for both the continuous and threshold models are lower than the model with no effect. Given that the marginal likelihood inherently penalizes against overfitting, this is strong evidence that multidimensionality explains a significant amount of variation within the data. Between these models, the threshold effect has the lowest marginal likelihood indicating that even a small amount of multidimensionality leads to a fundamentally different role of the pivotal voter. This result along with the uniform distribution of vote switches points to a chaotic result for more than 50% of the votes

TABLE 1. Logistic Regression Results for Krehbiel's Cloture Vote Switching Example (Krehbiel 1998, Chapter 5)

	Dependent variable:		
	Senator Vote Switch in Cloture Vote Pair		
	No Effect	Continuous	Threshold
f-Quartile	0.40	0.63	0.96
	(0.31,0.50)	(0.52,0.74)	(0.80,1.13)
f-Adj. Moderates	0.03	0.17	0.22
	(-0.06,0.14)	(0.04,0.28)	(0.03,0.40)
f-Adj. Extremists	0.06	0.25	0.52
	(-0.04,0.16)	(0.13,0.36)	(0.36,0.69)
President Side	-1.04	-1.03	-1.01
	(-1.14,-0.95)	(-1.13,-0.92)	(-1.12,-0.92)
Unified Government	-0.46	-0.45	-0.42
	(-0.55,-0.38)	(-0.54,-0.36)	(-0.51,-0.33)
Democrat	-0.40	-0.39	-0.36
	(-0.47,-0.33)	(-0.46,-0.31)	(-0.43,-0.28)
Multidimensional		1.34	1.10
		(1.07,1.63)	(0.94,1.26)
FQ $\times$ Multidim.		-2.10	-0.94
		(-2.64,-1.72)	(-1.15,-0.73)
AM $\times$ Multidim.		-1.03	-0.28
		(-1.47,-0.60)	(-0.50,-0.05)
AE $\times$ Multidim.		-1.61	-0.74
		(-2.08,-1.14)	(-0.97,-0.52)
Intercept	-2.11	-2.29	-2.78
	(-2.19,-2.02)	(-2.39,-2.19)	(-2.92,-2.63)
Observations	44,710	44,710	44,710
Log Marginal Likelihood	-11,609.54	-11,577.67	-11,464.47

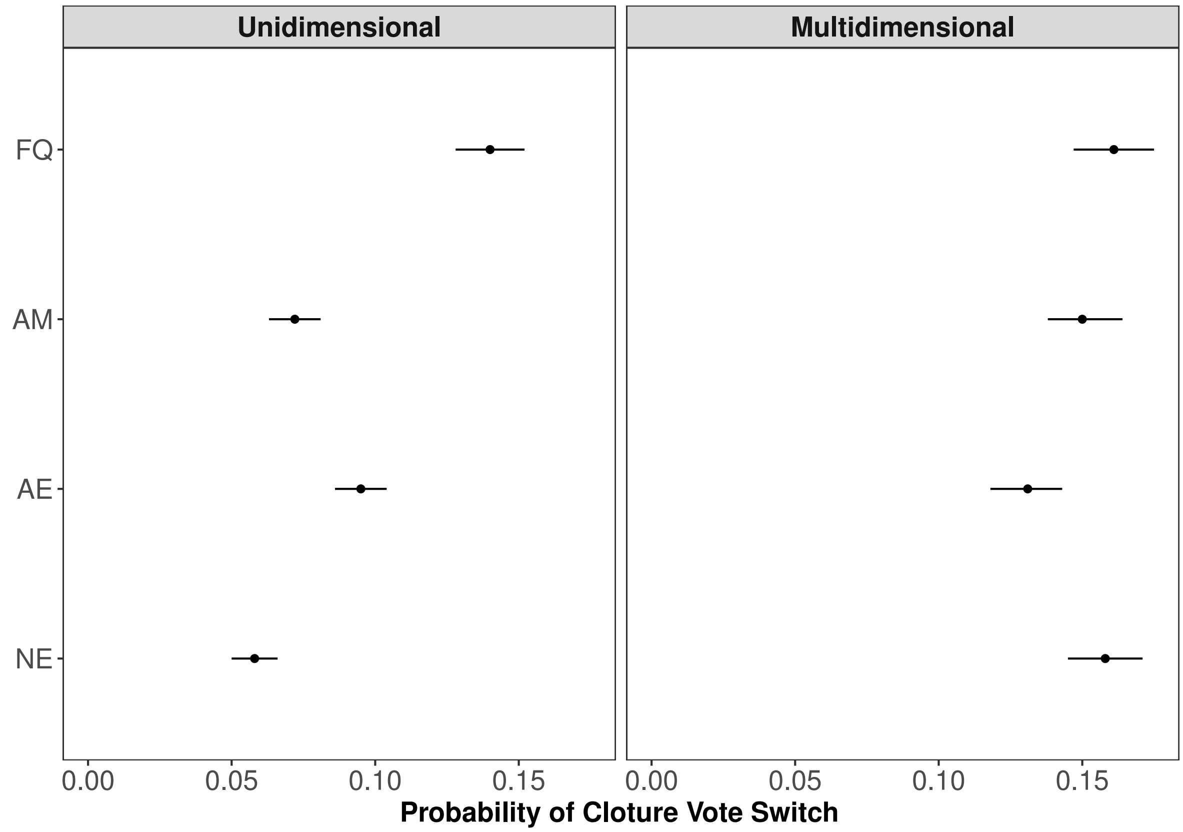
[1] The comparison group is Nonadjacent Extremists.
[2] Coefficient values are posterior medians and values in parentheses are 95% HPD intervals.
[3] Dimensionality under the threshold model is coded as 1 if the proportion of explained variance attributed to other dimensions is greater than .001 and 0 otherwise.

analyzed; the pivotal voter model provides no information about where vote switches will occur when the vote is multidimensional. This is one example of how properly accounting for multidimensionality in U.S. legislative voting can fundamentally change long held beliefs about legislative decision making.

A Theory on Party Control

In response to the work of Krehbiel (1998), scholars pointed to parties as another source that influences both the proposals that are made and the decisions made by their respective members. One example of

FIGURE 8. Probability of Cloture Vote Switch by Ideal Point Quartile



Note: Probabilities were calculated setting President Side, Unified Government, and Democrat to zero. Error bars show 95% HPD intervals.

this work is the party cartel theory of agenda control (Cox and McCubbins 2005). Under this model, rational voters in a unidimensional preference space want to select proposals that are close to their ideal points. However, their desire to be reelected also influences their vote choices and they frequently delegate to the central authority of the party to make vote decisions. This leads to strong party control in vote choice which can lead to votes that are against their individual best actions conditional on their ideal points.

Party cartel theory leads to a number of empirically testable predictions about agenda control in the U.S. Congress. One specific example relates to final passage votes in the U.S. House (Cox and McCubbins 2005, Chapter 5). For each final passage vote, the result can be classified by the proportion of members from the majority and minority parties voting in support of passage: if less than 50% of the minority party votes in support of passage, the vote is considered a minority roll with similar conditions defining a majority party roll. Under the theory of party cartels, these rolls occur in predictable ways. First, majority rolls should be rare and uniformly distributed conditional on the distance between the ideal point of the chamber median and the ideal point of the median member of the majority party. On the other hand, minority rolls should occur often with the frequency increasing as the distance between the floor median and the minority party median increases. Using unidimensional NOMINATE scores as estimates for the ideal points, Cox and McCubbins (2005, Chapter 5) find empirical evidence for both predictions.

As with Krehbiel (1998), I expect that this empirical evidence is colored by the usage of unidimensional ideal point estimates and the result is again conditional on the dimensionality of a vote. It is reasonable to believe that close alignment between the majority party median and the floor median indicates high levels of party loyalty and, in turn, produces strong agreements within the majority party on votes that can be thoroughly explained by party bloc voting. On the other hand, multidimensionality in a vote points to sources other than party voting and it is reasonable to expect that these votes have less predictable outcomes.[14]

To assess the role of multidimensionality, I examined the set of all final passage votes for the $84^{th} - 114^{th}$ sessions of the U.S. House.[15] This data was retrieved from the Political Institutions and Public Choice Roll-Call Database (Crespin and Rohde 2012). I determined if each final passage vote was a majority party roll, minority party roll, or if no roll had occurred. This led to $4,429$ observations of final passage votes with 127 majority party rolls and $1,743$ minority party rolls.[16] I used ideal point estimates from the first dimension of BPIRT for each session and recorded the location of the floor median and the respective party medians for each vote. The absolute difference between these two metrics constitutes the distance between medians.

As before, the three theories of multidimensionality are tested. Under the theory of no effect, the probability of a roll is tested only as a function of distance between the respective medians. The continuous effect was tested by utilizing the PVE for dimensions beyond the first on a given vote, the distance between medians, and a multiplicative interaction between the two. Finally, similar to the continuous model, the threshold effect was tested by classifying any vote where the PVE for dimensions beyond the first was greater than .001 as multidimensional. Under the threshold model, 68.8% of final passage votes were classified as multidimensional.

Figure 9 shows the set of final passage votes analyzed. Votes are compared based on the proportion of "Yea" votes cast by each party in each case. Votes are then classified as unidimensional or multidimensional using the threshold model. The difference in vote proportions between the two classes of votes is stark. On unidimensional votes, there are two general outcomes: near unanimous support by the entire set of voters or minority rolls with nearly unanimous support from the majority party. Majority party rolls on unidimensional votes are incredibly rare - only 4 out of 1,382 (.002%) unidimensional final passage votes result in majority party rolls. On the other hand, the results for multidimensional votes are more varied; the rate of minority rolls appears to be equivalent to the rate of votes where the minority party is not rolled and the rate of majority rolls (4%) is much higher than in the unidimensional case.

