Disagreement and Dimensionality: A Beta Process Approach to Estimating Dimensionality of Ideal Points in the U.S. Congress

Kevin McAlister[1]

[1] *Department of Political Science, University of Michigan, Ann Arbor 48109*

Abstract

Roll call scaling techniques, such as NOMINATE and IDEAL, are empirical standards for studies of voting behavior within legislative bodies. Though ideal point estimation techniques are frequently used, the theoretical implications of assumptions made in order to empirically estimate ideal points provide cause for concern. One assumption that is frequently leveraged in studies of legislatures is that ideal points are best represented in one or two dimensions. This assumption is key for further usage of ideal points in formal models and examinations of elite polarization. Despite the importance of this assumption, the dimensionality of the ideal point space is often simply

fixed to fit theoretical expectations or tested using subjective post-hoc tests. In this paper, I propose a method for properly modeling the dimensionality of the ideal point space using a beta-Bernoulli Bayesian nonparametric prior. This prior structure allows for the number of dimensions and ideal point estimates to be modeled simultaneously. I apply this model to all sessions of the U.S. House (1^{st}-114^{th}) and show that there is little evidence for the the low dimension conjecture in U.S. roll call scaling models. While the dimensionality of the policy space has decreased over recent sessions, the current estimates are not atypical and previous periods of the U.S. Congress have exhibited similar levels of high party level voting. This paper provides a meaningful examination of dimensionality in the U.S. Congress and shows how seemingly innocuous assumptions in the scaling procedures can lead to inappropriate inferences about legislative behavior.

1 Introduction

Studies of legislative behavior focus upon the relationship between legislative preferences, institutional structure, and legislative outcomes. Spatial models are a frequently used tool for studying these relationships. In a spatial model of voting bodies, policies are represented geometrically and votes occur as a function of individual legislators' *ideal points*. An ideal point represents a legislator's most preferred policy outcome and competing policies are judged based upon their distances from her most preferred policy. Under the assumption of utility-maximizing, rational legislators, the spatial model provides a consistent method for researchers to understand how ideal points and policy lead to specific legislative outcomes.

A common task in the legislative behavior literature is to estimate the set of ideal points for matrix of *roll call* data. In this data, the votes for each legislator on a variety of different proposals are recorded. Then, a scaling procedure is used to determine the ideal points for each legislator (Poole and Rosenthal, 1997; Clinton et al., 2004). Scaling procedures typically seek to represent each policy votes on in the roll call set in a low-dimensional Euclidian space. In turn, this allows estimation of ideal points in the same Euclidian space. Thus, the scaling procedure admits a consistent space in which all votes within the roll call set can be represented. Scaling roll call votes in this way implies that there exists a single policy space in which represents all roll call votes within the analyzed roll call set.

The policy space uncovered by scaling methods encompasses the various complexities of the legislator voting behaviors. While the ideal points, themselves, are generally of interest, the uncovered policy space is also substantively interesting. For example, McCarty et al. (2016) utilize ideal points estimated using NOMINATE methodology (Poole and Rosenthal, 1984) and the corresponding policy space to show increased polarization in elite voting over time. This result (and many others like it) relies on the assumption that meaningful parts of the policy space exist only in one dimension. This low-dimensionality conjecture is a key part of numerous theories relating to changes in Congress over time and is key to many other theories which utilize ideal point estimates.

McCarty et al. (2016) argue that there are between one and two dimensions in most session of Congress. The first dimension projects legislators' votes to a "liberal-conservative" dimension which corresponds mostly to economic issues. The second dimension, if needed, corresponds to social issues of the time, typically questions related to race. Over time, NOMINATE shows that the need for a second dimension has disappeared and most roll call voting behavior can be described by the liberal-conservative dimension. The single dimension argument has been the basis for many formal models and empirical findings about Congress (Aldrich and Battista, 2002; Bafumi and Herron, 2010; Binder, 1999; Cameron, 2000; Cox and McCubbins, 2005; Jessee, 2009, 2010; Krehbiel, 1992). However, many of these results are incredibly sensitive to changes in this assumption; if the dimensionality of the congressional vote choice model is any value greater than one, then median voter theorem no longer holds and the results no longer hold (Kramer, 1973). Thus, strong evidence for the one-dimensional model should be in place.

The low dimensional conjecture has many recent challenges in the roll call literature. Crespin and Rohde (2010) and Roberts et al. (2016) analyze roll call votes in specific issue areas and uncover substantial evidence for a multidimensional voting strategy. Norton (1999) shows the same when the analysis is concentrated on bills related to gender issues. These works along with numerous other studies argue that NOMINATE scores are not necessarily measuring ideology, rather they are measuring a summary of the observed voting behavior of a legislator (Clinton, 2012; Snyder Jr, 1992); a legislator votes in accordance to party pressures and the actual roll call votes are "endogenous to legislative context" (Shepsle and Weingast, 1986).

Research about how party leaders and party structure is particularly important in this area as it posits how the structures within Congress can alter the votes that legislators actually make (Dougherty et al., 2014; Aldrich, 1995). Hurwitz (2001) shows evidence of multidimensionality on agricultural votes which have long been posited to follow geographic bounds rather than strict party bounds. Jenkins (1999, 2000) examines the Confederate Congress to show that without strong party influence, the structure and low-dimensionality results about roll call votes disappears. Even in state legislatures where strong two party systems do no exist, the low-dimensional results do not seem to have much evidence (Welch and Carlson, 1973; Wright and Schaffner, 2002). In a comprehensive review of the U.S. Congress, Lee (2009) argues that much of the structure of scaling estimates is a result of partisan teamsmanship and does not truly represent ideology.

Aldrich et al. (2014) examines this notion and produces a number of simulations which examine how scaling procedures like NOMINATE estimate number of dimensions. Of particular note, the authors show that as the distance between two parties increases in the common space, scaling procedures tend to underestimate the true dimensionality of the space when using proportional reduction of error metrics for making the decision. In other words, when members of parties have a strong "teamsmanship", the dimensionality estimated tends to one. To mitigate these problems, the authors suggest two more subtle approaches to estimating the dimensions of the scaling model and the corresponding dimensionality: rigorous statistical testing for the dimensionality of a space and allowing structural zeroes to be introduced in the loadings matrix in order to prevent the assumption that all roll calls load on all dimensions.

In this paper, I propose a model that addresses the two approaches suggested by Aldrich et al. (2014). Built in a way similar to the Bayesian roll call scaling procedure introduced by Clinton et al. (2004), I propose a beta process item-response theory model. BPIRT provides an alternative scaling technique which explicitly estimates the number of dimensions needed to best model the data along with the meaningful structural parameters. Using a Bayesian specification, this method utilizes a nonparametric beta process prior on the loadings matrix to address both the question of dimensionality and the introduction of structural zeroes that specifically indicate which bills do not load on a given dimension. In selecting the number of factors needed to describe the latent policy space, the beta process prior selects dimensions which have loadings on any of the questions which are statistically different from zero. This technique is in stark contrast to the methods for selecting NOMINATE dimensions, which utilizes *aggregate proportional reduction in error* and a Scree-like procedure to select whether or not a dimension is needed. While dimensions may not add much in terms of reduction in error, this does not mean that they add nothing to the model; this incorrect conflation of small effect size and lack of meaningful contribution to the overall model leads NOMINATE to unfairly favor low-dimensional models.

BPIRT also relaxes the standard requirement that all questions load on all dimensions. While the Scree plot procedure implies that all bills and ideal points must exist in $\mathbb{R}^K$, the beta process prior approach allows for a more nuanced interpretation where each bill can exist in a subset of $\mathbb{R}^K$. This approach allows bills that differ in meaningfully in their respective ideal point content to potentially be represented by different sets of dimensions. Substantively, BPIRT models dimensions that point to disagreement within parties - if the first dimension of the roll call scaling procedure most closely models party teamsmanship (Lee, 2009), the dimensions beyond the first should measure departures from the strict party model.

2 A Model for Roll Call Analysis

For a legislature, assume there are N voting members that vote on P bills over the course of time analyzed. For any given vote $j \in (1, P)$, legislator $i \in (1, N)$ must choose between two alternatives: the proposed bill (A_j) or the status quo (S_j). In the standard context, a "Yea" vote corresponds to a vote for A_j and a "Nay" vote corresponds to a vote for 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 A_j and S_j within the space and a K-dimensional *ideal point*, ω_i, which encapsulates the policy preferences of legislator i.

Legislator i must choose whether to vote for A_j or 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:

$$U_i(A_j) = -\|\omega_i - A_j\|^2 + \eta_{i,j}$$

$$U_i(S_j) = -\|\omega_i - S_j\|^2 + \nu_{i,j}$$

where $\eta_{i,j}$ and $\nu_{i,j}$ are stochastic elements of the utility functions. In other words, she votes for A_j if and only if $U_i(A_j) > U_j(S_j)$.

This model is completely specified if a known structure is placed on $\eta_{i,j}$ and $\nu_{i,j}$. There are three choices in structure that have been used extensively in the literature. Perhaps the best known application, Poole and Rosenthal (1997) specify that these error structures follow a Type-I extreme value distribution giving the model a similar structure to the standard logistic regression model. In another instance, Heckman and Snyder Jr (1996) specify that this error structure has a uniform structure. Finally, Clinton et al. (2004) specify Gaussian errors giving the model a probit structure. Any of these choices lead to a tractable model.

Aside from the choice of error distribution, this model requires assumptions about the error structure in order to be tractable. First and foremost, it is assumed that given the choice of error distribution, each $\eta_{i,j}$ and $\nu_{i,j}$ $\forall\ i \in (1, N)$, $j \in (1, P)$ is an independent and identically drawn value from the respective distribution. Second, the error structures can be jointly structured where $E[\eta_{i,j}] = E[\nu_{i,j}]$ and $V[\eta_{i,j} - \nu_{i,j}] = \sigma_j^2$. These assumptions allow this model to be estimated using corresponding data and statistical specification.

Let $y_{i,j}$ be the vote choice that legislator i makes on proposal j - $y_{i,j} = 1$ if legislator i votes for A_j and $y_{i,j} = 0$ if legislator i votes for S_j. Given the model construction, the probability that legislator i votes for A_j can represented as:

$$
\begin{aligned}
P(y_{i,j} = 1) &= P(U_i(A_j) > U_i(S_j)) \\
&= P(-\|x_i - A_j\|^2 + \eta_{i,j} > -\|x_i - S_j\|^2 + \nu_{i,j}) \\
&= P(\nu_{i,j} - \eta_{i,j} < -\|x_i - S_j\|^2 - \|x_i - A_j\|^2) \\
&= P(\nu_{i,j} - \eta_{i,j} < 2(A_j - S_j)'x_i + S_j'S_j - A_j'A_j) \\
&= \Xi(\lambda_j'\omega_i - \alpha_j)
\end{aligned}
$$

where $\Xi(\cdot)$ is the CDF associated with the chosen error structure, $\alpha_j = (A_j'A_j - S_j'S_j)/\sigma_j^2$, and $\lambda_j = 2(A_j - S_j)/\sigma_j^2$. Note that α and $\mathbf{\Lambda}$ are functions of the difference in point locations of the proposed alternative and the status quo.

This construction admits a corresponding statistical model that allows for estimation of the *structural parameters* α and $\mathbf{\Lambda}$ and the latent variables, $\mathbf{\Omega}$. Using the i.i.d. assumptions about the error structures, a likelihood function can be derived:

$$
\mathcal{L}(\alpha, \mathbf{\Lambda}, \mathbf{\Omega}|\mathbf{Y}) = \prod_{i=1}^{N}\prod_{j=1}^{P} \Xi(\lambda_j'\omega_i - \alpha_j)^{y_{i,j}} \times \left(1 - \Xi(\lambda_j'\omega_i - \alpha_j)\right)^{1-y_{i,j}}
$$

This empirical model is closely related to the item-response theory model that is frequently utilized in the psychometrics and education literatures. Bayesian implementations of this model place priors on all of the structural parameters and estimation proceeds using Markov Chain Monte Carlo methods.

