The Ties that Bind: A Grouped Approach to Ideal Point Estimation

Kevin McAlister

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. Current scaling techniques ignore the role of group-level dependencies within the data. Assumptions about independence of observations in the scaling model ignore the possibility that members of the voting body have shared incentives to vote as a group. In turn, this leads to potential biases in the estimated values of the ideal points and underestimation of the number of dimensions needed to model the ideal point space. In this paper, I propose a new ideal point model that explicitly allows for group contributions in the underlying spatial model of voting. I derive a corresponding empirical model that utilizes flexible Bayesian nonparametric priors to estimate group effects in ideal points and the corresponding dimensionality of the ideal points. I apply this model to the 114^{th} U.S. House and show how group dynamics can be uncovered using only a set of roll call votes. This model provides insights into open questions related to group dynamics in legislative voting and has important implications for literature that utilizes ideal point estimates.

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.

Along with concerns about the dimensionality of the policy space, there has been recent

research that examines the notion of independence that is required for roll call scaling models to make estimates. In particular, there are questions about the role of groups within the data. A base assumption that must be made for procedures like NOMINATE is that the errors associated with each individual vote are independent and identically distributed conditional on the ideal point. While roll call scaling techniques are attempting to decompose dependence among the votes to find ideal points, dependence among errors is still a potential problem. This problem is discussed in depth by Spirling and Quinn (2010), showing that there is cause to be concerned about correlated errors in parliamentary voting in the U.K. This problem is mostly due to party incentives and can cause dependence among the votes that is not accounted for in the standard spatial model. Spirling and Quinn (2010) utilize an infinite mixture model approach which models the policy space as a number of latent clusters, but forgo the standard ideal point interpretation which is often used in studies of U.S. legislatures.

Continuous measurements of ideal points are key to numerous studies of the U.S. Congress, particularly those related to party control and party effects in the legislative chambers. Following Krehbiel (1992) and Krehbiel (2010), understanding the effects of party membership and the role that parties play in voting became an important endeavor in legislative studies. Through careful qualitative accounts and empirical studies, theories of party control emerged and posited that members of Congress have incentives to vote with their party and the groups assess situations where strong control over members is needed (Aldrich and Rohde, 2000; Aldrich, 1995; Cox and McCubbins, 2007). These studies and many that followed utilized NOMINATE scores to empirically test these points. NOMINATE scores were also used to show that parties do not have much influence over the voters of their members (Cox and Poole, 2002). These opposite findings raise causes for concern when using NOMINATE scores to test theories of party effects.

The main problem that arises when using NOMINATE scores and other roll call scaling techniques to test theories of party control is that, *a priori*, groups are assumed to be orthogonal noise. These stimation procedure require that errors are assumed to be independent and identically distributed across both individuals and bills, conditional on the uncovered latent variables. Given that there appears to be grouping that arises according to party membership in these latent scores, the possibility that group effects exist should, at least, be considered a possibility when estimating the model. Thus, flexible models that allow for less restrictive assumptions are needed in order to better test theories of voting in Congress.

Using the clustering approach proposed by Spirling and Quinn (2010) as a starting point, I propose a new model which allows for both individual ideal points and group ideal points and membership to be measured. This model, Clustered Beta Process Confirmatory Factor Analysis (CBPCFA) utilizes a combination of beta process priors on the number of dimensions for the uncovered ideal point space (Knowles and Ghahramani, 2011) and Dirichlet process priors (Ferguson, 1973) on group membership within the set of roll call votes. This model flexibly estimates continuous level ideal points for each individual voter with unknown dimensionality while also uncovering a set of discrete groupings within the data and estimating the group effect on voting. This model uses Bayesian nonparametric approaches, meaning that no theory of the mechanisms of voting are specified and the data dictates whether groups or dimensions are needed to best model the voting behavior of members of

Congress. By not constraining the model to fit a rigid set of assumptions, the data dictates which potential mechanisms of vote behavior are important and estimates the corresponding measures.