I used a logistic regression to examine the relationship between party median distance, multidimensionality, and party rolls. In order to best replicate the test from Cox and McCubbins (2005), I chose to include zero-mean session-level random intercepts estimated via a normal random effect. I report the variance of these parameters with the results from these regressions. Diffuse normal spike-and-slab priors with a spike at zero were placed on the regression coefficients.

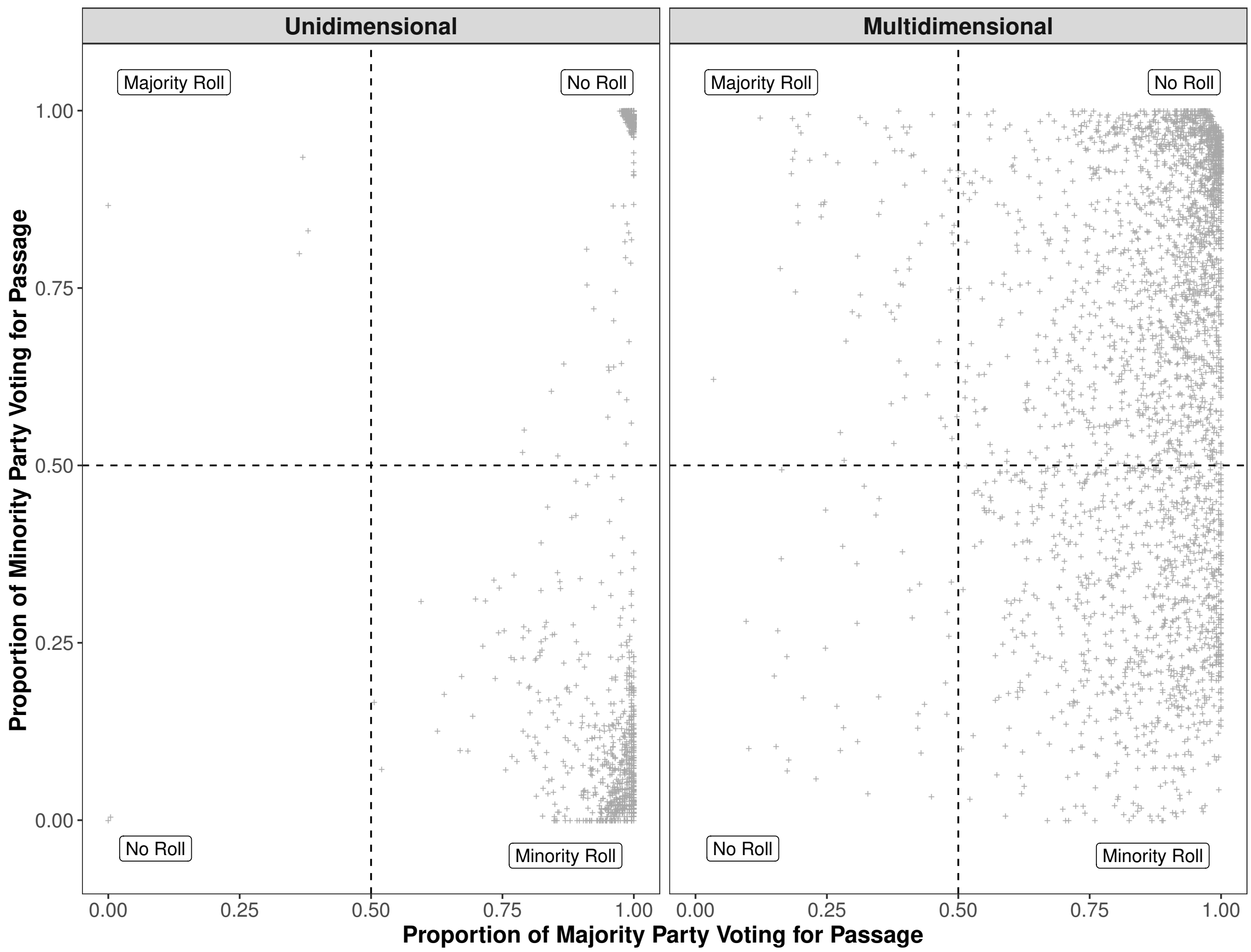
Table 2 shows the results from these regressions. First, examining the model of no effect, the results from Cox and McCubbins (2005) are replicated. When all votes are assumed to be unidimensional, the probability that a final passage vote results in a majority party roll shows no evidence that it is influenced by the distance between medians. Similarly, there is a large positive correlation between the probability of a minority party roll and the distance between the minority party median and the

[14]It is worth noting that multidimensionality is mentioned by Cox and McCubbins (2005) and this possibility is explicitly considered, but not examined in depth.

[15]In line with Cox and McCubbins (2005), I restrict this set to votes which required a simple majority for passage.

[16]The rate of passage for final passage votes is around 98%.

FIGURE 9. Results of Final Passage Votes in the U.S. House



floor median. However, there is strong evidence that controlling for the dimensionality of a vote explains more variation in both sets of party rolls. First, the DIC, a proxy for the marginal likelihood in hierarchical models that penalizes the addition of new parameters, is lower for the models that control for multidimensionality in both majority and minotiry party rolls. While this decrease is moderate in the majority rolls case, the minority rolls case shows a massive decrease in DIC. In each regression, the threshold model shows the smallest DIC, implying that a model that treats votes with even a small amount of variation that can be explained by dimensions beyond the first differently than unidimensional votes provides the best fit to the data.

Figure 10 shows the predicted probabilities of majority and minority party rolls as a function of distance between the respective medians and the dimensionality of a vote under the threshold model. These results point to a characterization of negative agenda control that is conditional on the dimensionality of the vote. First, the probability of a majority party roll is low in both unidimensional and multidimensional votes. While certainly lower in the unidimensional case, multidimensional votes show a predicted probability of around .06 in the case of the maximum observed distance between the floor and majority party medians. However, there is a statistically meaningful increase in the probability of majority rolls in the multidimensional case - while negative agenda control from the majority party is apparent, it seems that there are potentially other factors at play (Aldrich and Rohde 2000).

On the other side, minority party voters benefit from multidimensionality. Under strict unidimensional

TABLE 2. Logistic Regression Results for Cox and McCubbins' Final Passage Vote Example (Cox and McCubbins 2005, Chapter 5)

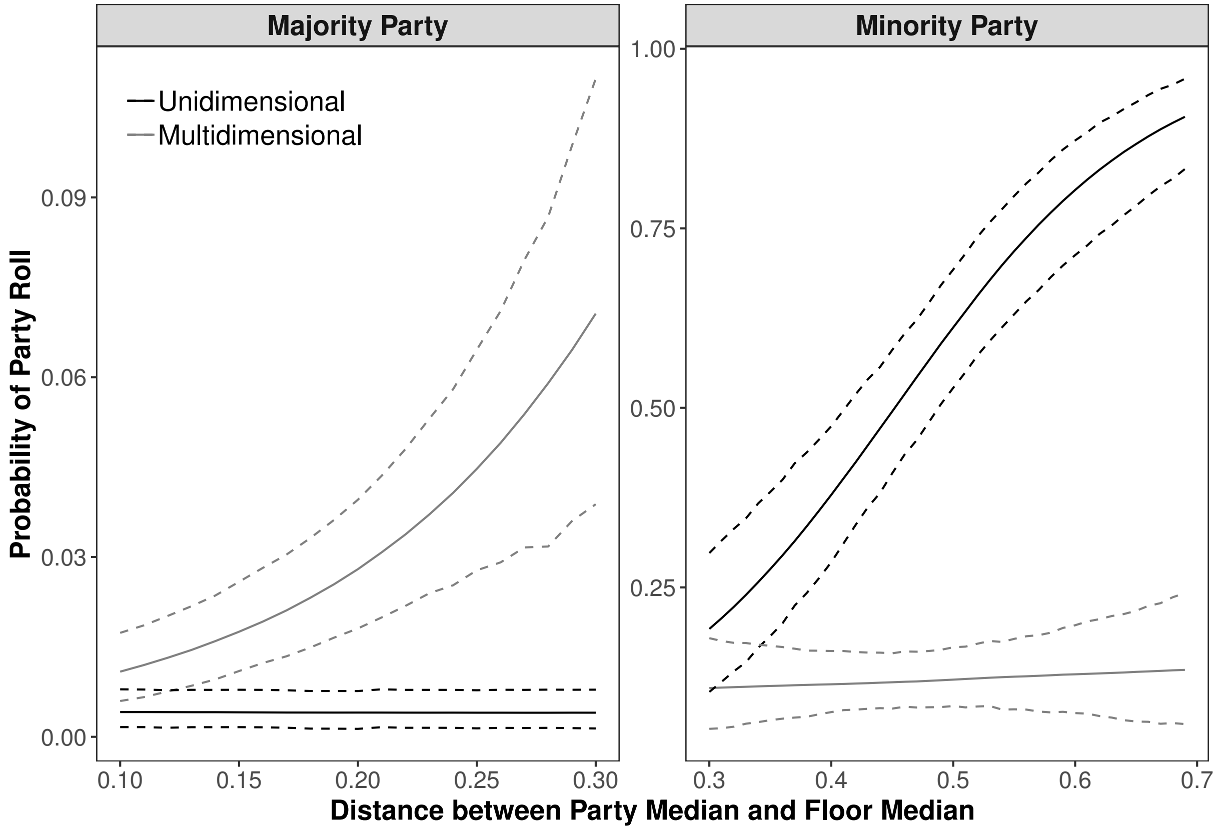
	Dependent variable:		
	Majority Party Roll on Final Passage Vote		
	No Effect	Continuous	Threshold
Distance	0	0	0
	(0,2.31)	(-2.63,0)	(-3.5,0)
Multidimensional		0	0
		(0,0)	(0,0)
Dist. × Multidim.		10.08	9.75
		(6.68,13.80)	(6.51,13.27)
Intercept	-3.79	-5.55	-5.48
	(-4.32,-3.38)	(-6.35,-4.72)	(-6.28,-4.66)
Variance of	0.72	0.68	0.66
Random Effects	(0.43,1.05)	(0.4,.99)	(0.38,0.97)
Observations	4,429	4,429	4,429
DIC	1131	1096	1077