3 Uncertainty in the Number of Factors

In many previous constructions of roll call scaling models, it has inherently been assumed that the dimensionality, K, of the uncovered latent space is known. While theory can dictate the choice of this value, it is often the case that assumptions about the dimensionality of the uncovered space are poorly motivated. Similarly, the dimensionality of the latent space can be of as much interest as the estimates for the structural parameters. In order for proper inferences to be made about dimensionality, tests should be used to estimate this quantity.

Inference on the number of dimensions in a latent variable model is both computationally and conceptually challenging. Perhaps the oldest and most commonly used method for determining the number of dimensions needed in a latent variable model is using metrics related to error reduction and plotting them to perform a Scree test (Cattell, 1966). If eigendecomposition is used to uncover the latent variables, then plotting the eigenvalues associated with each dimension against the dimension number and finding the "elbow" in the plot is used to determine the appropriate number of factors. Similarly, proportion error or proportion reduction in error metrics can be used to make the plot and the same procedure is followed. These tests can be used to make claims about dimensionality, but they are not comparable to the familiar statistical testing framework and can lead to poor inferential choices for number of factors. More statistically rigorous methods are needed to answer this question in a meaningful way.

One of the most intuitive approaches to estimating the number of factors in a Bayesian context is to directly calculate the marginal likelihood for a number of different values of K and select the model with the highest posterior probability (Ando, 2009). While this approach is certainly the most familiar approach, analytically calculating the marginal likelihood is difficult and requires making a number of convenient assumptions with choice of priors. Similarly, given the atypical functional form of the marginal distributions in latent variable models with discrete margins, it would be difficult to create an analytical framework that appropriately models the marginal likelihood.

Other more modern approaches use simulation methods to approximate the Bayes factor, or the ratio of the marginal likelihoods of two competing models (Lee and Song, 2002; Ghosh and Dunson, 2009). These approaches include path sampling and stochastic grid searches. While these approaches require less assumptions than direct calculation of the marginal likelihoods, they require running h different models when comparing h different values of K. Though many advances have been made in the area of Bayesian computing, estimation of factor analysis models can still be time consuming, especially when N becomes large. As N grows larger, we should allow the models to be more complex, and trading computational gains with limitations of the search space can lead to incorrect conclusions about the dimensionality of the latent traits being modeled.

In contrast to the above approaches, the other branch of research in this area has allowed K to vary as a part of the MCMC process used to estimate the factor analysis decomposition. An early approach to this problem utilizes reversible jump MCMC as an estimation method for K (Lopes and West, 2004). RJMCMC works and allows the distribution of K to be explicitly modelled. However, RJMCMC is computationally inefficient and can add significant time to any MCMC procedure that utilizes the method.

3.1 A Bayesian Nonparametric Approach

The most promising approach to solving this problem is to utilize nonparametric priors on the number of factors and allow K to stochastically vary. Bayesian nonparametric priors such as the Dirichlet process prior (Ferguson, 1973) are frequently used to learn the number of features when each observation can belong to one feature only. These priors have strict probabilistic properties that make usage in applied statistics attractive. The problem of continuous latent variable modeling, however, is not that of clustering; since a single observation or item can be represented on more than one dimension, a more nuanced approach is required.

One fully nonparametric approach is presented by Bhattacharya et al. (2011), who propose a sparseness inducing normal-gamma process prior on K. This approach allows K to be determined in a fully stochastic manner and is estimable through full Gibbs sampling. While this approach would be appropriate in the case of only estimating a reduced dimensionality covariance matrix, the fact that it does not estimate identifiable structural parameters makes this method unattractive for measurement models in the social sciences. The approach by Bhattacharya et al. (2011) does not invoke sparseness of the same set of factors each time the model is run; though K is consistently estimated, the location of the K dimensions is completely determined by the starting values of the corresponding MCMC estimation method. While this is not problematic for reduced dimensionality covariance estimation, theory driven measurement models rely on structural parameters for theory testing.

An alternative approach is presented by Kim et al. (2018), who count dimensionality in a large sample. joint text and roll call space by utilizing a regularization approach. Similar to Bayesian nonparametric approaches, regularization produces a sparse decomposition of the observed data and measures the number of dimensions as those which have nonzero elements. While this approach is computationally efficient, regularization prevents the ability to get a full posterior distribution for the structural parameters of the latent variable model. If the dimensionality related to specific items or individual latent variables is of interest, then the full posterior is desirable. For very large data sets, regularization provides a reasonable approach. However, sets of roll call votes are generally a reasonable size. Thus, an approach that fully characterizes posterior distributions is desirable.

A purely Bayesian nonparametric approach which allows us to probabilistically model the number of necessary factors in the factor model utilizes the beta process (Hjort, 1990). A beta process is a Levy process which can be defined as follows:

Definition. *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:*

$$H(B_k) \sim Beta(\alpha H_0(B_k), \alpha(1 - H_0(B_k))) \tag{3.1}$$

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:

$$H(\nu) = \sum_{k=1}^{\infty} \pi_k \delta_{\nu,k}(\nu) \tag{3.2}$$

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

$$r_{i,k} \sim Bern(\pi_k) \tag{3.3}$$

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.

Sampling from an infinite beta process is difficult, but a marginalized approach exists (Paisley and Carin, 2009) that allows for a relatively simple sampling scheme. Define the Bernoulli process dictating the values of the infinite matrix, $\mathbf{R}$, as:

$$\begin{aligned} \pi_k &\sim \text{Beta}\left(\frac{a}{K}, \frac{b(K-1)}{K}\right) \\ r_{i,k} &\sim \text{Bern}(\pi_k \delta_{i,k}) \end{aligned} \tag{3.4}$$

where a and b are hyperparameters and $\delta_{i,k}$ is an associated probability measure.

A *beta process prior* can be seen as a prior on an infinite binary matrix, $\mathbf{R}$. $\mathbf{R}$ is assumed to be a $P \times \infty$ matrix where a one indicates the presence of a feature and a 0 indicates its absence. This is a *sparsity* inducing prior, meaning that most columns of $\mathbf{R}$ will be inactive, $r_j = \mathbf{0}$.

The beta process prior of this form constitutes a simple stochastic process. Integrating out π, this process is the *two parameter Indian Buffet Process* (IBP) (Ghahramani and Griffiths, 2006). This process can be imagined as follows (Thibaux and Jordan, 2007):

1. The first customer enters an Indian buffet with an infinite number of dishes.
2. She helps herself to the first Pois(α) dishes.
3. The $j^{th} \in (1, ..P)$ customer helps himself to each dish with probability $\frac{m_k}{\beta+P-1}$, where m_k is the number of times dish $k \in (1, .., \infty)$ was previously chosen.
4. The j^{th} customer tries Pois $\left(\frac{\alpha\beta}{\beta+j-1}\right)$ new dishes.

The two IBP parameters play different roles in this setting. α dictates how many dishes a customer tries assuming no other customers have visited the buffet. β dictates the *a priori* probability that a customer tries a given dish. IBP is *infinitely exchangeable*, so it can always be assumed that the current customer is the last customer. Infinite exchangeability also allows there to exist a strict duality between a beta process and the IBP by de Fenetti's theorem (Diaconis and Freedman, 1980).

IBP has two notable properties for producing sparse matrices. First, the number of dishes tried by a single customer is Pois(α). However, over the entire set of P customers, the number of dishes sampled is Pois $\left(\sum_{j=1}^{P} \frac{\alpha\beta}{\beta+j-1}\right) \approx \mathcal{O}(\log(P))$. As $P \to \infty$, the number of dishes sampled also grows to infinity. This shows that as P increases, the implied complexity of the model also increases. Second, IBP exhibits a "rich get richer" property; as a dish becomes more popular, the probability that it is samples in proceeding iterations increases. In turn, dishes that are not popular amongst the customers 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 dishes to dominate the feature space. This allows the IBP prior to prevent against overfitting.

A finite approximation to the beta process can be made by setting the maximum number of columns in $\mathbf{R}$ to a large, but finite, value (Doshi et al., 2009). For standard values of P, setting the maximum number of features to 100 shows good behavior. As P increases, the approximation performs more accurately. Similarly, when the number of features in the data is small, the finite approximation shows even fewer losses. One way to minimize loss due to the finite approximation is to allow any MCMC procedure to run for some number iterations with an information-free prior. In this case, this equates to setting the number of features that hold a feature to $\frac{P}{2}$. Allowing a model utilizing this prior to mix for some period of time shows improvement on selecting the correct number of dimensions.

Computational efficiency is an important part of any procedure considered and while the infinite nature of the IBP is needed to ensure posterior consistency, searching for new features is a costly algorithm (Knowles and Ghahramani, 2011). Similarly, poor choices for α and β can lead to the introduction of *spurious* features. The finite procedure does not add new features to the feature space beyond the initial over-encompassing set. Likewise, the marginalized beta process does not require these hyperparameters when computing the posterior probabilities for existing features. Therefore, α and β are not needed in the finite specification. Given the increase in computational efficiency and the relatively similar performance between the infinite and finite specifications, the finite specification is explored in this paper. For a more thorough discussion of the infinite specification and applications to latent feature models, see Knowles and Ghahramani (2011) and Doshi et al. (2009).

4 A Beta Process IRT Model

Utilizing the theory above, a scaling procedure for binary outcomes utilizing a sparsity inducing beta process prior can be made. Let $\mathbf{X}$ be a latent mapping of the observed manifest data $\mathbf{Y}$ achieved via data augmentation such that:

$$x_{i,j} \sim \begin{cases} \mathcal{TN}_{-\infty,0}(\lambda_j\omega_i - \alpha_j, 1) \text{ if } y_{i,j} = 0 \\ \mathcal{TN}_{0,\infty}(\lambda_j\omega_i - \alpha_j, 1) \text{ if } y_{i,j} = 1 \\ \mathcal{N}(\lambda_j\omega_i - \alpha_j, 1) \text{ if } y_{i,j} \text{ is missing} \end{cases} \tag{4.1}$$

then we can define a beta process IRT model as (BPIRT) as:

$$x_{i,j} = (r_j \odot \lambda_j)\omega_i - \alpha_j + \epsilon_{i,j} \tag{4.2}$$

where $\mathbf{R}$ is a $P \times K$ binary matrix. $\odot$ is Hadamard multiplication, which is equivalent to elementwise multiplication.

This specification induces a *spike and slab* prior on the matrix of loadings, $\mathbf{\Lambda}$. Using the beta process notation, the induced prior of $\lambda_{j,k}$ is:

$$P(\lambda_{j,k}|r_{j,k}, \gamma_k) \sim r_{j,k}\, \mathcal{N}(0, \gamma_k) + (1 - r_{j,k})\delta_0 \tag{4.3}$$

where δ_0 is a point mass PDF at zero (Dirac, 1981). Thus, when $r_{j,k} = 1$, $\lambda_{j,k}$ takes on a non-zero value. This prior promotes *sparsity* in the loadings matrix by allowing elements of Λ to take non-zero values if and only if a non-zero value adds something over fixing the value at zero.

This construction allows us to define a full model:

$$\begin{aligned} &P(x_{i,j}|-) \sim \mathcal{N}_p((r_j \odot \lambda_j)\omega_i - \alpha_j, 1) \\ &P(\omega_i) \sim \mathcal{N}_K(0, I_K) \\ &P(\lambda_{j,k}|r_{j,k}) \sim r_{j,k}\mathcal{N}_p(0, \gamma_{j,k}^{-1}) + (1 - r_{j,k})\delta_0 \\ &P(r_{j,k}) \sim \text{Bern}(\pi_k) \\ &P(\pi_k) \sim \text{Beta}(a/K, b(K-1)/K) \\ &P(\gamma_{j,k}) \sim \text{Gamma}(c, d) \\ &P(d) \sim \text{Gamma}(c_0, d_0) \end{aligned} \tag{4.4}$$

where $i \in (1, ..., N)$, $j \in (1, ..., P)$, and $k \in (1, ..., K)$. a, b, c, d, c_0, d_0 are prior hyperparameters. In this construction, normal priors are assumed on the factor loadings and Gamma priors are assumed on the precisions for the various structural parameters.