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*, ξ_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:

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

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

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:

$$\begin{aligned} P(y_{i,j} = 1) &= P(U_i(A_j) > U_i(S_j)) && (2.1)\\ &= P(-\|\xi_i - A_j\|^2 + \eta_{i,j} > -\|\xi_i - S_j\|^2 + \nu_{i,j}) && (2.2)\\ &= P(\nu_{i,j} - \eta_{i,j} < -\|\xi_i - S_j\|^2 - \|\xi_i - A_j\|^2) && (2.3)\\ &= P(\nu_{i,j} - \eta_{i,j} < 2(A_j - S_j)'x_i + S_j'S_j - A_j'A_j) && (2.4)\\ &= \Xi(\lambda_j'\xi_i - \alpha_j) && (2.5) \end{aligned}$$

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:

$$\mathcal{L}(\alpha, \mathbf{\Lambda}, \mathbf{\Omega}|\mathbf{Y}) = \prod_{i=1}^{N}\prod_{j=1}^{P} \Xi(\lambda_j'\xi_i - \alpha_j)^{y_{i,j}} \times \left(1 - \Xi(\lambda_j'\xi_i - \alpha_j)\right)^{1-y_{i,j}} \tag{2.6}$$

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.

McAlister (2018) derives a similar model that allows for appropriate estimation of the dimensionality of the resulting ideal point space. Unlike previous approaches that require post-hoc testing of the number of dimensions, often erroneously leading to conclusions that one dimension is appropriate for describing the policy preferences of members of the U.S. Congress, a Bayesian nonparametric approach is used that places a *prior* on the number of dimensions in the model. Using the Indian Buffet Process prior for the number of dimensions needed to represent the ideal point space, a slightly altered formulation is proposed:

$$P(y_{i,j} = 1) = \Xi((r_j \odot \lambda_j)'\xi_i - \alpha_j) \tag{2.7}$$

where r_j is a binary vector that takes a value of zero if bill j does not influence a dimension and a one otherwise. This specification induces a spike-and-slab prior on the bill loadings and the infinite nature of the priors allows for estimation of a sparse set of dimensions.

The main workhorse of the BPIRT model is the *Indian Buffet Process*, which places a prior on the number of dimensions needed to best explain the latent ideal point space. The Indian Buffet Process is a version of a stochastic Beta process which is a sparsity inducing prior. This approach allows for each bill-dimension pair to be explicitly tested against the hypothesis of no effect; a test that tells whether the collection of bill j provides any information about

the location of the ideal point on dimension k. The dimensionality of the set of ideal points is a substantively meaningful variable (McCarty et al., 2016) and proper estimation is a meaningful task (Aldrich et al., 2014). For these reasons, this specification of the ideal point model will serve as a starting point for the model derived in this paper.

2.1 A Model for Roll Call Analysis with Group Influence

Now, legislator i must choose whether to vote for A_j or S_j. Like before, she uses a utility maximization model that assumes quadratic loss in distance from her ideal point. However, assume that she belongs to a group, $g_i \in (1, ...G)$ that influences her vote choice. Members of the same group have similar interests and share a desire to pass legislation that is beneficial to the group. Let τ_i be the *ideal point* associated with g_i and assume that this effect is additively separable from the individual ideal point. Formally, this implies that the new utility functions are:

$$U_i(A_j) = -\|\xi_i + \tau_i - A_j\|^2 + \eta_{i,j}$$
$$U_i(S_j) = -\|\xi_i + \tau_i - S_j\|^2 + \nu_{i,j}$$

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

As before, these equations can be rearranged to give the standard IRT formulation:

$$P(y_{i,j} = 1) = \Phi((r_j \odot \lambda_j)'(\xi_i + \tau_i) - \alpha_j) \tag{2.8}$$

where $\Phi(\cdot)$ is the standard normal CDF.

This equation can be expressed in an alternative form using an *auxiliary variable* for $y_{i,j}$. Let $x_{i,j}$ be the auxiliary variable and let $\mathcal{N}(x_{i,j}; \lambda_j'(\xi_i + \tau_i) - \alpha_j, 1)$ be the probability density function associated with the auxiliary variable. Then:

$$P(y_{i,j} = 1) = \int_0^\infty \mathcal{N}(x_{i,j}; (r_j \odot \lambda_j)'(\xi_i + \tau_i) - \alpha_j, 1) dx_{i,j} \tag{2.9}$$

where $E[x_{i,j}] = (r_j \odot \lambda_j)'(\xi_i + \tau_i) - \alpha_j$. Using $x_{i,j}$ and linking it to the linear factor analysis predictor, the IRT formulation can be rearranged as:

$$P(y_{i,j} = 1) = P((r_j \odot \lambda_j)'(\xi_i + \tau_i) - \alpha_j + \epsilon_{i,j} > 0) \tag{2.10}$$

where $\epsilon_{i,j} \sim \mathcal{N}(0, 1)$. Rearranging a final time:

$$P(y_{i,j} = 1) = P((r_j \odot \lambda_j)'\xi_i + (r_j \odot \lambda_j)'\tau_i - \alpha_j + \epsilon_{i,j} > 0) \tag{2.11}$$

This construction allows the latent variable model to be written in a familiar regression-like specification - the observed value is connected to a set of coefficients and **unobserved** covariates. Instead of directly observable covariates, this model implies that there are two voter level covariates that are indirectly observable: a group level ideal point and a structured individual level departure from the group level ideal point.