	Dependent variable:		
	Minority Party Roll on Final Passage Vote		
	No Effect	Continuous	Threshold
Distance	7.78	10.04	9.44
	(4.82,10.45)	(7.17,12.91)	(6.52,12.48)
Multidimensional		3.60	2.00
		(2.16,4.88)	(1.22,2.83)
Dist. × Multidim.		-15.22	-8.88
		(-18.26,-12.46)	(-10.71,-7.23)
Intercept	-4.15	-4.38	-4.26
	(-5.48,-2.74)	(-5.75,-2.95)	(-5.72,-2.82)
Variance of	0.29	0.24	0.25
Random Effects	(0.2,0.41)	(0.15,0.35)	(0.15,0.37)
Observations	4,429	4,429	4,429
DIC	4994	4264	4158

[1] Coefficient values are posterior medians and values in parentheses are 95% HPD intervals.
[2] These models were estimated with spike-and-slab priors on the coefficients. The spike was placed at zero. Posterior values equal to zero arise when the coefficient is not statistically distinguishable from zero.
[3] Congress-level random effects were estimated for the intercept terms only.
[4] Dimensionality under the threshold model is a binary variable coded as 1 if the PVE of the first dimension is less than .001 and 0 otherwise.

votes, the minority party roll rate is positively correlated with the distance between medians. However, in multidimensional votes, knowing the distance between the minority party median and the floor

FIGURE 10. Probability of Party Rolls as a Function of Distance between the Party Median and the Floor Median



Note: Probabilities were calculated only within the range of observed outcomes for majority and minority party distances. Dotted lines show 95% HPD intervals.

median provides no information about the probability of a minority party roll. In other words, the predictions from party cartel theory relating to the minority party only apply to around 30% of final passage votes. This finding does not invalidate the party cartel theory. Rather, it points to multidimensionality creating cross-party support for bills that are not necessarily correlated with the number of times that a member votes with the majority party. Along with theory that strategic proposers use multidimensionality to create passing votes that cater to their own preferences, this opens a new door for research into the role of agenda control in the U.S. Congress under potentially highly multidimensional bills.

CONCLUSION

Roll call scaling and the operationalization of the spatial model is critical to the scientific examination and development of theories about how members of the U.S. Congress cast votes. While existing methods produce scores that appear to be best represented in a single dimension, I show that this finding is due to the tests used to determine dimensionality. In turn, I develop a varying dimensional representation of the spatial model and show a corresponding estimation technique that allows for estimation of both the aggregate-level dimensionality of the ideal point space as well as vote-specific sets of dimensions. Using this model, I show that there is little evidence for unidimensional ideal points in the U.S. House and U.S. Senate and that historical voting demonstrates multidimensional patterns. I present a set of ideal points that bridge the gap between the common aggregate methods and the more subject-specific examinations that are present in the roll call scaling literature. Under this new model, I then show that multidimensionality is an important aspect to be considered for further models of U.S. legislative voting that rely on ideal points as summaries of the preferences of voters. All in all, I show that BPIRT is a powerful procedure for determining the dimensionality of latent variables used in social science applications.

While the work in this article is catered to the study of roll call scaling, the models presented here are widely applicable to any study where latent variables are estimated via an item-response theory model. With light changes, Indian Buffet Process priors can be used to test dimensionality in a variety of settings and can be used as validation for claims related to the dimensionality of a latent space. Similarly, there are numerous extensions which can be made to the model presented in this paper. Under the Bayesian nonparametric framework, it is possible to examine how dimensions change over time by providing sufficient changes to the underlying priors. Similarly, different methods of clustering can be used to find interesting similarities in voting between both voters and the topics of the votes, themselves. This model serves a starting point into more complex examinations of what all can be learned from the spatial model of voting and scaling techniques.

REFERENCES

Aldrich, John H , Jacob M Montgomery, and David B Sparks (2014). Polarization and ideology: Partisan sources of low dimensionality in scaled roll call analyses. *Political Analysis 22*(4), 435–456.

Aldrich, John H and David W Rohde (2000). *The logic of conditional party government: Revisiting the electoral connection*. PIPC.

Baron, David P and John A Ferejohn (1989). Bargaining in legislatures. *American political science review 83*(4), 1181–1206.

Bateman, David A , Joshua D Clinton, and John S Lapinski (2017). A house divided? roll calls, polarization, and policy differences in the us house, 1877–2011. *American Journal of Political Science 61*(3), 698–714.

Bhattacharya, Anirban , David B Dunson, and Others (2011). Sparse Bayesian infinite factor models. *Biometrika 98*(2), 291.

Binder, Sarah A (1999). The dynamics of legislative gridlock, 1947 96. *American Political Science Review 93*(3), 519–533.

Bonica, Adam (2014). Mapping the ideological marketplace. *American Journal of Political Science 58*(2), 367–386.

Carroll, Royce , Jeffrey B Lewis, James Lo, Keith T Poole, and Howard Rosenthal (2009). Measuring bias and uncertainty in DW-NOMINATE ideal point estimates via the parametric bootstrap. *Political Analysis 17*(3), 261–275.

Cattell, Raymond B (1966). The scree test for the number of factors. *Multivariate behavioral research 1*(2), 245–276.

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 and Keith T Poole (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 and David Rohde (2012). Political institutions and public choice roll-call database. retrieved from https://ou.edu/carlalbertcenter/research/pipc-votes/.

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

Ghahramani, Zoubin and Thomas L Griffiths (2006). Infinite latent feature models and the Indian buffet process. In *Advances in neural information processing systems*, pp. 475–482.

Harbridge, Laurel (2015). *Is Bipartisanship Dead?: Policy Agreement and Agenda-setting in the House of Representatives*. Cambridge University Press.

Heckman, James J and James M Snyder Jr (1996). Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators. Technical report, National Bureau of Economic Research.

Hurwitz, Mark S (2001). Distributive and partisan issues in agriculture policy in the 104th House. *American Political Science Review 95*(4), 911–922.

Jessee, Stephen and Neil Malhotra (2010). Are congressional leaders middlepersons or extremists? yes. *Legislative Studies Quarterly 35*(3), 361–392.

Kim, In Song , John Londregan, and Marc Ratkovic (2018). Estimating Spatial Preferences from Votes and Text. *Political Analysis 26*(2), 210–229.

Knowles, David and Zoubin Ghahramani (2011). Nonparametric Bayesian sparse factor models with application to gene expression modeling. *The Annals of Applied Statistics*, 1534–1552.

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

Lauderdale, Benjamin E and Alexander Herzog (2016). Measuring political positions from legislative speech. *Political Analysis 24*(3), 374–394.

Lee, Frances E (2009). *Beyond ideology: Politics, principles, and partisanship in the US Senate*. University of Chicago Press.

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

McKelvey, Richard D (1976). Intransitivities in multidimensional voting models and some implications for agenda control. *Journal of Economic theory 12*(3), 472–482.

Norton, Noelle H (1999). Uncovering the dimensionality of gender voting in Congress. *Legislative Studies Quarterly*, 65–86.

Paisley, John and Lawrence Carin (2009). Nonparametric factor analysis with beta process priors. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 777–784.

Poole, Keith T and Howard Rosenthal (1997). *Congress: A political-economic history of roll call voting*. Oxford University Press on Demand.

Poole, Keith T and Howard L Rosenthal (2011). *Ideology and congress*, Volume 1. Transaction Publishers.

Ramey, Adam (2016). Vox populi, vox dei? Crowdsourced ideal point estimation. *The Journal of Politics 78*(1), 281–295.

Riker, William H (1980). Implications from the disequilibrium of majority rule for the study of institutions. *American Political Science Review 74*(2), 432–446.

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

Schofield, Norman (1978). Instability of simple dynamic games. *The Review of Economic Studies 45*(3), 575–594.

Shepsle, Kenneth A (1978). *The giant jigsaw puzzle: Democratic committee assignments in the modern House*. University of Chicago Press Chicago.

Shepsle, Kenneth A and Barry R Weingast (1981). Structure-induced equilibrium and legislative choice. *Public choice 37*(3), 503–519.

Smith, Steven S (2007). *Party influence in Congress*. Cambridge University Press.

Tahk, Alexander (2018). Nonparametric ideal-point estimation and inference. *Political Analysis 26*(2), 131–146.

Tausanovitch, Chris and Christopher Warshaw (2017). Estimating candidates' political orientation in a polarized congress. *Political Analysis 25*(2), 167–187.

Part I

Appendix(For Online Publication Only)

Table of Contents

THE BPIRT MODEL AND ESTIMATION

Beta Processes

A beta process is a random discrete measure that is completely described by a countably infinite set of atoms, where each atom has a finite mass determined from a stick-breaking process (Hjort 1990). Unlike the well-known Dirichlet process (Ferguson 1973), the probabilities that an individual unit belongs to a set of potential groups do not have to sum to one. Rather, the masses must only be between zero and one. The beta process is then used as a base measure for a Bernoulli process. In other words, a beta process yields a stochastic process for binary random variables or *feature selection*.