This model has many of the same features as the standard factor analysis model. The manifest variables are decomposed into the loadings matrix, $\mathbf{\Lambda}$, and the latent variables, $\mathbf{\Omega}$. The dimensions of the latent variables are assumed to be orthogonal, *a priori*. Marginally, $P(x_i - \alpha) \sim \mathcal{N}_K(0, \mathbf{\Lambda}'\mathbf{\Lambda} + I_K)$. The new additions, however, provide interesting properties for the latent variable estimation. First, the addition of the infinite binary matrix, $\mathbf{R}$, allows us to learn about the true number of latent dimensions within the manifest data. $\mathbf{R}$ represents whether feature $k \in (1, ..., K)$ is a meaningful summary of the data. Similarly, it provides a computational advantage in that many of the computations needed to estimate Λ are not needed at each step; when $r_{j,k} = 0$, there is no need to estimate $\lambda_{j,k}$. All in all, the beta process IRT approach provides a computationally efficient approach to the dimensionality problem that also provides interpretable estimates of the structural parameters.

Under this specification, estimation of the model using Gibbs sampling proceeds in a relatively straightforward manner with sampling steps are outlined in Appendix A. Setting good initial values can greatly increase the overall performance of this procedure. In particular, good starting values for $\mathbf{\Lambda}$ and $\mathbf{\Omega}$ can significantly speed up the convergence of the MCMC chains. In order to do this, begin by setting the initial values of the latent copula random variables, X, in accordance with their ordering. On this set of initial values, perform singular value decomposition to get good starting values for Λ and Ω. K is initially set to a value that is believed to be much larger than the true dimensionality of the latent space, typically around 100. The variance terms and other structural parameters are all initialized at random.

4.1 Posterior Inference

A major problem in the standard latent variable specification is that the estimates for the structural parameters are not uniquely identified without further constraints. We can obtain an identical $\mathbf{\Omega}$ by multiplying $\mathbf{\Lambda}$ by an orthonormal matrix, $\mathbf{M}$, such that $\mathbf{M}\mathbf{M}' = \mathcal{I}$. Following a common convention to ensure identifiability, many implementations of Bayesian factor analysis assume that Λ has a full-rank lower triangular structure with positive elements on the diagonal (Geweke and Zhou, 1996). The spirit of this recommendation relies on the notion that the researcher can effectively place structural zeroes in the loadings matrix in accordance with his theory of the latent space in mind. However, this is rarely achievable as the theory behind a latent space is difficult to put in terms of the loadings matrix. Similarly, the resulting prior on $\mathbf{\Lambda}$ is no longer *exchangeable*. When these constraints are placed in an ad-hoc manner, they can lead to significant dependencies and multimodalities in the resulting posterior. Note that similar constraints can be placed on the matrix of latent variables, $\mathbf{\Omega}$ (Clinton et al., 2004).

BPIRT avoids the issue of rotational invariance through the inducement of spike and slab priors on the factor loadings. In short, the spike and slab priors on the factor loadings prevents their respective posteriors from having both positive and negative support. A more in-depth rationale of this reasoning is provided by Bhattacharya et al. (2011).

One of the key inferential questions answered by this model pertains to the number of orthogonal factors required to best summarize the manifest data. The beta process prior and corresponding infinite binary matrix provides a method for making this inference; once the Markov Chains have converged to the stationary distributions, the distribution of the number of factors can be sampled from Monte Carlo draws. Under the finite specification used in this paper, the distribution of the number of factors converges to a single positive integer value. Theoretically, this can be seen as a conservative, lower-bound on the number of dimensions needed to best describe the manifest set.

Inference on the structural parameters of BPIRT can be done as normal. The posterior distribution of the factor scores, $\mathbf{\Omega}$, provide a measure of the latent scores for each observation. Over the set of observations, they describe a projection of the manifest set on the latent space. Over the collection of latent factors, each factor is characterized by the factor loadings, $\mathbf{\Lambda}$, which load highly on them; this can be used to determine which covariates influence each latent dimension. Similarly, $\mathbf{R}$ can be used to see which items have a non-zero contribution to each dimension. This is in direct contrast to standard factor analysis procedures which require that all questions explicitly load on all dimensions.

5 Simulation Results

To assess the quality of BPIRT, estimates of the structural parameters are compared against known quantities using a synthetically created data set. For the first known data set, $n = 1000$, $p = 500$, and $k = 10$. The data set is generated with known $\mathbf{\Lambda}$, $\mathbf{\Omega}$, and idiosyncratic error variances. $\mathbf{\Lambda}$ is also designed with specific structural zeroes, thus a known $\mathbf{R}$ is also simulated. 1000 thinned Monte Carlo draws are taken from the posterior after a burn-in of 5000 iterations over 4 chains. Standard convergence tests were used and showed no convergence issues.

A first check is to ensure that the model appropriately estimates the number of latent dimensions. Figure 1 shows number of dimensions estimated by the model during each iteration of the MCMC procedure over the 4 chains. For each of the 4 chains, the MCMC chain converges to the correct value of 10. For this simulation, it only took around 200 iterations for the chain to converge to this value. Thus, we can see that BPIRT converges to the appropriate number of dimensions given sufficient N and P.

Since the number of dimensions is correctly identified, $\mathbf{R}$ can also be examined to determine how accurately the model discovers the structural zeroes. With relation to $\mathbf{\Lambda}$, structural zeroes correspond to covariates that contribute only noise variance on a specific latent dimension. Figure 2 shows the true $\mathbf{R}$ against the median of the posterior draws for $\mathbf{R}$. In this case, $\mathbf{R}$ was recovered with approximately 94% accuracy. This performance is quite good and shows that BPIRT also does a great job at locating structural zeroes in the loadings matrix.

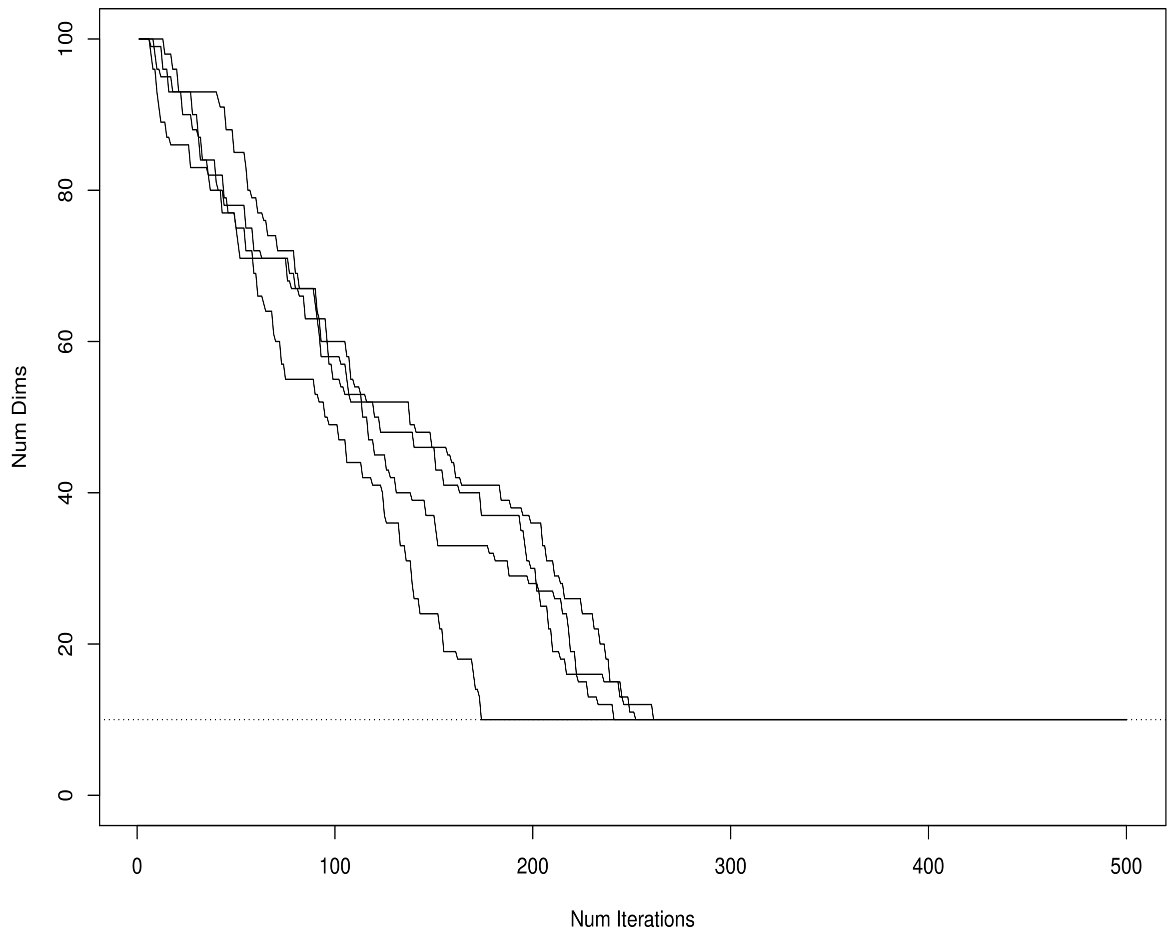

Figure 1: Number of Dimensions over Iterations

An important comparison to make is to mixed factor analysis (Quinn, 2004), a standard latent variable model for mixed margins. The two methods should perform similarly on the simulated data. Given that mixed factor analysis doesn't have a build in test for correct dimensionality, the methods are compared using proportional error and proportional reduction in error [1] Figure 3 shows both measures for both models. The models perform similarly on these metrics, as they should. However, the novelty of the IBP priors can be seen most meaningfully in the proportional reduction in error plot. When attempting to select a dimensionality using a Scree-like procedure, the goal is to look for the proportional reduction

[1]Define the baseline predictions for the model as the error associated with using the mean response for continuous margins and the modal response for discrete margins (i.e. $k = 0$). For models with positive dimensionality (i.e. $k \geq 1$), define the predicted values given point estimators for Λ and Ω as $\dot{Y} = \dot{\Lambda}\dot{\Omega}$. Define proportional error for a k-dimensional model as $PE(k) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{p} |y_{i,j} - \dot{y}_{i,j}|}{\sum_{i=1}^{n}\sum_{j=1}^{p} y_{i,j}}$. Define proportion reduction in error for a k-dimensional model as $PRE(k) = PE(k) - PE(k-1)$.