This construction allows the difference between the two models to be easily seen. The IRT construction of the group effects model can be directly compared to the IRT construction for the standard model:

$$P(y_{i,j} = 1) = P((r_j \odot \lambda_j)'\xi_i - \alpha_j + \epsilon_{i,j} > 0) \tag{2.12}$$

If the standard model is assumed, then the group effect is assumed to not exist and the model is estimated under the original specification. Assuming that the grouped model is correct, then estimation using the standard model pushes the group estimation to the *error term*. Much like the regression context, this may lead to *omitted variable bias* in the estimation of Λ and Ω. Specifically, if ξ_i and τ_i are correlated, then the estimates of the ideal point under the standard specification will be biased. Given that the ideal points are a key quantity of interest, it is important that the point estimator used for each value is unbiased.

There is also evidence that misspecification of group effects in the ideal point model can lead to poor inference about the number of dimensions needed to model the ideal point space. Aldrich et al. (2014) show that in simulations of ideal point spaces with a known number of policy dimensions, the standard roll call scaling model tends to underestimate the dimensionality of the space when there are strong group effects. This problem can be mitigated by scaling any known groups separately. While the American legislative context may appear to lend itself to prior specification of voting groups, this assumption might be too strong. In order to appropriately estimate dimensionality, systematic correlations among votes due to groups should be taken into account.

Substantively, this model addresses questions of group influence within the ideal point model. Like regression models, problems arise if the ideal point is correlated with the error function. Given that the errors corresponding to the latent predictors are assumed to be independent and identically distributed in the standard specification, any covariate which is not controlled in the model specification that is correlated with the errors can lead to poor estimation of the structural parameters. In the context of roll call voting, one situation which can lead to this outcome is difference in group strategies. Imagine that two parties exist in a legislature and one party exerts strong party control on a specific bill type such that all members of the party must vote together on that bill. On the other hand, the other party allows its members to vote for their own preferred outcome, regardless of the party preference. If the ideal points are assumed to be estimates of the true preferences of the voters, then the errors associated with this vote will be correlated with party - the strong force party members will exhibit errors that are larger than those related to members of the weak force party. This difference in strategy leads to a violation of the standard model assumptions and, in turn, leads to biased estimates of the ideal points from the estimation procedure.

Beyond problems of dependence in errors, roll call scaling using the standard model also suffers from attempting to place a discrete latent variable on a continuous scale. Theories of the role of groups in the U.S. Congress posit that there is a tiered approach to vote decision making - first members of the party consult with the party desire then decide whether or not they should depart from the party wishes (Aldrich and Rohde, 2000; Rohde, 2010; Cox and McCubbins, 2007). While there are certainly cases where this model is not appropriate, there is evidence that party plays a role in the vote decision. The first dimension of the NOMINATE model is correctly interpreted as party loyalty (Lee, 2009), and is often made up of distinctive clusters of ideal points. These clusters are interpreted as parties and the distances are used to make statements related to the behavior of parties in Congress (McCarty et al., 2016). However, this usage of ideal points is not supported by the theory of the model - if party is a cause of votes, then party should be accounted for in the ideal point estimates. However, party is a discrete covariate. Thus, the continuous model is unequipped to properly estimate these effects. The effect of mapping a discrete latent variable to a continuous manifold is explored by Aldrich et al. (2014) and its shown that this leads to an understatement of the dimensionality of the ideal point space. The party effect model derived above provides an approach to properly estimating the underlying group latent variables.

There are two significant challenges in estimating and interpreting the group model. First, it is too strong of an assumption to explicitly specify that group effects exist in a roll call model. While there are theories of voting in legislative chambers that posit that various groupings matter, there are corresponding theories which say the opposite. For this reason, it is important that any estimation procedure that attempts to model group effects is flexible and allows for the possibility that group effects may not exist - specifying group effects when group effects do not exist leads to inefficient estimations of the ideal points, which lead to inefficient estimates of the other structural parameters in the roll call model.