Definition 1 *Let Ω be a measurable space and $\mathbb{B}$ be its σ-algebra. Let H_0 be be a continuous probability measure on $(\Omega, \mathbb{B})$ and α a positive scalar. Assume that Υ can be divided into K disjoint partitions, $(B_1, B_2, ..., B_K)$. The corresponding beta process is generated as:*

$$H(B_k) \sim Beta(\alpha H_0(B_k), \alpha(1 - H_0(B_k))) \tag{A1}$$

where $Beta(\cdot, \cdot)$ corresponds to the standard two-parameter beta distribution. Allow $K \to \infty$ and $H_0(B_k) \to 0$, then $H \sim BP(\alpha H_0)$.

The beta process can be written in set-function form:

$$H(\nu) = \sum_{k=1}^{\infty} \pi_k \delta_{\nu,k}(\nu) \tag{A2}$$

where $H(\nu_i) = \pi_i$ and $\delta_{\nu,k}(\nu)$ is an arbitrary measure on ν. In the case of the beta process, π does not serve as a PMF. Rather, π serves as part of a new measure that parameterizes a Bernoulli process:

Definition 2 *Let the column vector, r_j, be infinite and binary with the k^{th} value, $r_{j,k}$:*

$$r_{i,k} \sim Bern(\pi_k) \tag{A3}$$

The new measure on the measurable space, Υ, is drawn from a Bernoulli process.

By arranging the samples for a set of infinite vectors as a matrix, we can see that a beta process is a prior over an infinite binary matrix with each row corresponding to a location in the measurable space.

BPIRT Full Model Specification

Beginning with the likelihood of the data, a full model specification for the BPIRT model can be defined. First, recall that the binary random variable is projected to a latent continuous space through data augmentation (Albert and Chib 1993) such that:

$$x_{i,j} \sim \begin{cases} \mathcal{TN}_{-\infty,0}(\boldsymbol{\lambda_j \omega_i} - \alpha_j, 1) \text{ if } y_{i,j} = 0 \\ \mathcal{TN}_{0,\infty}(\boldsymbol{\lambda_j \omega_i} - \alpha_j, 1) \text{ if } y_{i,j} = 1 \\ \mathcal{N}(\boldsymbol{\lambda_j \omega_i} - \alpha_j, 1) \text{ if } y_{i,j} \text{ is missing} \end{cases} \tag{A4}$$

Then, the BPIRT model can be defined as:

$$\begin{aligned} &P(x_{i,j}|-) \sim \mathcal{N}((\boldsymbol{r}_j \odot \boldsymbol{\lambda}_j)\boldsymbol{\omega}_i - \alpha_j, 1) \\ &P(\boldsymbol{\omega}_i) \sim \mathcal{N}_K(\boldsymbol{0}, \boldsymbol{I}_K) \\ &P(\lambda_{j,k}|r_{j,k}) \sim r_{j,k}\mathcal{N}_p(\lambda_{j,k}; 0, \gamma_k^{-1}) + (1 - r_{j,k})\delta_0 \\ &P(r_{j,k}) \sim \text{Bern}(\pi_k) \\ &P(\pi_k) \sim \text{Beta}(a/K, b(K-1)/K) \\ &P(\gamma_k) \sim \text{Gamma}(c, d) \end{aligned} \tag{A5}$$

In all cases, intentionally vague or improper uniform priors are used on the structural parameters. Similarly, conjugate priors are used for convenience in estimation. While there is debate as to the impact of these choices (Murray et al. 2013), simulation shows that these choices are relatively innocuous given the size of the standard roll call data set.

Under a large, but finite K that approximates an infinite dimensional representation of $\boldsymbol{R}$, the model can be estimated with an explicit beta-Bernoulli prior on the elements of the binary matrix (Paisley and Carin 2009; Doshi et al. 2009). If $K = \infty$, then we use the Indian Buffet Process prior (Ghahramani and Griffiths 2006). I choose to use the explicit infinite approach in this article.

MCMC For BPIRT

Estimation of the BPIRT model uses the following sampling routine (Knowles and Ghahramani 2011):

1. **Sample Continuous Mappings for the Binary Random Variables, $\boldsymbol{X}$.** For each $i \in (1, \ldots, N)$ and $j \in (1, \ldots, P)$, sample $x_{i,j}$ from a truncated normal distribution according to:

$$x_{i,j} \sim \begin{cases} \mathcal{TN}_{-\infty,0}(\boldsymbol{\lambda_j \omega_i} - \alpha_j, 1) \text{ if } y_{i,j} = 0 \\ \mathcal{TN}_{0,\infty}(\boldsymbol{\lambda_j \omega_i} - \alpha_j, 1) \text{ if } y_{i,j} = 1 \\ \mathcal{N}(\boldsymbol{\lambda}_j \boldsymbol{\omega}_i - \alpha_j, 1) \text{ if } y_{i,j} \text{ is missing} \end{cases} \tag{A6}$$

2. **Sample R and Λ jointly.**
Sampling Currently Observed Features
Define K^+ as the current number of active features. For each $j \in (1, \ldots, p)$ and $k \in (1, \ldots, K^+)$ define:

$$\begin{aligned} t_{j,k} &= \frac{P(r_{j,k} = 1|\boldsymbol{X}, -)}{P(r_{j,k} = 0|\boldsymbol{X}, -)} \\ &= \frac{P(\boldsymbol{X}|r_{j,k} = 1, -)}{P(\boldsymbol{X}|r_{j,k} = 0, -)} \frac{P(r_{j,k} = 1)}{P(r_{j,k} = 0)} \end{aligned} \tag{A7}$$

$$\frac{P(\boldsymbol{X}|r_{j,k}=1,-)}{P(\boldsymbol{X}|r_{j,k}=0,-)}=\sqrt{\frac{\gamma_k}{\gamma}}\exp\left(\frac{1}{2}\gamma\mu^2\right) \tag{A8}$$

$$\frac{P(r_{j,k}=1)}{P(r_{j,k}=0)}=\frac{m_{-j,k}}{P-m_{-j,k}+1} \tag{A9}$$

where $\gamma=\boldsymbol{\omega}_k'\boldsymbol{\omega}_k+\gamma_k$, $\mu=\frac{1}{\gamma}\boldsymbol{\omega}_k'\hat{\boldsymbol{E}}_j$, $\hat{\boldsymbol{E}}_j=\boldsymbol{x}_j-\boldsymbol{\lambda}_j\boldsymbol{\Omega}+\alpha_j$ setting $\lambda_{j,k}=0$, and $m_{-j,k}=-r_{j,k}+\sum_{h=1}^{p} r_{h,k}$. Let

$$p_{r=1}=\frac{t_{j,k}}{1+t_{j,k}}$$

then sample $P(r_{j,k}|-)\sim\text{Bern}(r_{j,k};p_{r=1})$. If $r_{j,k}=1$, then sample $P(\lambda_{j,k}|-)\sim\mathcal{N}(\lambda_{j,k};\mu,\gamma^{-1})$. Otherwise, set $\lambda_{j,k}=0$.

Sampling New Features

Begin by sampling the two Indian Buffet Process hyperparameters, a and b. Sample a from the full conditional:

$$P(a|-)\sim\text{Gamma}\left(a;K^+,H_P(b)\right) \tag{A10}$$

where $H_P(b)=\sum_{h=1}^{P}\frac{b}{b+h-1}$. b must be drawn via a Metropolis-Hastings step. Draw a new proposal:

$$P(b^*)\sim\text{Gamma}(b^*,1,1) \tag{A11}$$

Accept b^* with probability $\min(1,r_{b\to b^*})$:

$$r_{b\to b^*}=\frac{(ab^*)^{K^+}\exp[-aH_P(b^*)]\prod_{k=1}^{K^+}\mathbb{B}(m_k,P-m_k+b^*)}{(ab)^{K^+}\exp[-aH_P(b)]\prod_{k=1}^{K^+}\mathbb{B}(m_k,P-m_k+b)} \tag{A12}$$

where $\mathbb{B}(\cdot,\cdot)$ is the Beta function and $m_k=\sum_{j=1}^{P} r_{j,k}$.

For each $j\in(1,\ldots,P)$, sample the new number of dimensions to try:

$$P(\kappa_j)\sim\text{Pois}\left(\kappa_j;\frac{ab}{b+P-1}\right) \tag{A13}$$

Knowles and Ghahramani (2011) discuss ways to make this proposal explore the feature space in a faster way.

Draw values for each of the new potential dimensions. For $q\in(1,\ldots,\kappa_j)$:

$$P(\lambda_{j,q})\sim\mathcal{N}(\lambda_{j,q};0,1) \tag{A14}$$

which will be referred to collectively as $\boldsymbol{\lambda}_{j,\kappa_j}$.

Using this as the proposal for a Metropolis-Hastings draw, accept the new dimensions with probability $\min(1,r_{\eta\to\eta^*})$:

$$r_{\eta\to\eta^*}=(2\pi)^{(N\kappa_j)/2}|\boldsymbol{M}|^{-N/2}\exp\left[.5\sum_{i=1}^{N}\boldsymbol{m}_i'\boldsymbol{M}\boldsymbol{m}_i\right] \tag{A15}$$

where $\boldsymbol{M} = \boldsymbol{\lambda}'_{j,\kappa_j}\boldsymbol{\lambda}_{j,\kappa_j} + \boldsymbol{\mathcal{I}}_{\kappa_j}$, $\boldsymbol{m}_i = \boldsymbol{M}^{-1}\boldsymbol{\lambda}_{j,\kappa_j}\hat{E}_{i,j}$, and $\hat{\boldsymbol{E}} = \boldsymbol{X} - \boldsymbol{\Lambda\Omega} + \boldsymbol{\alpha}$.

If the new proposal is accepted, set $\boldsymbol{\lambda}_{j,(K^{+}+1:K^{+}+\kappa_j)}$ to the proposed values. Scheduling this part of the algorithm after updating values of $\boldsymbol{\Lambda}$ improves mixing. Set K^{+} to the new number of columns represented in $\boldsymbol{\Lambda}$.