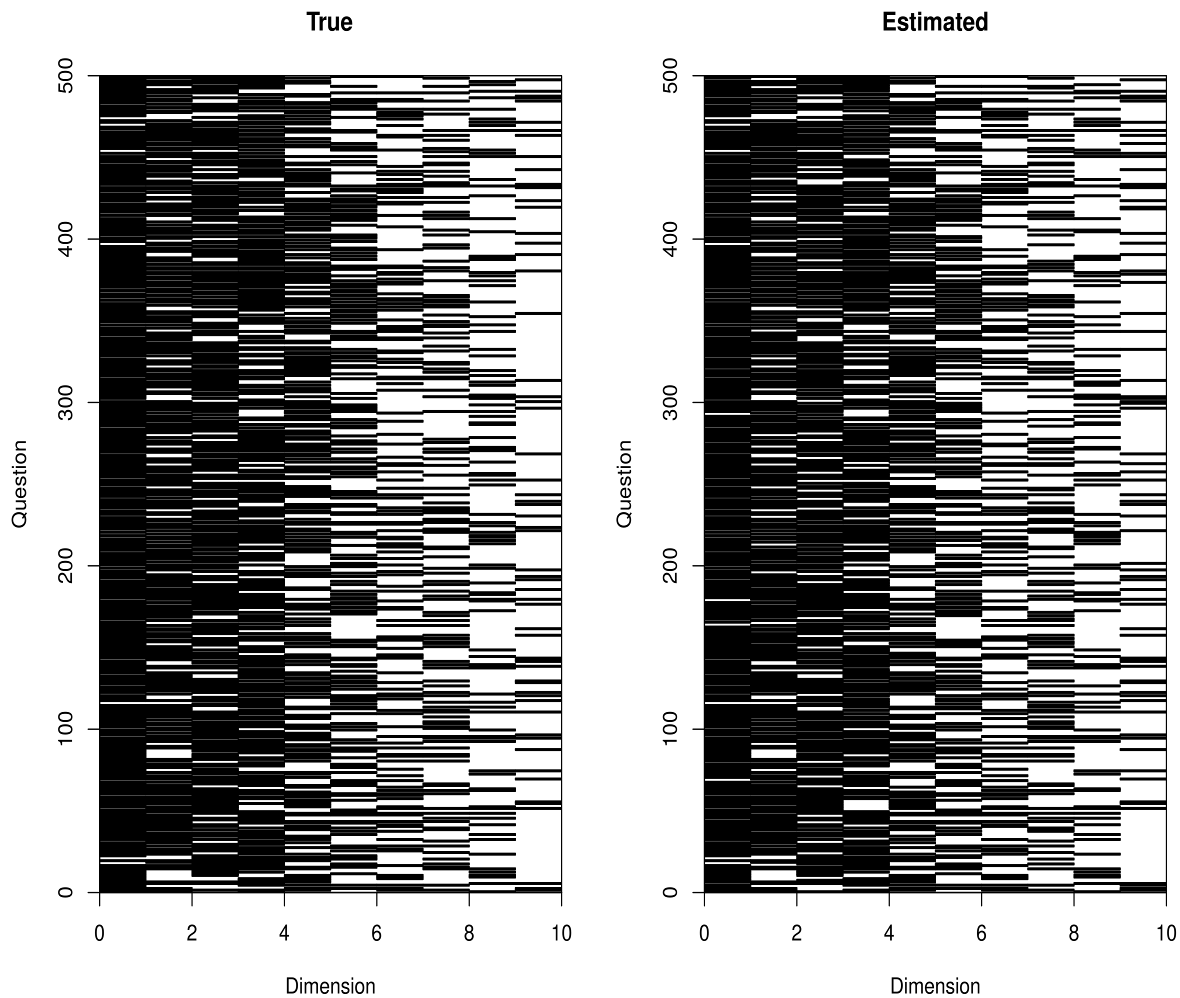

Figure 2: True $\mathbf{R}$ vs. Estimated $\mathbf{R}$

in error which the return from adding a new dimension gives "diminishing returns". In this simulation, a researcher could reasonably argue to keep between 5 and 8 dimensions. Similarly, if the researcher was only running models until she saw a plateau in the plot, she might stop at 5. Given that this data is known to have 10 orthogonal dimensions, the Scree procedure is likely to cause underestimates of the true dimensionality of the space. This shows the importance of using a statistical procedure like the IBP prior when making a decision of this sort.

One point that deserves further exploration is the effect of N and P on the ability of BPIRT to estimate the correct number of dimensions in a model. Recall that the Indian Buffet Process allows the *a priori* complexity of the model to grow as P increases; the expected number of features in the prior is approximately $\log(P)$. N also influences the implied complexity of the model by providing more information for each estimate of $\lambda_{j,k}$. In order to see the extent to which N and P influence the model's estimated dimensionality, the same simulations are run with varying N and P. The results from these simulations

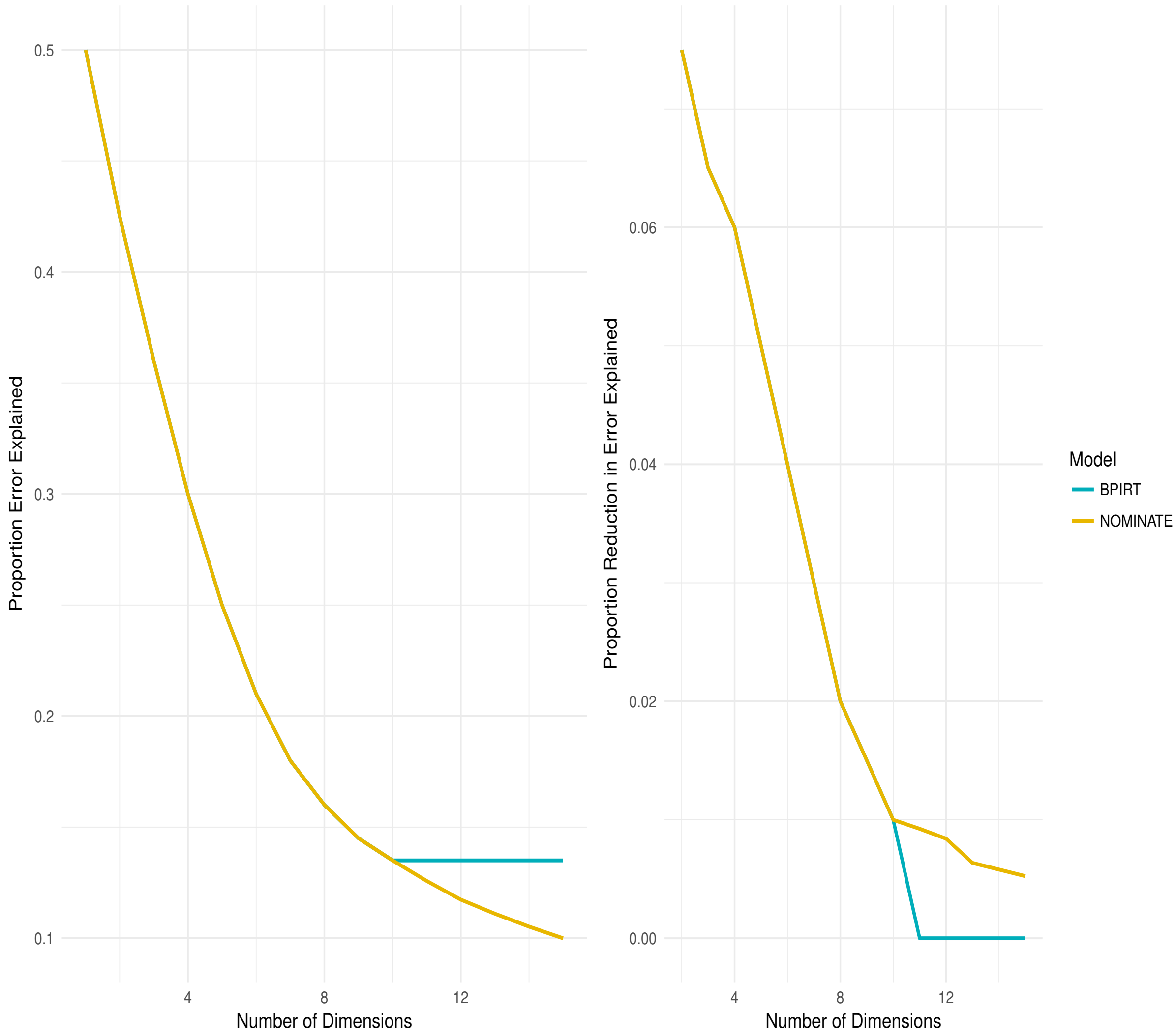

Figure 3: Proportion Error and Proportional Reduction in Error for BPCFA vs. Mixed Factor Analysis

	P = 100	P = 200	P = 300	P = 400	P = 500
N = 100	4	5	6	6	6
N = 200	6	7	8	8	8
N = 300	6	7	8	8	8
N = 400	6	8	10	10	10
N = 500	7	9	10	10	10
N = 1000	9	10	10	10	10

Table 1: Number of Dimensions Recovered from Simulated Data with 10 Orthogonal Dimensions

can be seen in Table 1. As P increases, the dimensionality approaches a plateau associated with specific N, which makes sense given that the expected number of prior dimensions is a function of P. On the other hand, as N increases, the dimensionality of the model is allowed

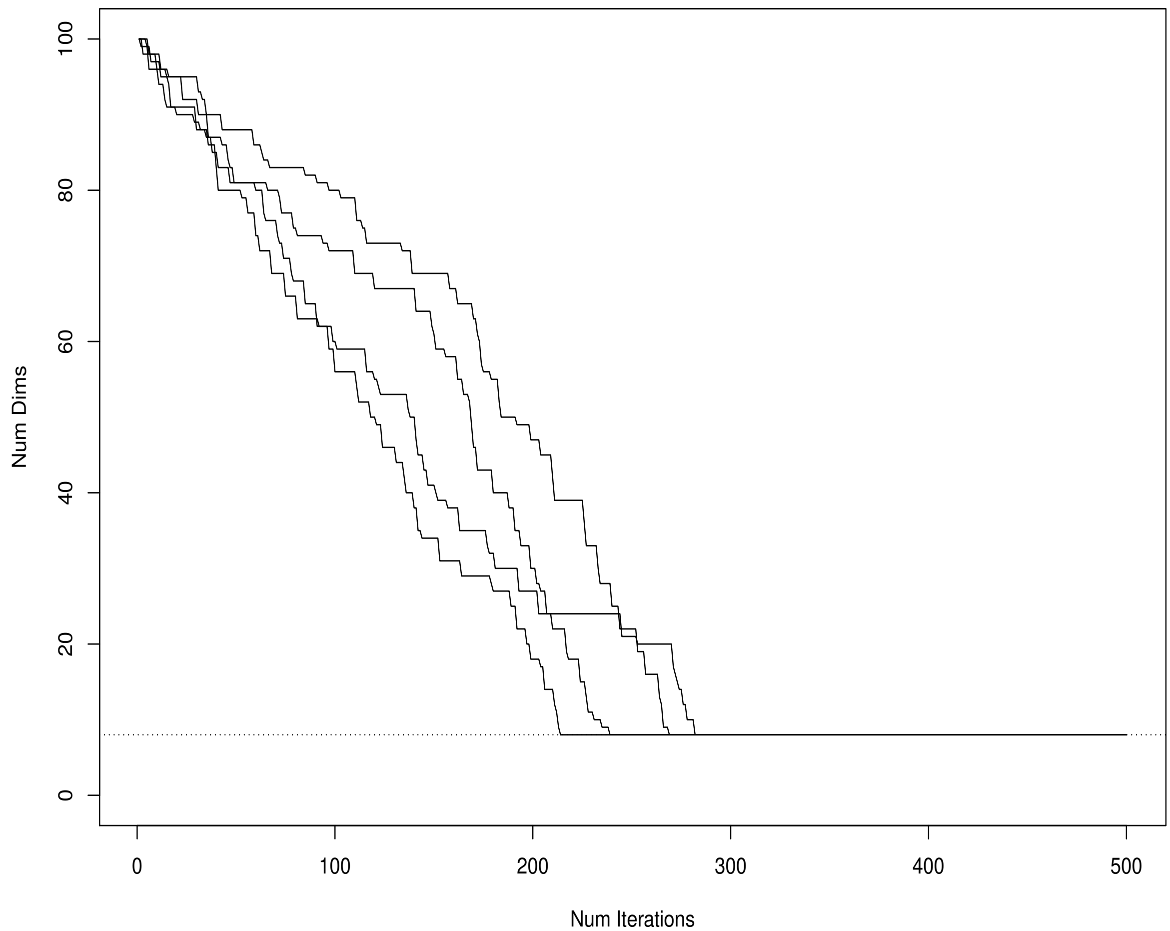

Figure 4: Convergence to Eight Dimensions over Iterations

to increase and approaches the true value. However, this value never exceeds to the true value. Thus, large N and P ensure that the true dimensionality of the model is uncovered. On the other hand, when N and P are not sufficiently large, BPIRT is conservative and returns a dimensionality lower than the truth.