Second, given the problems of rotational invariance present in estimates of the structural parameters of the IRT model, the estimates for ξ_i and τ_i are not uniquely identifiable. Under the standard IRT formulation, ξ_i has infinite solutions without further constraints. For similar reasons, attempting to estimate two components of an additive effect leads to more problems and requires more stringent constraints in order to identify a unique solution. This is problematic, as the constraints placed on structural parameters in latent variable models are often too restrictive and, necessarily, lead to solutions that are contingent on the ordering of the observations and items (Bhattacharya et al., 2011). For these reasons, it is worthwhile to derive an estimation strategy for the group effects model that allows for *infinitely exchangeable* prior specification. Similarly, rather than estimating ξ_i and τ_i separately, the sum of the two is estimated, $\omega_i = \xi_i + \tau_i$. ω_i should differ from the estimate of ξ_i in the standard model by properly inducing conditionally exchangeable errors - $P(\epsilon_{i,j}|\omega_i, g_i) \perp P(\epsilon_{h,j}|\omega_j, g_h)$.

3 Uncovering Groups in the Ideal Point Model

Mentioned above, a key assumption for the latent variable models previously discussed is that each of the N observations are *independently and identically distributed*. Each of the i random vectors associated with the manifest set are modeled separately and assumed to constitute a joint likelihood:

$$P(Y|-) = \prod_{i=1}^{n} P(y_i|-)$$

At face value, this assumption seems relatively innocuous. However, much of the estimation procedure used to estimate these models relies heavily on this assumption and warrants further exploration and validation.

While i.i.d assumptions are frequently referenced, a lighter assumption will be explored here - the assumption of exchangeability. Exchangeability (De Finetti, 1992) can be best defined by a thought experiment - can the labels on each observation be permuted and the same end results achieved? Note that this is a necessary condition for i.i.d to hold. Exchangeability violations lead to a violation of the assumption of idiosyncratic errors. In other words, the assumption that the distribution of errors, ϵ, is equal across observations is violated when exchangeability does not hold.

In the context of ideal point estimation, the exchangeability assumption arises in the estimation of the latent variables, Ω. *A priori*, ω_i is often assumed to follow:

$$\omega_i \sim \mathcal{N}_k(0, I_k) \tag{3.1}$$

meaning that regardless of underlying covariates, the latent variables are drawn from the same distribution.

Immediately, this assumption provides points for concern in the context of roll call scaling. Linked back to the formal model that gives rise to the scaling model, there is an assumption that the errors associated which each utility calculation are independent of one another. However, this is not likely to be true - work on parties and geographic dependencies across votes shows that there is likely some influence in voting across members of groups. This inconsistency has led to questions about the appropriateness of the roll call scaling model and has given rise to alternative implementations that do not require this assumption to hold (Spirling and Quinn, 2010).

To understand the impact of dependencies within the data, imagine that there exists groupings amongst the observations in the observed data set. The set of groups, G, contains distinct groups, g, such that $g \in 1 \leq g \leq G$. Each group is associated with a different error function such that the probability of the observed data under the latent variable model is:

$$P(X) \propto \int_{\infty}^{-\infty} \int_{\infty}^{-\infty} \int_{G} \int_{\Theta} P(X|G, \Omega, \epsilon, \Theta) P(\Omega|G) P(\epsilon|G) P(G) P(\Theta) d\Theta dG d\Omega d\epsilon \tag{3.2}$$

where Θ is used to denote all other parameters in the model not explicitly denoted in the equation and X is the matrix of auxiliary variables associated with Y as defined above.

This function can be compared to the probability of the manifest set when no groups are assumed to exist:

$$P(X) \propto \int_{\infty}^{-\infty} \int_{\infty}^{-\infty} \int_{\Theta} P(X|\epsilon, \Theta) P(\Omega) P(\epsilon) P(\Theta) d\Theta d\Omega d\epsilon \tag{3.3}$$

These two functions are equivalent if and only if there is only **one** group within the data. This formulation creates an immediate symmetry between exchangeability and omitted variables - if a covariate that correlates with the errors produced by the error function is omitted from the model specification, then the errors no longer meet the exchangeability conditions needed for the model to be properly specified. While this notion has been explored heavily in other areas of applied statistics, violations in latent variable models are no less damaging than with other models.