3. **Remove Inactive Features and Normalize $\boldsymbol{\Lambda}$.** For each $k \in (1, ..., K^{+})$, if $r_{j,k} = 0 \;\forall\; 1 \leq j \leq p$, remove k from the analysis. Recalculate K^{+}. Post-process $\boldsymbol{\Lambda}$ to normalize the variance. For each $j \in (1, ..., p)$ and $k \in (1, ..., K^{+})$ set $\boldsymbol{\lambda}_{j,k}$:

$$\lambda_{j,k} = \frac{\lambda_{j,k}}{\sqrt{1 + \sum_{h=1}^{K^{+}} \lambda_{j,h}^{2}}} \tag{A16}$$

4. **Sample $\boldsymbol{\Omega}$.** For each $i \in (1, ..., n)$, sample $\boldsymbol{\omega}_i \in \mathbb{R}^{K^{+}}$ from:

$$P(\omega_i|-) \sim \mathcal{N}_{K^{+}}(\omega_i; (\boldsymbol{\Lambda}'\boldsymbol{\Lambda} + \boldsymbol{\mathcal{I}}_{K^{+}})^{-1}\boldsymbol{\Lambda}'\boldsymbol{y}_i, (\boldsymbol{\Lambda}'\boldsymbol{\Lambda} + \boldsymbol{\mathcal{I}}_{K^{+}})^{-1}) \tag{A17}$$

5. **Sample Item Level Intercepts, $\boldsymbol{\alpha}$.** For each $j \in (1, ...p)$, sample the item level intercept from:

$$P(\alpha_j|-) \sim \mathcal{N}\left(\bar{\mu}_j, \frac{1}{N^2}\sum_{i=1}^{N}(\mu_{i,j} - \bar{\mu}_j)^2\right) \tag{A18}$$

where $\mu_{i,j} = \boldsymbol{\lambda}'_j\omega_i - x_{i,j}$ and $\bar{\mu}_j = \frac{1}{N}\sum_{i=1}^{N}\mu_{i,j}$.

6. **Sample Factor Precisions, γ_k.** For each $k \in (1, ..., K^{+})$, sample γ_k from:

$$P(\gamma_k|-) \sim \text{Gamma}\left(\frac{m_k}{2}, \sum_{j=1}^{p}\lambda_{j,k}^2\right) \tag{A19}$$

where m_k is the number of sources for which feature $k \in (1, ..., K)$ is active.

Identification of Structural Parameters in BPIRT

One model consideration which requires further examination relates to identification of the structural parameters. It is well known that ideal points estimated using latent variable models are unidentified with further constraints (Rivers 2003). Identification can be induced by placing constraints $K(K+1)$ ideal points or requiring that the matrix of discrimination parameters have a lower-block triangular structure with positive elements on the main diagonal (Geweke and Zhou 1996).

Using BPIRT, neither of these approaches are viable - since the number of columns in the matrix of discrimination parameters is technically infinite, placing *a priori* constraints is not possible. Fortunately, the sparsity inducing beta process prior on the discrimination parameters ensures identification. First, the number of votes which take on a feature necessarily decreases as the cardinality of the feature increases (Bhattacharya et al. 2011). For example, if $r_{j,1} = 1 \;\forall\; j \in (1, ..., P)$ and $\sum_{j=1}^{P} r_{j,1} = P$, then $\sum_{j=1}^{P} r_{j,1} < P$. This ensures that $\sum_{j=1}^{P}(1 - r_{j,k}) \geq k \;\forall\; k \in (1, ..., \infty)$. Second, the spike-and-slab priors on the matrix of discrimination parameters ensures that each element of this matrix has a posterior distribution completely contained on one side of zero (Knowles and Ghahramani 2011). Together, these two features effectively mimic the requirements for identification presented by Geweke and Zhou (1996) and ensure that all structural parameters estimated using BPIRT are uniquely identified. This is echoed by examining the posterior distributions for the ideal points, which never exhibit bimodality, evidence that the sign-switching problem is not present in MCMC estimation.

Assessing Convergence for the BPIRT Algorithm

Convergence of the vote level difficulty parameters can be assessed using routine convergence diagnostics included in the `R` package `superdiag` (Tsai et al. 2012). The number of dimensions retrieved from the BPIRT procedure can also be monitored using standard convergence diagnostics. However, due to the discrete nature of this value, it is often more beneficial to assess convergence using visual inspections. Assessing the convergence of the difficulty parameters, ideal points, and elements of the binary matrix proves more challenging due to the varying dimension nature of the set of estimates that make up each set of structural parameters. Convergence for these parameters is not directly assessed. Rather, convergence diagnostics are performed on the mean of the latent distribution passed to the data augmentation step - $(\boldsymbol{r}_j \odot \boldsymbol{\lambda}_j)\omega_i - \alpha_j$. Similarly, I monitor the log-likelihood of the data given the implied model at each step and use this to assess convergence of the posterior distribution of the log-likelihood of the data. Under a converged model, the log-likelihood should be 1) unimodal and approximately symmetric and 2) show behavior that appears as a random walk over iterations. Both of these conditions are generally met when allowing the procedure to have a long burn-in. There is relatively low autocorrelation between draws when the stationary distribution is reached, so the parameter space is explored relatively quickly. Similarly, the posterior distributions of interest are normally distributed due to the conjugacy of the model. Thus, the number of draws needed to truthfully represent the posterior distributions are relatively low.

Methods for Achieving Faster Convergence

As with many MCMC procedures, setting good starting values is a key aspect to achieving fast convergence to the stationary posterior distribution. Using the matrix of binary random variables, I used principal components to put together a reasonable set of starting values. For any values that were missing, I used multiple imputation to quickly fill-in the missing values. I then ran PCA on this full matrix and kept Pois(1) of the scores and loadings as the latent variables and discrimination parameters, respectively. I always started the difficulty parameters at 0 and set 50% of the elements of each column of $\boldsymbol{R}$ to 1. The variance parameters are always started at 1.

One potential problem arises at the beginning of a MCMC chain - if the starting values are particularly bad and the RNG is not giving favorable initial draws from the infinite part of the feature sampler, it is possible for the number of features at the end on an MCMC iteration to move to zero. This is problematic. In order to prevent this behavior, I chose to begin each chain of the MCMC procedure with 100 iterations where the IBP prior 1) did not look at the number of other votes which took on a feature when determining if it would take the feature (i.e. setting Equation A9 equal to 1 for all votes) and 2) did not use the infinite search over the feature space. This creates a period where the model simply learns the ideal points over a fixed number of dimensions. Once the ideal points begin to sort, the rest of the model runs smoothly and there is never less than 1 dimension in the analysis.

Simulation Exercises

In order to understand how BPIRT estimates the binary matrix that encodes each vote's dimensionality and, in turn, the dimensionality of the underlying policy space, I ran simulations on data sets with known parameters and examined how BPIRT uncovers the true underlying latent structure of the data.

The purpose of these simulations is two-fold. First, it is necessary to examine how accurately BPIRT recovers the binary matrix associated with the matrix of discrimination parameters from a known data generating process as a function of the number of voters, number of votes, and the true underlying dimensionality of the latent policy space. Given that the goal of BPIRT is to recover the vectors that dictate which votes correspond to which dimensions of the underlying policy space, it is important to understand when the procedure succeeds in providing an accurate representation of this data. Second, and closely related to the first goal, it is important to understand the properties of the Indian Buffet Process prior across varying numbers of votes, voters, and true underlying dimensionalities. In particular, it is important to examine the ability of the IBP prior to recover the true number of dimensions and to ensure that BPIRT does not uncover *spurious* sets of features that do not correlate with the true underlying feature set. While theory dictates that the IBP will uncover the exact solution when the number of votes and voters is large, roll call data sets are inherently limited in the number of voters and the number of votes made within a session. Thus, understanding the small and medium sample properties of this nonparametric prior is of interest.

In order to simulate data that has a similar structure to actual roll call data, I used PCA on a set of 928 votes made by 428 members of the 105^{th} session of the U.S. House to estimate a seven-dimensional covariance matrix. This covariance matrix was used to generate simulated roll call data sets with 100/250/400 voters, 100/450/900 votes, and 1/3/5/7 true underlying dimensions in the latent policy space. These simulated data sets were then passed to an implementation of BPIRT and the structural parameters were estimated. Each Markov Chain Monte Carlo routine was run with a burnin of 5000 iterations and 1000 unthinned draws were taken from the stationary posterior distribution over 2 chains. There were no indications of convergence issues in these simulations.

I first examine the relationship between number of voters, number of votes, the true dimensionality of the vote set, and the dimensionality uncovered by BPIRT using the Indian Buffet Process prior. The number of dimensions uncovered by BPIRT for each simulation set can be seen in Table A1. On first glance, it is easy to see that the behavior of the IBP prior to uncover the correct number of dimensions expectedly depends on the number of votes. As shown by Ghahramani and Griffiths (2006), the number of features represented by the prior increases in $O(\log(P))$. This property is apparent in the simulation sets as the ability to estimate the true dimensionality is contingent on having a large number of votes. This relationship is also seen in the number of voters, though not as strongly. This property makes sense, as one would expect that more observations would lead to more accurate estimation of model parameters. However, the number of dimensions estimated appears to be capped in the number of votes.