6 Application to Congressional Roll Call Votes

To examine these theories and show the benefits of the BPIRT model, roll calls from the 1^{st}-114^{th} sessions of the U.S. House are examined and the latent policy spaces are estimated for each individual session, independently. For each session, four chains were run with a burn-in of 5000 draws. 250 draws were collected from each chain thinned to take every 10^{th} draw.

Convergence was assessed using the diagnostics presented by the `R` package `superdiag` (Tsai et al., 2012). There was no evidence indicating problems of convergence in these models.

6.1 The 107^{th} U.S. House

To demonstrate the power of this model, the 107^{th} of the U.S. House is examined, in depth. The 107^{th} session took place from January 3^{rd}, 2001 to January 3^{rd}, 2003. The 107^{th} Congress was characterized by the September 11^{th} terrorist attacks and subsequent changes to transportation, immigration, and defense policies. In response to the attacks, Congress voted to allow the president to invade Iraq, which was an issue that created friction within both major parties. Over the course of the session, Republicans had a slim majority in the House while the Senate was changed leadership three times. This lack of stability created interesting party dynamics and, in turn, led to interesting returns in roll call votes.

Running BPIRT for the 107^{th} House yields a policy space with eight dimensions. Figure 4 shows the number of dimensions estimated by the model as a function of iteration. As with the simulation, the model settles in on a value for the dimensionality in around 300 iterations. In order to get an idea of what each of the dimensions corresponds to, bill summaries and titles were scraped from `voteview.com` (Poole and Rosenthal, 2012). For a given dimension, K, the summary of any bill, j, where $P(r_{j,k} = 1) > .95$ was Tf-Idf weighted over the set of included bills and most important words were discovered. These words can be seen in Figure 5.

Figure 5 shows dimension names and associated important words. The most important dimension is the liberal-conservative dimension, which heavily corresponds to the first dimension in NOMINATE scores (correlation = .98). The other seven dimensions correspond to different issue areas or bill types. General issues emerge as important dimensions in the roll call record, like rural issues, social issues, trade, domestic spending, and defense spending. On the other hand, this algorithm was able to go through and pick out a set of procedural votes, which include votes on approving the Journal at the end of every meeting, celebrating the 50^{th} anniversary of the constitution of Puerto Rico, celebrating the completion of a new railroad system, etc. The final dimensions is specific to the Congress, as it corresponds to bills that were related to the 9/11 terrorist attacks. These bills discussed border control, counterterrorism efforts, FAA reform, and numerous other topics related to changes after the attacks. Along with the procedural dimension, these two dimensions provide great evidence that BPIRT is picking theoretically important and distinct issue dimensions in the common space estimation procedure.

Figure 6 shows plots of the mean estimate of legislators' locations in the latent spaces for four of the issue dimensions plotted against their corresponding liberal-conservative scores. These plots show one of the big advantages of using the BPIRT procedure over NOMINATE methodology - the ability to see within-party dynamics. For example, looking at the rural issues dimension, there is little correlation between party and position on the rural dimension. Looking deeper in the actual data, it is easy to see that legislators that scores highly on the

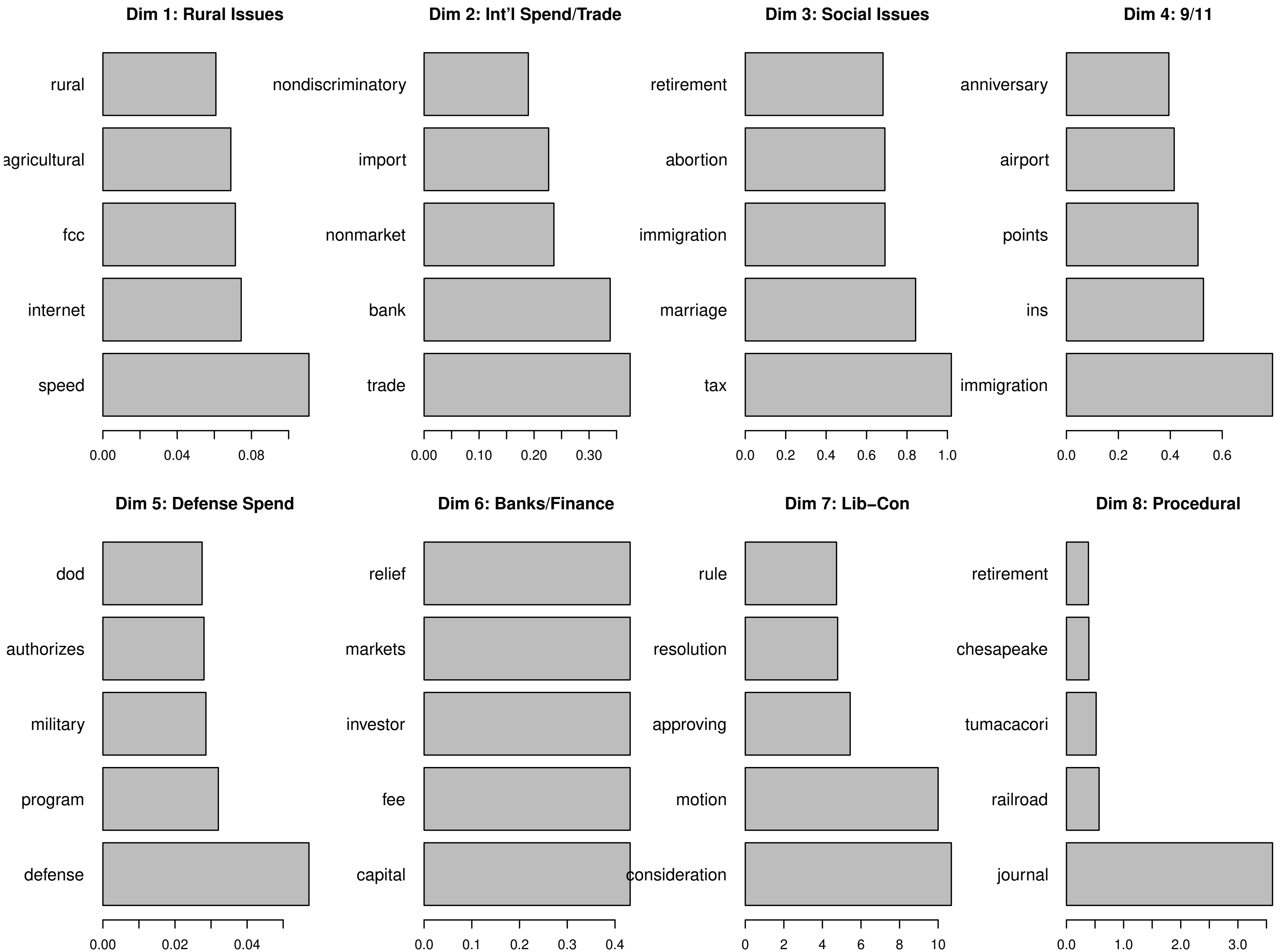

Figure 5: Dimension Names and Important Words for 107^{th} House

rural dimension come from rural districts in Mississippi, Texas, New York, etc. Given that the majority of bills that fall into this category for this Congress relate to infrastructure, affordable Internet access, and agriculture, it is no surprise that legislators who are based in rural areas vote differently than their urban-based counterparts. A similar, non-partisan dynamic can be seen on the procedural dimension, with most legislators voting "yes" on almost all bills in the dimension. However, there are a set of legislators that are not in line with the rest of the House. While procedural votes are rarely not passed, legislators use these votes as "protest" votes to show lack of faith in the parties. The legislators that load highly on this dimension were vocal with their displeasure with the party at some point in the session, so this variation makes sense.

Perhaps more interesting, the policy scores for social issues show different spreads across the two parties. Looking at the plots for social issues and defense spending, there is a noticeable difference between the spreads of the latent scores for Democrats and Republicans; Democrats have much more spread in their scores than the Republicans. This can be attributed to the power of the party as a coalition, or how frequently the party members vote together on a specific set of issues. Figure 7 shows the proportion of members from each party voting together on each bill that has a non-zero loading ($P(r_{j,k} = 1) > .95$) on social and defense spending issues, respectively. The difference between parties is stark; on every bill in these issue areas, the Republican party was more unified than the Democratic party. This feature of differing spread is unique to BPIRT and is not seen in the corresponding NOMINATE scores, letting BPIRT answer a variety of inner-party questions that other methods cannot.

The advantages of BPIRT over NOMINATE are not only substantive in nature, the scores estimated by this method show significant increases in the ability to classify roll call votes. An important comparison to make is between the ability of the BPIRT IBP prior to select dimensionality and the PRE method used by NOMINATE. As discussed previously, BPIRT only selects dimensions in accordance to the beta-Bernoulli process while NOMINATE selects according to a Scree-like procedure using proportional reduction in error. Figure 8 shows error rates and proportional reduction in error for both BPIRT and NOMINATE up to a 15 dimensional model. On first glance, it is obvious that the two models produce estimates which are relatively similar in reduction in error. However, there are some minor, but important differences. First, the error rate curve is shallower for BPIRT, implying that there is more predictive power in higher dimensions in under the BPIRT model. Second, in the proportion reduction in error plot, it is easy to see that BPIRT stops adding dimensions when they no longer add predictive power that can be distinctly distinguished from statistical noise. This makes selecting an appropriate number of dimensions quite simple. On the other hand, looking at the curve for NOMINATE would reasonably lead one to select a one-dimensional model; after the first dimension, the increase in predictive power decreases quite a bit. However, this confuses small contributions with insignificant contributions and leads to underestimation of the dimensionality of the model.

A more fine tuned way to examine these gains is to examine individual bills and examine changes in classification ability. Figure 9 plots error rates for BPIRT against one-dimensional NOMINATE. For a number of bills, NOMINATE and BPIRT perform similarly, implying

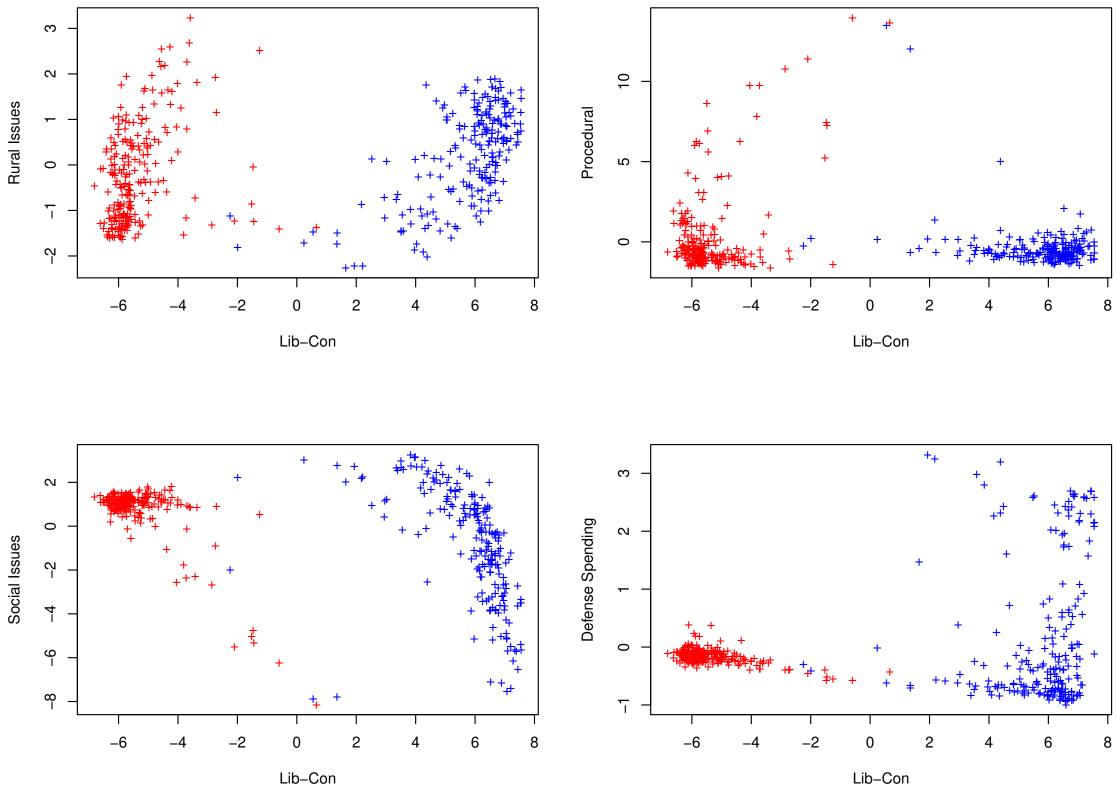

Figure 6: Policy Space Plots for 107^{th} U.S. House

	Rural	Trade	Social	9/11	Defense	Domestic	Lib-Con	Proced.	PRE Gain
V0364	1	0	0	0	0	0	0	0	0.220
V0606	1	0	0	0	0	0	0	0	0.146
V0218	1	0	0	0	0	0	0	0	0.138
V0629	1	0	0	0	0	0	0	1	0.131
V0368	1	0	1	0	0	0	0	1	0.121
V0847	0	1	0	0	0	0	0	1	0.119
V0828	0	0	1	0	0	0	0	1	0.114
V0673	0	0	0	0	0	1	0	0	0.112
V0098	0	0	0	0	0	1	0	0	0.110
V0832	0	0	0	0	0	1	0	0	0.107
V0547	0	0	0	0	0	1	0	0	0.107