Aside from potential bias in estimation for the factor loadings and error terms, violation of exchangeability of this form leads to biased inference on the number and meanings of the dimensions uncovered by latent variable techniques. Let $x_{i,k} = y_i - \omega'_{i,-k}\Lambda_{-k} + \alpha$ be the residuals left for y_i when dimensions $1, 2, ..., k-1, k+1, ..., K$ taken into account. Marginally, $x_{i,k}$ under the standard specification follows:

$$P(x_{i,k}) \propto \int_{\Theta} P(x_{i,k}|g_i, -) P(g_i) d\Theta \tag{3.4}$$

which implies that:

$$P(x_{i,k}) = P(x_{i,k}, g_i) \tag{3.5}$$

where g_i is the group associated with observation i. By the definition of joint probability, (3.5) is true if and only if:

$$P(g_i|x_{i,k}) = 1 \tag{3.6}$$

which is only true when there is only one group. If not accounted for, the residual for $x_{i,k}$ is correlated with g_i. Assume that g_i accounts for some portion of the covariance of the manifest set, then these residuals will be mixed within the dimensions uncovered in the estimation process, meaning that one or more dimensions will correlate with group membership rather than explaining covariance in the manifest set related to a continuous manifold.

This, in and of itself, is a problem for ideal point estimation procedures. Given that the procedure is attempting to measure *continuous* manifolds, measuring a truly discrete manifold in this manner is a violation of the procedure's underlying assumptions. In turn, as this manifold is being improperly measured, the amount of variance explained by the "continuous" latent variable will be overestimated. This can lead to many problems in the

inferential task, most egregiously distorting the importance of the discrete manifolds. In other words, improperly measuring discrete manifolds will lead to underestimation of the number of latent dimensions needed to properly model the latent space.

As alluded to above, properly accounting for discrete manifolds in this context will lead to a properly specified model. One approach is to properly model the latent variable in relation to latent groups, $P(\omega_i|g_i)$ and, in turn, properly model the error terms, $P(\epsilon_i|g_i)$. This can be seen as an analogue to hierarchical regression models, where $P(\epsilon_i)$ is assumed to be *conditionally exchangeable* when conditioned on *cluster* membership.

This notion of a *hierarchical* formulation of the latent variable has been explored in recent years (Rai and Daumé, 2009; Gruhl et al., 2013), but the applied literature has given very little attention to this class of models. In particular, these model explicitly assume that the groupings are known *a priori*. While this makes more sense in the regression context, understanding the groupings of the latent variable can be an intellectually challenging task. For this reason, I present a model where neither the latent groupings nor the latent variables are known and are imputed by the estimation task leveraging the **Dirichlet process**.

3.1 Uncovering Unknown Latent Groups

I seek to derive an approach to estimating groups for the ideal point model:

$$P(y_i) = \int_0^\infty \int_\Theta \int_{\omega_i} \int_G \mathcal{N}_p(x_i; \Lambda\omega_i - \alpha, \Psi) P(\omega_i|g_i) P(g_i) dG d\omega_i d\Theta dx_i \tag{3.7}$$

where g_i is a categorical variable with probability densities corresponding to the probability that observation i is part of cluster $g \in 1 \leq g \leq G$.

In the case where g_i is unknown, but the number of clusters, C, is known, then a prior can be placed on g_i such that:

$$P(g_i) \sim Cat(\zeta) \tag{3.8}$$

where $Cat(\zeta)$ corresponds to a multinomial distribution where one draw is taken. ζ is a set of C probabilities that sums to 1.

However, the case of interest in this context is one where the number of clusters within the data is *unknown*. Thus, a standard categorical distribution is inappropriate. For this situation, a Bayesian nonparametric approach will be used and g_i assumed to follow a Dirichlet process (Ferguson, 1973):

Definition. *Let H be a distribution over Θ and α be a positive real number. Then for any finite measurable partition $A_1, ..., A_C$ of Θ the vector $(G(A_1), ..., G(A_C))$ is random since G is random. G is Dirichlet process distributed with base distribution H and concentration*

parameter α*, written* $G \sim DP(\alpha, H)$*, if*

$$P(G(A_1), ..., G(A_C)) \sim Dir(\alpha H(A_1), ..., \alpha H(A_C))$$

for every finite partition of Θ*. Draws from this process will be discrete with probability 1.*

The Dirichlet process can be seen as a prior over distributions. Each $g \in 1 \leq g \leq G$ constitutes a cluster with an associated distribution. The Dirichlet process prior admits a probability density over the membership for observation i. In other words:

$$P(g_i) \sim DP(\alpha, H)$$

Given this choice of prior, the posterior distribution, $P(G|\Theta)$ can be defined. Due to the conjugacy of the Dirichlet prior to the categorical distribution $P(\Theta|G)$, the resulting posterior follows a Dirichlet distribution.