A second important observation is that the model is **conservative** in its estimation of new dimensions when the number of votes or voters is low. In all cases, the number of estimated dimensions is lower than the truth with small roll call voting data sets resulting in a one dimensional posterior. On the other hand, when presented with a data set that is truly one-dimensional, BPIRT accurately estimates that only one dimension is needed. This finding should assuage concerns that BPIRT uncovers spurious dimensions. All in all, BPIRT provides a useful tool for estimating the dimensionality of the underlying

TABLE A1. Number of Dimensions Estimated From Simulated Data With Various Numbers of Voters, Votes, and Known Dimensionalities Using BPIRT

Voters \ Votes	100	450	900
100	1	1	1
250	1	1	1
400	1	1	1

(a) True Dimensionality = 1

Voters \ Votes	100	450	900
100	1	1	1
250	1	3	3
400	1	3	3

(b) True Dimensionality = 3

Voters \ Votes	100	450	900
100	1	1	1
250	1	4	4
400	1	4	5

(c) True Dimensionality = 5

Voters \ Votes	100	450	900
100	1	2	3
250	2	5	6
400	2	5	6

(d) True Dimensionality = 7

Note: Values reported are posterior modes. In almost every case, the posterior for number of dimensions converged to a single value.

TABLE A2. Proportion of Elements in R Correctly Classified From Simulated Data With Various Numbers of Voters, Votes, and Known Dimensionalities Using BPIRT

Voters \ Votes	100	450	900
100	0.93	0.99	1.00
250	0.97	1.00	1.00
400	0.97	1.00	1.00

(a) True Dimensionality = 1

Voters \ Votes	100	450	900
100	0.77	0.73	0.73
250	0.80	0.81	0.84
400	0.80	0.87	0.89

(b) True Dimensionality = 3

Voters \ Votes	100	450	900
100	0.81	0.77	0.77
250	0.82	0.84	0.86
400	0.83	0.89	0.91

(c) True Dimensionality = 5

Voters \ Votes	100	450	900
100	0.84	0.82	0.84
250	0.86	0.87	0.89
400	0.87	0.91	0.93

(d) True Dimensionality = 7

Note: Values reported are posterior medians.

policy space. In particular, it is well suited to examine whether or not a roll call data set requires only one dimension to explain variance within the data set.

I also examine BPIRT's ability to correctly uncover structural zeros in the binary matrix. Recall that zeros in this matrix imply that a specific vote does not rely on variance explained by a given dimension when explaining the underlying utility functions that lead to certain vote outcomes. The proportion of correctly classified elements of **R** for each simulation set is shown in Table A2.[1] The relationship between the number of votes, number of voters, and accuracy in recovering elements of the binary matrix is similar to the one seen in Table A1 - more votes and more voters results in more accurate estimation of the structural parameters in **R**. However, unlike in the previous case, the accuracy of estimation seems to be driven by the number of voters. This finding makes sense, however, as each estimate within **R** is related to a specific vote/dimension combination. Thus, more voters means more information about which sources of variation best describe the vote. Even in smaller samples, BPIRT provides an accurate representation of the binary matrix. This is especially apparent in estimations with a true one-dimensional model. All in all, these simulations show that BPIRT can accurately recover the underlying structures which drive voting under the varying dimensions model of vote choice.

MULTIDIMENSIONALITY IN THE U.S. CONGRESS OVER TIME

For analysis in this section, I examine the roll call voting data sets for the $1^{st} - 115^{th}$ (1789 – 2017) sessions of both chambers of the U.S. Congress, separately. Over the set of roll call votes in each session, I analyzed votes that had at least 5 votes in the minority and voters that registered roll call votes for at least 50% of the votes analyzed. I chose to run the BPIRT procedure on each roll call data set for a burnin of 20,000 iterations and capture 1000 draws of the parameters from the stationary posterior distribution over two independently initialized chains. Assessments of convergence both within and across chains showed no evidence of lack of convergence.

The proportion of variance explained by a dimension for a vote can be derived using the properties of the BPIRT model. Recall that the marginal probability of the augmented data under BPIRT is:

$$P(\boldsymbol{x}_i) \sim \mathcal{N}_P(\boldsymbol{\alpha}, (\boldsymbol{R} \odot \boldsymbol{\Lambda})(\boldsymbol{R} \odot \boldsymbol{\Lambda})' + \mathcal{I}_P) \tag{A20}$$

Note that under the marginal posterior, each voter has the same probability density function. This implies that the variance of the augmented data for a vote is:

$$V[\boldsymbol{x}_j] = \sum_{k=1}^{K} r_{j,k}\lambda_{j,k}^2 + 1 \tag{A21}$$

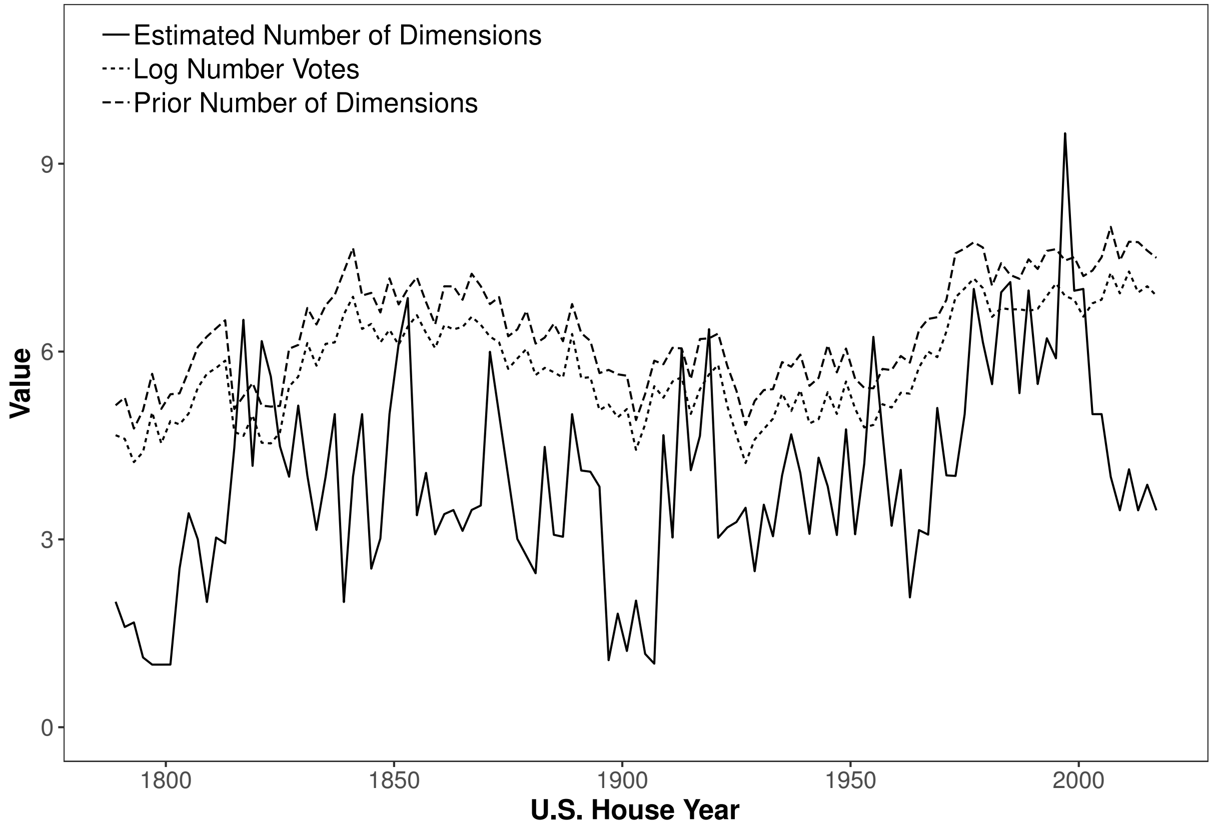
Then, the proportion of variance explained (PVE) by a dimension on a vote can be defined as:

$$\text{PVE}_{j,k} = \frac{r_{j,k}\lambda_{j,k}^2}{\sum\limits_{h=1}^{K} r_{j,k}\lambda_{j,h}^2} \tag{A22}$$

Since each of these quantities presented in this section are, themselves, uncertain measures, the posterior means are also associated with a 95% highest posterior density interval. In practice, these ranges are very tightly bunched around the posterior means. As these quantities do not plot well, I have chosen not to include the error bars in many of the figures. These quantities can be found in the replication materials. The inclusion of 95% HPD intervals do not change the overall conclusions made from any of the graphs included in this section.

[1]When constructing this data, I attempted to recreate the number and structure of dimensions that are typically seen in roll call voting. This led to seven dimensions with a decreasing number of votes which took on each dimension. Every vote took on the first dimension while other dimensions were only required for a proportion of votes. This leads to the high hit rate when the true number of dimensions is one and the decrease in hit rate between one and three.