Table 2: Proportion Error Gain for BPIRT over NOMINATE

that one-dimension is sufficient for explaining vote behavior for a number of bills. However, for the rest, there is non-zero contribution by the other dimensions. Table 2 shows the IBP matrix, **R**, for the twenty bills which show largest gains in classification power using BPFCA. The reason for this is easy to see - when bills do not load on the liberal-conservative dimension, the gains for BPIRT are large. This speaks directly to the theory posited by

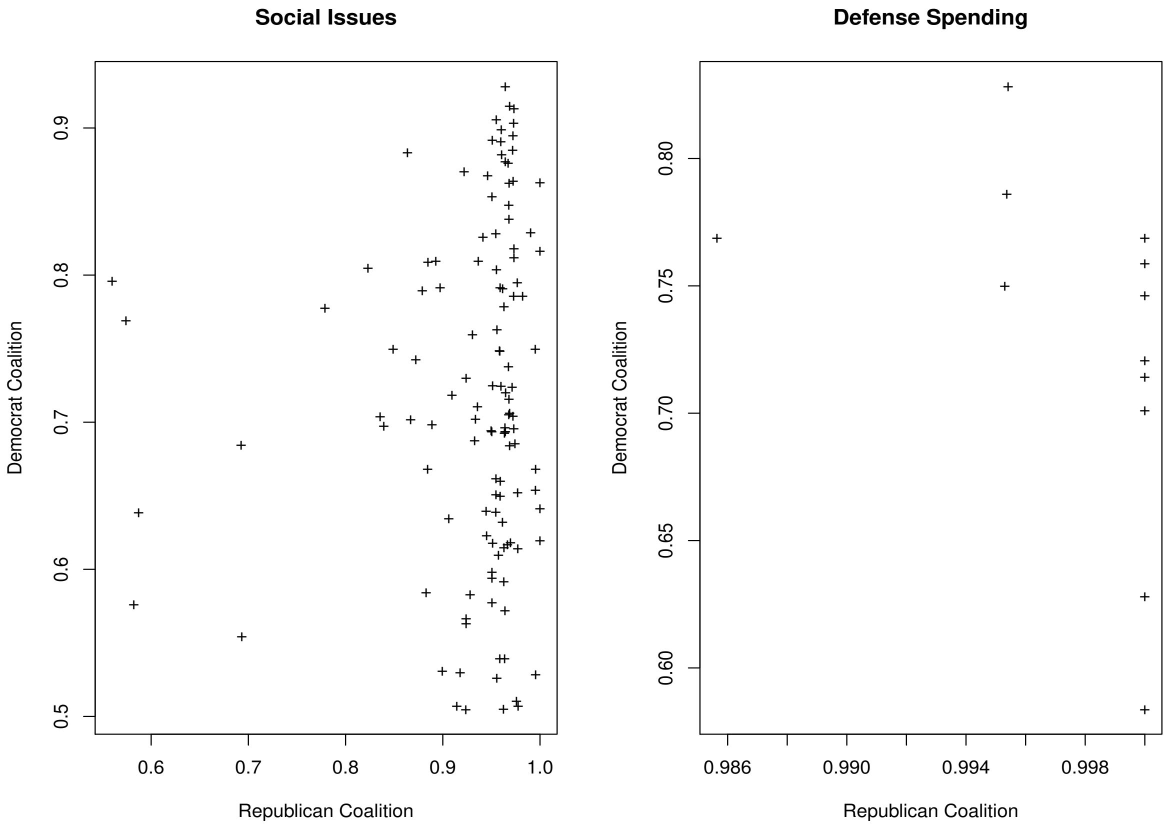

Figure 7: Republican Vote Similarity vs. Democrat Similarity for Defense and Social Dimensions in the 107^{th} U.S. House

Aldrich et al. (2014) that the low-dimensional conjecture is supported due to the requirement that all bills load on all dimensions. In turn, this results in significant increases in the quality of classification from the latent variable model. On the other hand, bills that load only on the liberal-conservative dimension perform similarly to NOMINATE, meaning that the estimates are essentially the same in low dimensions.

6.2 U.S. House over Time

Using BPIRT, the latent policy space was estimated for the U.S. House over the 1^{st} session to the 114^{th} session. Using these estimates, I examine how the dimensionality of the U.S. House has changed over time. McCarty et al. (2016) posits that polarization is increasing in the U.S. Congress over time. While the main argument made is due to parties moving further apart from one another in the latent space, they also argue that the dimensionality of Congress is decreasing over time. Even in the case of a low-dimensional policy space, the authors show that the need for higher dimensions decreases over time. As with all NOMINATE dimensionality decisions, evidence of this phenomenon is shown using proportional reduction in error as the main metric.

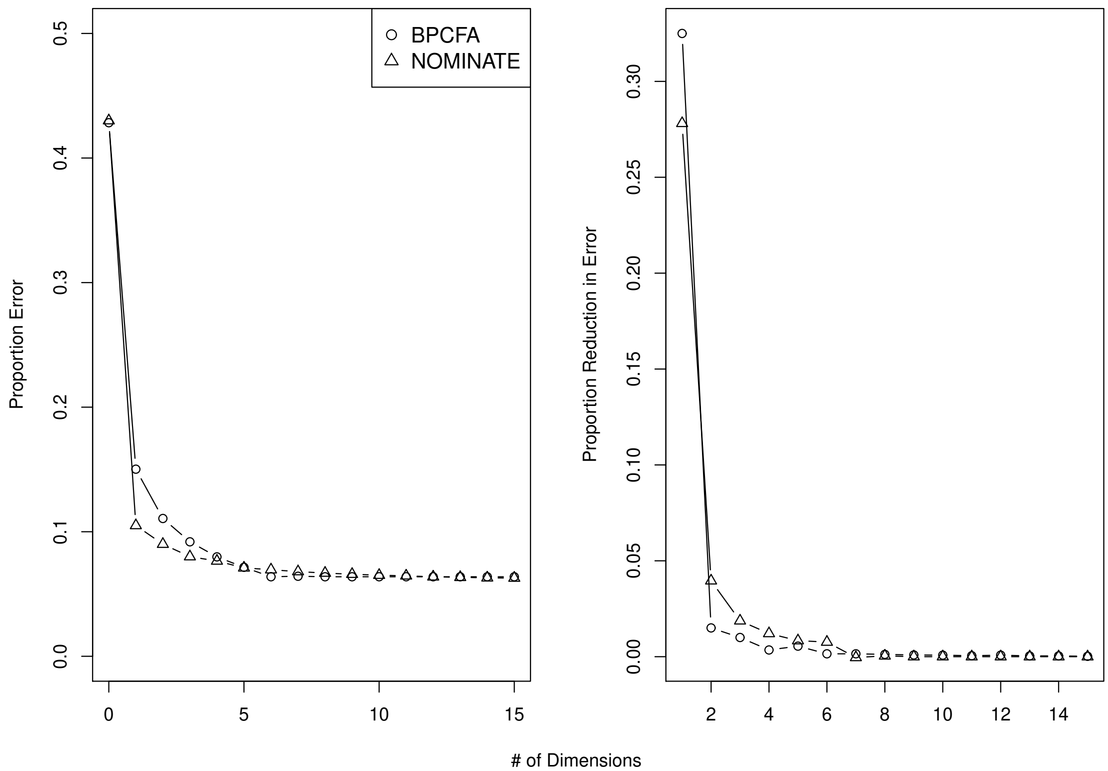

Figure 8: Proportion Error and Proportional Reduction in Error for 107th U.S. House

This notion is reexamined using BPIRT and allowing the model to dictate how many dimensions are needed to appropriately describe the latent space for each session. Figure 10 shows the estimated dimensionality of the policy space for each individual session of the U.S. House. Examining the results returned by the BPIRT method shows that there is little evidence for the low-dimensional conjecture over the course of all sessions of the U.S. House. Looking at the more recent 100^{th} - 114^{th} sessions, there appears to be a decrease in the overall dimensionality. However, this value never goes below 4 dimensions. In fact, this period of time appears to have a higher dimensionality than earlier sessions of the U.S. House. While this finding is certainly motivated by increases in the number of bills that are voted on in each session increasing over time (≈ 60 in the 58^{th} session to well over 1000 in recent sessions), there is still strong evidence that the one dimension argument is incorrect.

From the 100^{th} to the 114^{th} sessions, there are four dimensions which are shared across all sessions. In line with NOMINATE, a main liberal-conservative/party loyalty dimension is included in all spaces estimated. However, other dimensions are statistically important. In each session of the U.S. House, a procedural dimension appears which includes bills which are related to standard procedural votes. Along with this, two issue dimensions appear: defense spending and rural issues/rural spending. Given that these issues have long been debated in Congress and have caused significant friction within parties, their presence in each session is reasonable. On the other hand, a number of dimensions that were distinct in earlier sessions

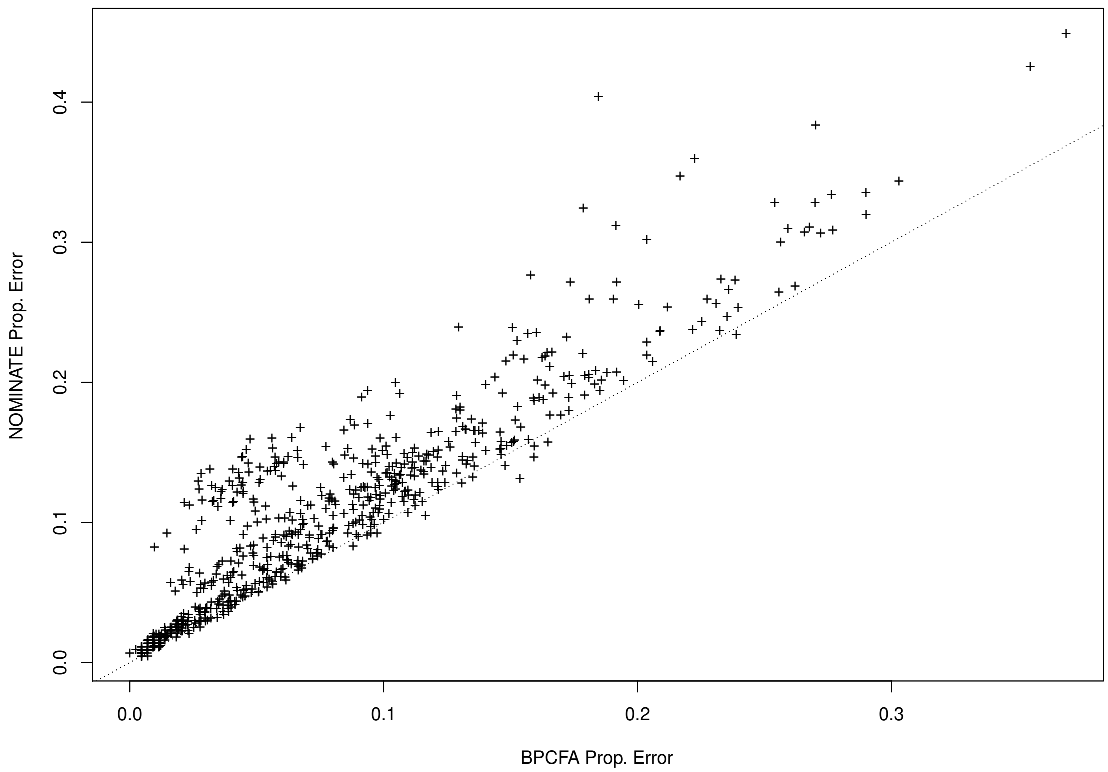

Figure 9: Proportion Errors in Classification for BPIRT vs. NOMINATE

disappear. For example, in the 100^{th} session, an important dimension related to tax policy. Over time, this dimension merges into a domestic spending dimension then collapses into the liberal-conservative dimension, indicating that the contents of the policy space are evolving over time. Though this analysis is brief, it points to a potential new usage of the BPIRT model to examine the evolution of the policy space over the course of U.S. History.

An important question to ask about the BPIRT ideal point estimates is how they compare to NOMINATE over time. A particular comparison that shows the value of added dimensions is to examine the difference in proportion of votes explained by each model. Comparison to a third case - the party only model [2] - is also useful. Figure 11 shows the proportion of votes correctly classified for each of the models. This figure provides a number of useful insights. First and foremost, this plot shows that votes in recent sessions of the U.S. House can largely be predicted using the party only model. Whereas the average session of the U.S. House shows that the party model can predict between 75% and 85% of votes, recent sessions show that the party model can predict above 90% of votes. However, this is not the only period in history where the party only model performs so well - party voting from

[2] For a given roll call vote, the party only model defines the proportion of votes that can be explained by party alone. Here, the party vote is defined as the most common vote made by members of each party for each roll call vote. Then, this vote is compared to the actual outcomes and the proportion correct is compared to BPIRT and NOMINATE - 1 Dimensional

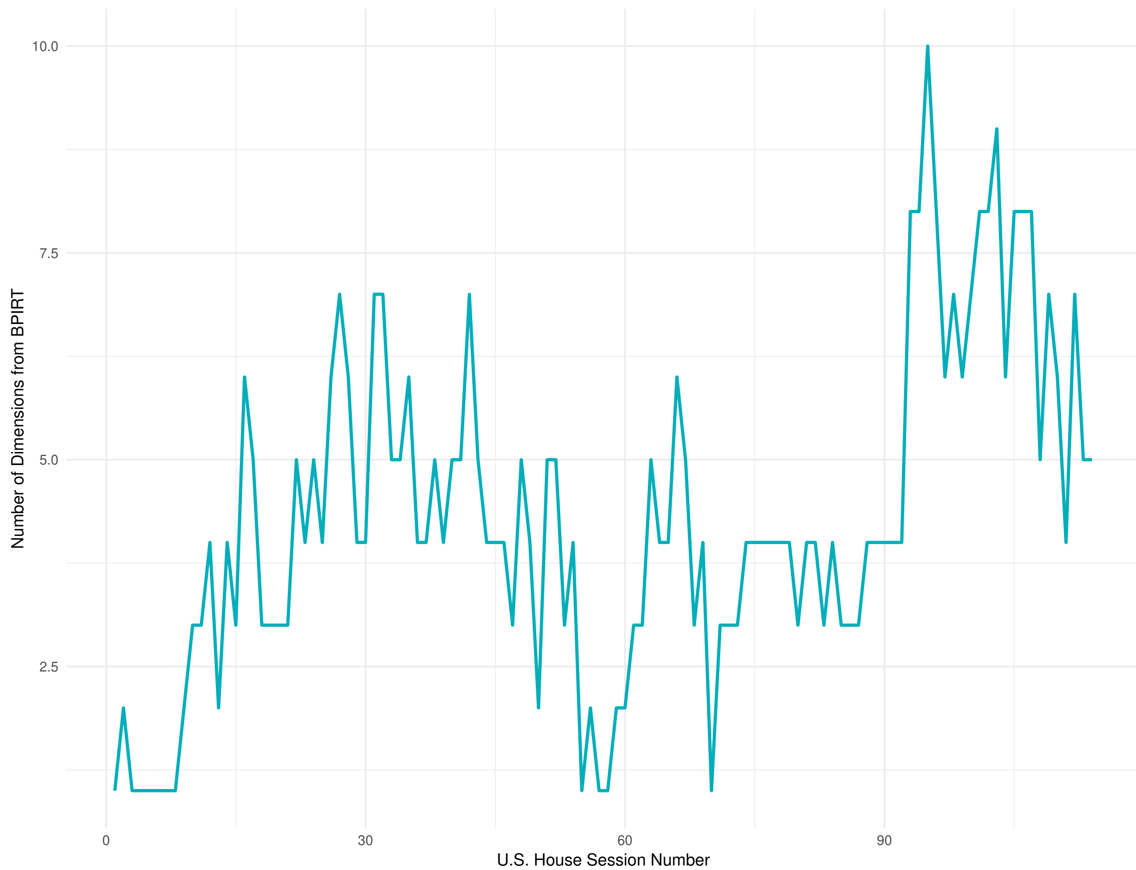

Figure 10: Dimensionality of the U.S. House over Time

the 45^{th} to the 60^{th} sessions showed a high classification ability. This finding supports the claims from McCarty et al. (2016) regarding the increase in party voting in recent times.

While the level of voting that can be explained by party is interesting, the role of roll call scaling models is to provides novel insights beyond the party only model; if knowing a member's party provides most of the information about respective votes, then there is no reason to estimate the richer metrics that the scaling procedures provide. For this reason, it is important to examine proportion classification using the party only model as a baseline. Placing the proportion of votes explained on this scale provides knowledge of what roll call scaling methods provide **beyond** party loyalty. Figure 12 shows the comparison of NOMINATE and BPIRT classification performance accounting for the strength of the party only model. [3]. This plot shows the value of added dimensions and the BPIRT procedure, overall. While it is true that the number of votes correctly classified by the party only model is increasing in recent times, the value of dimensions beyond the first is increased classification of the error beyond party loyalty - on the order of 10% - 20%, on average. Given that the value of roll call scaling techniques is the added intuition about how votes occur beyond simple party loyalty, this plot demonstrates the need to include other dimensions in the

[3]For a session of the U.S. House, let PO_h be the proportion of votes classified by the party only model in session h. Then, the party adjusted proportion votes explained is $PAPE_h = \frac{PE_h - PO_h}{1 - PO_h}$

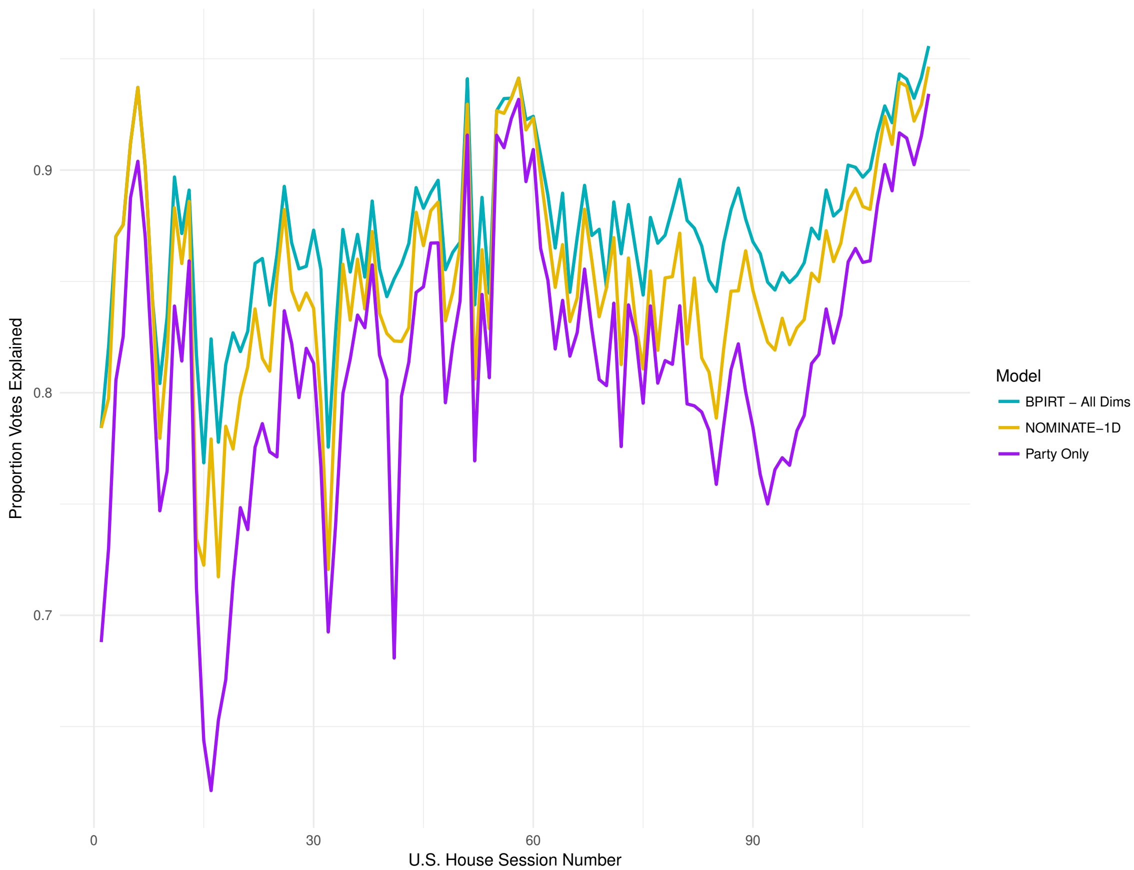

Figure 11: Proportion Votes Explained By Various Models over Time

representation of ideal points in the spatial model. This finding corroborates and extends the findings of Aldrich et al. (2014).

Note that this analysis treats each session as an independent entity and potential connections and similarities across time are not modeled. This analysis can be seen as a jumping off point for a more thorough discussion of time trends and polarization in Congress. With adjustments made to the modeling strategy to introduce dynamic preferences as introduced by Martin and Quinn (2002), BPIRT can be used to better answer questions about changes in representation by Congress over the course of U.S. history. In particular, improvement of the Indian Buffet Process priors to allow for random walks across time will provide better intuitions about the evolution of the policy space over the course of U.S. history.