FIGURE A1. Relationship Between Number of Votes, Estimated Number of Dimensions, and Prior Number of Dimensions Implied by IBP for each Session of the U.S. House



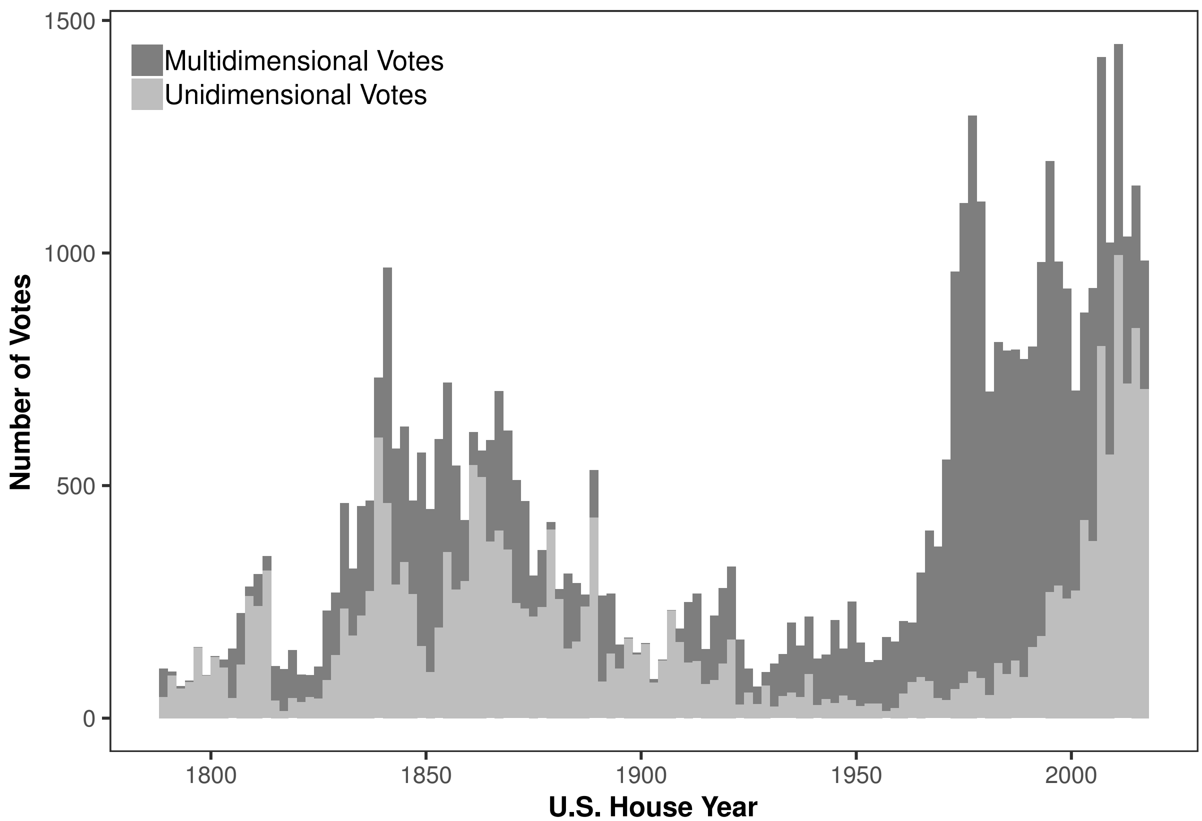
IBP and Dimensions Observed in the U.S. House

For many sessions of the U.S. House and U.S. Senate, the posterior distribution for the number of dimensions converged strongly on a single value. These sessions are indicated in Figure 1 by points with no error bars. Exploring the infinite set of features using the beta process prior is contingent on tuning parameters, which are outlined in the estimation procedure presented in the Appendix. Similarly, the discrete nature of the posterior distribution for number of dimensions can lead to results that appear to not have reached stationarity. Given the performance of BPIRT in simulation exercises, there is strong evidence that these values are equal to or below the truth and assuage any concerns related to not fully exploring the posterior. Other choices for these hyperparameters lead to results that share the same mode but have higher values included in the 95% HPDs.

The relationship between the number of voters, number of votes, and estimated dimensionality is well established in the simulation section. However, it is important to see how these relationships manifest in the actual roll call data used in these applied examples. Similarly, one might question whether or not these results are unduly driven by the choice of priors for the hyperparameters of the IBP prior. To address these concerns, I checked the relationship between the number of votes, the expected number of dimensions drawn from the IBP prior, and the estimated dimensionality.

Figure A1 shows the logarithm of the number of votes analyzed, the posterior mean number of dimensions estimated, and the number of dimensions implied by the IBP prior using posterior means values for the IBP hyperparameters for each session of the U.S. House. First, it is easy to see that

FIGURE A2. Number of Unidimensional and Multidimensional Votes Analyzed in Each Session of the U.S. House

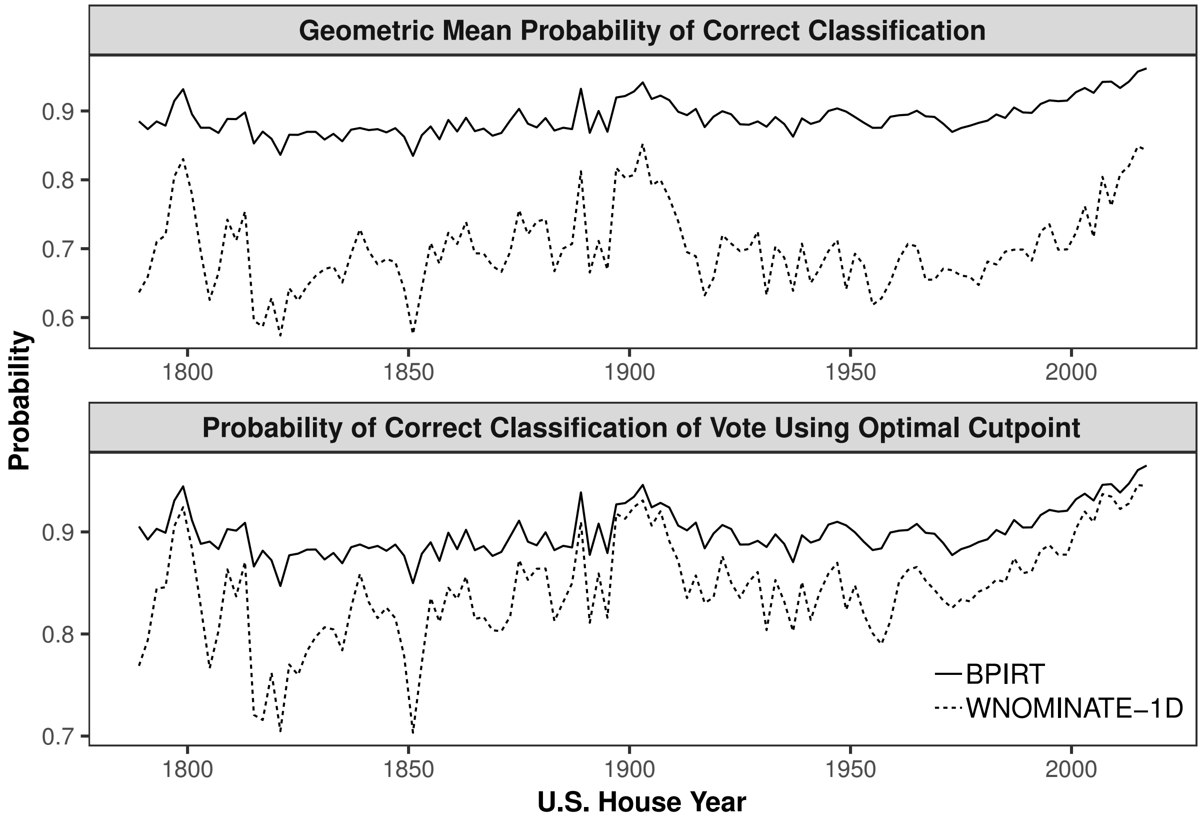


there is s strong dependence in the number of votes and the prior number of dimensions implied by BPIRT - the prior number of dimensions scales almost perfectly with the logarithm of the number of votes with an additive constant implied by the IBP hyperparameters. On the other hand, there is not a strong correlation between the posterior mean number of dimensions estimated by BPIRT and the prior expected number of dimensions. This implies that the choice of prior is not unduly influencing the number of dimensions estimated by BPIRT. However, the choice of prior does seem to place a cap on the number of dimensions which can be estimated; the number of dimensions estimated rarely goes above the number of dimensions implied by the prior. This behavior is expected due to the properties of the IBP prior discussed in the simulations.

More Summaries of Multidimensional Voting in the U.S. House

Figure A2 shows the number of votes that were analyzed for each session of the U.S. House. These votes are then classified as unidimensional or multidimensional by the posterior mean probability that a vote requires more than the first dimension to best explain individual vote variation. This figure demonstrates two key concepts. First, the sheer number of votes that occur within each session dramatically increase after 1950. This allows BPIRT to better estimate the number of dimensions needed to model the roll call data set and, in turn, allows for more multidimensionality in votes to appear. Second, this plot makes it clear that the proportion of votes that were classified as multidimensional by BPIRT also dramatically increases after 1950. While there is a downturn in this proportion in recent times, these

FIGURE A3. Aggregate GMP and Proportion Correctly Classified Votes for the 1st - 115th Sessions of the U.S. House



proportions are generally on par with the proportion of votes that were classified as unidimensional in the mid-1800s. This is another way of showing that the unidimensional, party bloc voting of recent times is not a unique occurrence.

GMP is only one way in which the quality of ideal point models is assessed. The most common way in which models are compared is through correct classifications metrics. Using the ideal points and other structural parameters, the proportion of votes that are correctly classified can be used to demonstrate the ability of the model to partition "Yea" and "Nay" votes appropriately in different situations. This metric is a natural fit for predictive models like NOMINATE, but does not necessarily account for the uncertainty associated with each of the model parameter estimates (Carroll et al. 2009). The proportion of votes correctly classified using optimal cutpoints for BPIRT was estimated using the mean of the augmented posterior - $\mu_{i,j} = (r_j \odot \lambda_j)\omega_i - \alpha_j$. If $\mu_{i,j} < 0$, then the vote was a predicted "Nay" and "Yea" otherwise. A similar calculation is used for NOMINATE.