7 Conclusion

The beta process IRT model is a powerful tool that estimates latent variables for measurement in the social sciences. IBP priors on the loadings matrix allow BPIRT to provide an automatic procedure for selection of dimensionality and placement of structural zeroes. This

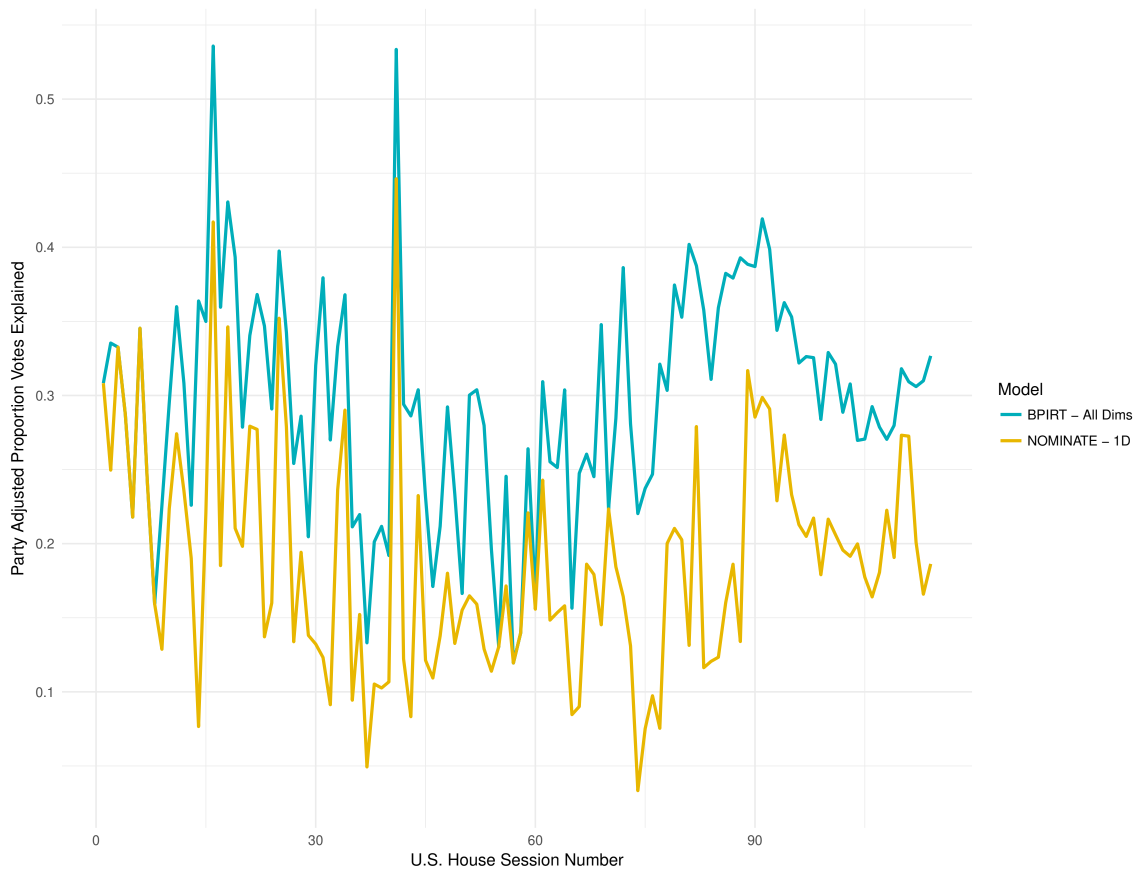

Figure 12: Proportion Party Adjusted Votes Explained By BPIRT and NOMINATE

nonparametric technique loosens a number of the assumptions that make factor analysis a difficult and overly subjective procedure. This subjectiveness is demonstrated by analyzing sessions of the U.S. House and showing that there is general tendency to *underestimate* the dimensionality of the policy space implied by roll call scaling procedures. In turn, properly modeling dimensionality can result in conclusions about voting behavior in the U.S. House that is substantively different than conclusions made using post-hoc testing procedures.

This model is ripe for extension. As mentioned previously, allowing for dynamic modeling of the latent space is a meaningful next step to allow for time trends to be appropriately considered over a set of manifest variables. There are also numerous situations where hierarchical priors would be appropriate; clustering within the matrix of ideal points would allow for rich inference about the number of groups within the policy space. In particular, allowing the number of clusters to follow a Chinese Restaurant Process prior would allow for modeling the number of groups in the latent space without making prior assumptions about what groups may exist within voting bodies. Another area where clustering is beneficial is in the loadings matrix; requiring all questions in a manifest set to have the same loadings across groups can be relaxed and this can be modeled hierarchically. Bayesian nonparametric tree priors like Kingman's coalescent provide a reasonable approach to introducing this model dynamic.

Substantively, these extensions will continue to provide ways to test theories related to the U.S. Congress in a flexible empirical framework. Unlike other roll call scaling models which have strict sets of assumptions that have strong theoretical implications for resulting inference, Bayesian nonparametric priors provide a way to minimize assumptions and provide relatively implication free estimation. In turn, these models can provide approaches that can alter how research about the U.S. Congress is done.

References

Aldrich, J. H. (1995). *Why parties?: The origin and transformation of political parties in America*. University of Chicago Press.

Aldrich, J. H. and J. S. C. Battista (2002). Conditional party government in the states. *American Journal of Political Science*, 164–172.

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

Ando, T. (2009). Bayesian factor analysis with fat-tailed factors and its exact marginal likelihood. *Journal of Multivariate Analysis 100*(8), 1717–1726.

Bafumi, J. and M. C. Herron (2010). Leapfrog representation and extremism: A study of american voters and their members in congress. *American Political Science Review 104*(3), 519–542.

Bhattacharya, A., D. B. Dunson, et al. (2011). Sparse bayesian infinite factor models. *Biometrika 98*(2), 291.

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

Cameron, C. M. (2000). *Veto bargaining: Presidents and the politics of negative power*. Cambridge University Press.

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

Clinton, J., S. Jackman, and D. Rivers (2004). The statistical analysis of roll call data. *American Political Science Review 98*(2), 355–370.

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

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

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

Diaconis, P. and D. Freedman (1980). Finite exchangeable sequences. *The Annals of Probability*, 745–764.

Dirac, P. A. M. (1981). *The principles of quantum mechanics*. Number 27. Oxford university press.

Doshi, F., K. Miller, J. V. Gael, and Y. W. Teh (2009). Variational inference for the indian buffet process. In *International Conference on Artificial Intelligence and Statistics*, pp. 137–144.

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

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

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

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

Ghosh, J. and D. B. Dunson (2009). Default prior distributions and efficient posterior computation in bayesian factor analysis. *Journal of Computational and Graphical Statistics 18*(2), 306–320.

Heckman, J. J. and J. 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.

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

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

Jenkins, J. A. (1999). Examining the bonding effects of party: A comparative analysis of roll-call voting in the us and confederate houses. *American Journal of Political Science*, 1144–1165.

Jenkins, J. A. (2000). Examining the robustness of ideological voting: evidence from the confederate house of representatives. *American Journal of Political Science*, 811–822.

Jessee, S. A. (2009). Spatial voting in the 2004 presidential election. *American Political Science Review 103*(1), 59–81.

Jessee, S. A. (2010). Partisan bias, political information and spatial voting in the 2008 presidential election. *The Journal of Politics 72*(2), 327–340.

Kim, I. S., J. Londregan, and M. Ratkovic (2018). Estimating spatial preferences from votes and text. *Political Analysis 26*(2), 210–229.

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

Kramer, G. H. (1973). On a class of equilibrium conditions for majority rule. *Econometrica: Journal of the Econometric Society*, 285–297.

Krehbiel, K. (1992). *Information and legislative organization*. University of Michigan Press.

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

Lee, S.-Y. and X.-Y. Song (2002). Bayesian selection on the number of factors in a factor analysis model. *Behaviormetrika 29*(1), 23–39.

Lopes, H. F. and M. West (2004). Bayesian model assessment in factor analysis. *Statistica Sinica*, 41–67.

Martin, A. D. and K. M. Quinn (2002). Dynamic ideal point estimation via markov chain monte carlo for the us supreme court, 1953–1999. *Political Analysis 10*(2), 134–153.

McCarty, N., K. T. Poole, and H. Rosenthal (2016). *Polarized America: The dance of ideology and unequal riches*. mit Press.

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

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

Poole, K. T. and H. Rosenthal (1984). The polarization of american politics. *The Journal of Politics 46*(4), 1061–1079.

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

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

Quinn, K. M. (2004). Bayesian factor analysis for mixed ordinal and continuous responses. *Political Analysis*, 338–353.

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

Shepsle, K. A. and B. R. Weingast (1986). Institutional foundations of committee power. Technical report, mimeo, Washington University, St. Louis.

Snyder Jr, J. M. (1992). Committee power, structure-induced equilibria, and roll call votes. *American Journal of Political Science*, 1–30.

Thibaux, R. and M. I. Jordan (2007). Hierarchical beta processes and the indian buffet process. In *International conference on artificial intelligence and statistics*, pp. 564–571.

Tsai, T.-h., J. Gill, and J. Ripkin (2012). *superdiag: R Code for Testing Markov Chain Nonconvergence*. R package version 1.1.

Welch, S. and E. H. Carlson (1973). The impact of party on voting behavior in a nonpartisan legislature. *American Political Science Review 67*(3), 854–867.

Wright, G. C. and B. F. Schaffner (2002). The influence of party: Evidence from the state legislatures. *American Political Science Review 96*(2), 367–379.

A Estimation of the BPIRT Model

Estimation of the BPIRT model uses the following Gibbs sampling routine:

1. **Sample the latent variable,** X**.** For each $i \in (1, ..., n)$ and $j \in (1, ..., p)$, sample $x_{i,j}$ from a truncated normal distribution according to:

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

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

$$\begin{aligned} t_{j,k} &= \frac{P(r_{j,k} = 1 | Y, -)}{P(r_{j,k} = 0 | Y, -)} \\ &= \frac{P(Y | r_{j,k} = 1, -)}{P(Y | r_{j,k} = 0, -)} \frac{P(r_{j,k} = 1)}{P(r_{j,k} = 0)} \end{aligned} \tag{A.2}$$

$$\frac{P(Y | r_{j,k} = 1, -)}{P(Y | r_{j,k} = 0, -)} = \sqrt{\frac{\gamma_k}{\gamma}} \exp\left(\frac{1}{2} \gamma \mu^2\right) \tag{A.3}$$

$$\frac{P(r_{j,k} = 1)}{P(r_{j,k} = 0)} = \frac{m_{-j,k}}{p - m_{-j,k} + 1} \tag{A.4}$$

where $\gamma = \omega_k' \omega_k + \gamma_k$, $\mu = \frac{1}{\gamma} \omega_k' \hat{E}_j$, $\hat{E}_j = x_j - \lambda_j \Omega + \alpha_j$ 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 $P(r_{j,k}|-) \sim \text{Bern}(p_{r=1})$. If $r_{j,k} = 1$, then sample $P(\lambda_{j,k}|-) \sim \mathcal{N}(\mu, \gamma^{-1})$. Otherwise, set $\lambda_{j,k} = 0$.

3. **Remove Inactive Features and Normalize Λ.** 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 Λ to normalize the variance. For each $j \in (1, ..., p)$ and $k \in (1, ..., K^+)$ set $\lambda_{j,k}$:

$$\lambda_{j,k} = \frac{\lambda_{j,k}}{\sqrt{1 + \sum_{h=1}^{K^+} \lambda_{j,h}^2}} \tag{A.5}$$

4. **Sample Ω.** For each $i \in (1, ..., n)$, sample ω_i from:

$$P(\omega_i|-) \sim \mathcal{N}_{K^+}(\Lambda'\Lambda + I_{K^+})^{-1}\Lambda' y_i, (\Lambda'\Lambda + I_{K^+})^{-1}) \tag{A.6}$$

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

$$P(\alpha_j|-) \sim \mathcal{N}\left(\lambda_j \Omega - x_j, \frac{1}{n}\right) \tag{A.7}$$

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

$$P(\gamma_k|-) \sim \text{Gamma}\left(c + \frac{m_k}{2}, d + \sum_{j=1}^{p} \lambda_{j,k}^2\right) \tag{A.8}$$

where m_k is the number of sources for which feature K is active.

7. **Sample d.** Sample d from:

$$P(d|-) \sim \text{Gamma}\left(c_0 + cK^+, d_0 + \sum_{k=1}^{K^+} \gamma_k\right) \tag{A.9}$$