Figure A3 shows the aggregate GMP and proportion correct classification under the optimal cutpoint for both BPIRT and one-dimensional WNOMINATE model. Beginning with correct classification, it is easy to see that BPIRT and a unidimensional WNOMINATE model yield similar results throughout much of U.S. history, especially in recent sessions. While this could certainly be taken as evidence that the unidimensional model is sufficient, correct classification done in this sense is theoretically deficient and ignores the inherent uncertainty associated with ideal point measures while encouraging heavily overfit models (Aldrich et al. 2014; Roberts et al. 2016). Ideal points are rarely used to predict new votes; in fact, ideal points are almost always used to explain the voting behavior given the entirety of

TABLE A3. Correlation between Dimensions Estimated for the 107th U.S. House

	Party Loyalty	Procedural	Security	Budget	Rural	Foreign	DOD
Party Loyalty	1.00	0.30	0.28	0.33	-0.07	-0.09	0.17
Procedural	0.30	1.00	0.35	0.05	-0.03	-0.07	0.25
Security	0.28	0.35	1.00	0.10	0.09	-0.04	0.16
Budget	0.33	0.05	0.10	1.00	0.13	-0.01	0.05
Rural	-0.07	-0.03	0.09	0.13	1.00	0.04	-0.06
Foreign	-0.09	-0.07	-0.04	-0.01	0.04	1.00	0.01
DOD	0.17	0.25	0.16	0.05	-0.06	0.01	1.00

Note: Values reported are posterior means for the Pearson correlation coefficient between estimated ideal points.

votes for an analyzed period. It is imperative that ideal point fits are treated with the same probabilistic rigor that any other inferential technique requires when attempting to explain behavior and, thus, important to consider more statistically rigorous approaches when making choices about the underlying parameters of ideal point models.

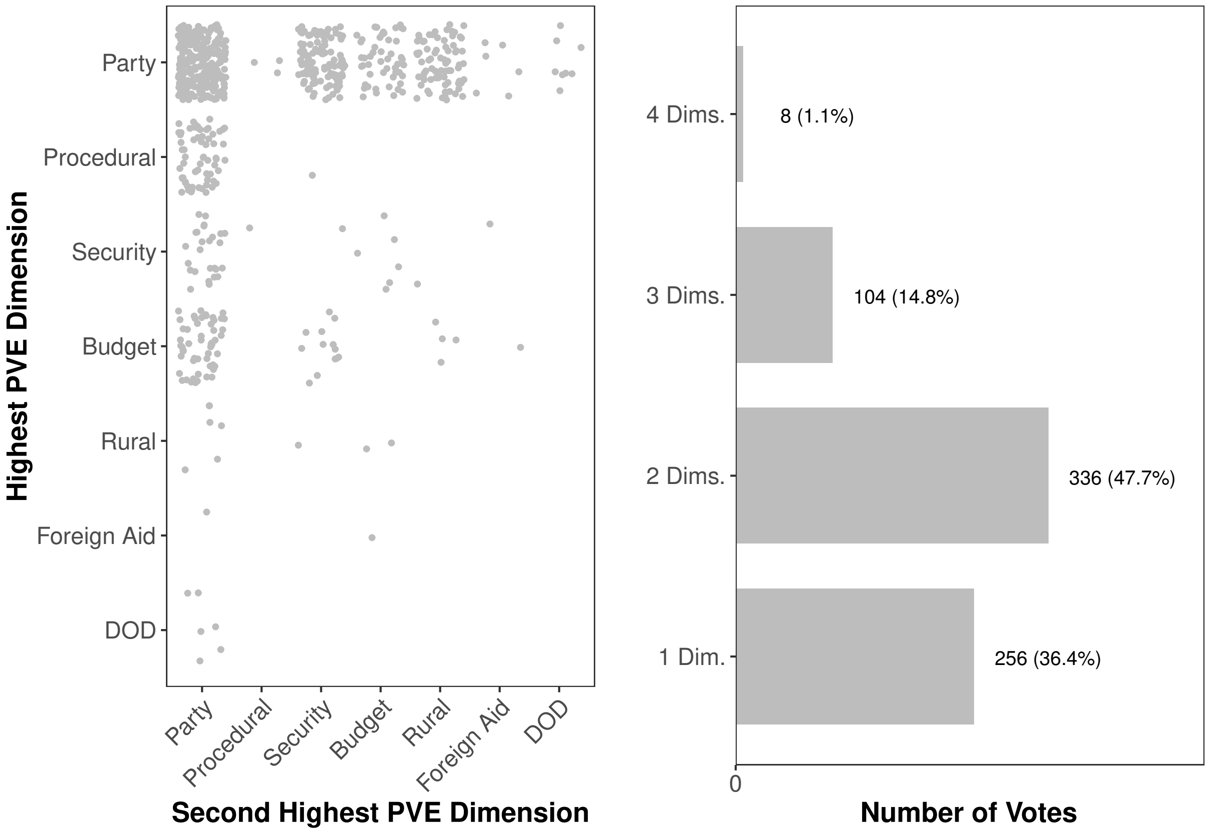
Proportion reduction in error metrics are similar to correct classification metrics (Carroll et al. 2009; Aldrich et al. 2014; Roberts et al. 2016). These approaches are central to previous discussions of how many dimensions are needed to model a roll call voting set. While these approaches provide some information of this topic, they are post-hoc statistics that require *a priori* assumptions about the structure of the underlying latent space. BPIRT estimates the dimensionality and necessity of dimensions at a vote-level within its statistical procedure and, therefore, is incompatible with the notion of adding or subtracting whole dimensions from the latent space. For this reason, I choose to avoid inappropriate attempts to compare APRE and MPRE achieved from BPIRT.

More Summaries for the 107th U.S. House

Names were provided for each of the dimensions estimated by BPIRT for the 107^{th} session of the U.S. House using a suite of tools. First, I used vote classifications from `voteview.com` to analyze the set of votes which had non-zero contributions to a dimension and saw general trends in the content of votes that loaded in each dimension (Poole and Rosenthal 2012). Second, I utilized non-negative matrix factorization and regularized logistic regression to extract important words from the bill summaries associated with each vote on each set of votes. These tools created a general picture of what each dimension was modeling. While still somewhat ad-hoc, this approach defines a general method that will be useful for future research attempting to find trends in what each dimension means, over time.

BPIRT tends to extract dimensions which run relatively orthogonal to one another. One of the advantages of the IBP prior is that it generally prevents against dimensions that are highly correlated with one another from arising during estimation. This can be attributed to the IBP draws, which check to see if there is any new information added by a new dimension while conditioning on previously existing dimensions. Table A3 shows the pairwise correlations between dimensions estimated for the 107^{th} session of the U.S. House. Correlations for this session range from .01 to .35. While this indicates that there are dimensions which are not orthogonal to one another, the correlation is still relatively low.

FIGURE A4. Highest PVE Dimensions and Number of Dimensions For Each Vote in the 107th U.S. House



Note: Values reported are estimates of the binary matrix from the iteration of the MCMC procedure with the highest complete-data likelihood. Points on the main diagonal of the PVE graph required only one dimension to best model variation within the vote.

One advantage that comes from utilizing an IBP prior is that votes are allowed to take on a collection of dimensions rather than one, and only one, dimension. Given the complexity of the topics that are being considered when votes are cast, this is a desirable property. BPIRT models this complexity and presents the set of dimensions which each vote requires to best model its variance in voting. To this end, it is interesting to examine the number of dimensions, and which dimensions, are chosen for votes within the analyzed set. Figure A4 illustrates this dynamic by showing the dimensions with the highest and second highest PVE for each vote analyzed. While around 35% of votes only require the party loyalty dimension to model the roll call votes, the other 65% of votes require at least one other dimension. However, all votes require a non-zero contribution from the party loyalty dimension. This dynamic is partially due to the "rich get richer" property of the IBP prior, the fact that most votes require the party loyalty dimension as the highest or second highest PVE dimension fits well with theories of party control in U.S. legislative voting.

REFERENCES

Albert, James H and Siddhartha Chib (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association 88*(422), 669–679.

Aldrich, John H , Jacob M Montgomery, and David B Sparks (2014). Polarization and ideology: Partisan sources of low dimensionality in scaled roll call analyses. *Political Analysis 22*(4), 435–456.

Bhattacharya, Anirban , David B Dunson, and Others (2011). Sparse Bayesian infinite factor models. *Biometrika 98*(2), 291.

Carroll, Royce , Jeffrey B Lewis, James Lo, Keith T Poole, and Howard Rosenthal (2009). Measuring bias and uncertainty in DW-NOMINATE ideal point estimates via the parametric bootstrap. *Political Analysis 17*(3), 261–275.

Doshi, Finale , Kurt Miller, Jurgen V Gael, and Yee W Teh (2009). Variational inference for the Indian buffet process. In *International Conference on Artificial Intelligence and Statistics*, pp. 137–144.

Ferguson, Thomas S (1973). A Bayesian analysis of some nonparametric problems. *The annals of statistics*, 209–230.

Geweke, John and Guofu Zhou (1996). Measuring the pricing error of the arbitrage pricing theory. *The review of financial studies 9*(2), 557–587.

Ghahramani, Zoubin and Thomas L Griffiths (2006). Infinite latent feature models and the Indian buffet process. In *Advances in neural information processing systems*, pp. 475–482.

Hjort, Nils Lid (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *The Annals of Statistics*, 1259–1294.

Knowles, David and Zoubin Ghahramani (2011). Nonparametric Bayesian sparse factor models with application to gene expression modeling. *The Annals of Applied Statistics*, 1534–1552.

Murray, Jared S , David B Dunson, Lawrence Carin, and Joseph E Lucas (2013). Bayesian Gaussian copula factor models for mixed data. *Journal of the American Statistical Association 108*(502), 656–665.

Paisley, John and Lawrence Carin (2009). Nonparametric factor analysis with beta process priors. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pp. 777–784.

Poole, Keith T and Howard Rosenthal (2012). Voteview. *University of California, San Diego. www. voteview. com. Poole, Keith T*.

Rivers, Douglas (2003). Identification of multidimensional spatial voting models. *Typescript. Stanford University*.

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

Tsai, Tsung-han , Jeff Gill, and Jonathan Ripkin (2012). superdiag: R Code for Testing Markov Chain Nonconvergence.