Given the form of the posterior distribution, we know the joint probability that $G = G^*$ for some permutation of group labels. However, this is not the quantity of interest. Rather, the inferential task asks to find $P(g_i|\Theta)$. In order to do this, define g_{-1} as the group labels for the observations **other** than observation i. Given that the Dirichlet process is *infinitely exchangeable*, a predictive distribution can be defined and used to examine the individual probability distributions:

$$P(g_i) = \frac{1}{\alpha + n} \left(\alpha H + \sum_{j=1}^{n-1} \delta_{g_j} \right) \tag{3.9}$$

where H is the base measure, α_{CRP} is the DP hyperparameter, and $\delta_{g_j} = 1$ if observation j is in group g. In words, this says that the probability that $i \in g$ is proportional to the number of other observations that are already in group g. There is also the probability that i belongs in its own cluster, which is quantified by the base measure H.

Using this structure, a process can be defined for the case of interest- a case where the number of groups is unknown *a priori*. By driving $C \to \infty$, a process can defined where g_i is modeled as an *infinite mixture*. In words, this means that there are infinite possible cluster components within the data. In order for the model to be tractable, a finite number of these components should be used within the data. Thus, of the $C = \infty$ possible clusters that g_i can join, only C^+ have a positive probability of occurrence. This process constitutes the *Chinese Restaurant Process* (Rasmussen, 2000; Pitman et al., 2002; Blackwell and MacQueen, 1973; Escobar and West, 1995). Under this prior, each observation is assumed to take on a currently defined cluster with probability proportional to the number of other observations that are currently in that cluster and take a new cluster assignment with probability proportional to α. This prior is sparsity inducing as the expected number of clusters grows with n:

$$E[C^+] \approx O(\log(n)) \tag{3.10}$$

This implies that the complexity of the inferred mixture can only be as complex as the sample size allows. This is desirable as it prevents against overfitting.

As with all mixture models that are estimated using a generative model, the cluster assignments achieved using the CRP are unique up to a permutation of labels; the model is identified, but the individual cluster assignments are not. However, the groupings are identifiable. For this reason, inference performs not on the individual cluster labels, themselves, but the probability that $g_i = g_j \forall i \neq j$.

3.2 Latent Groups within Ideal Point Estimation

Using the previous theory, an ideal point model is developed allowing there to be latent groupings within the latent variable. This model can be defined as:

$$P(y_i = 1) = \int_0^\infty \int_\Theta \int_{\omega_i} \int_G \mathcal{N}_p(x_i; \Lambda\omega_i - \alpha, \mathcal{I}_p) P(\omega_i|g_i) P(g_i) dG d\omega_i d\Theta dx_i \tag{3.11}$$

Note that g_i only arises in the estimation of the latent variable ω_i as the errors are assumed to be conditionally exchangeable given g_i.

The key quantity of interest in this model is $P(g_i|g_{-i})$. Using Bayes rule, this quantity can be defined as:

$$P(g_i|g_{-i}) \propto \int_{\omega_i} \int_G \mathcal{N}_p(x_i; \Lambda\omega_i - \alpha, \mathcal{I}_p) P(\omega_i|G) d\omega_i dG \tag{3.12}$$

Without further specification, this model is intractable.

For the purposes of ideal point analysis, multivariate normal clusters will be assumed. For each $g \in G$, the corresponding cluster is assumed to follow a K^+ dimensional multivariate normal cluster with mean vector μ_g and diagonal covariance matrix Σ_g, where K^+ is dictated by the beta process prior on the loadings matrix. Standard conjugate priors are assumed on these parameters. Plugging this into (3.12), we can redefine the model as:

$$P(g_i = g|g_{-i} \in g) \propto \int_{\omega_i} \int_{\mu_g} \int_{\Sigma_g} \mathcal{N}_p(x_i; \Lambda\omega_i - \alpha, \mathcal{I}_p) \mathcal{N}_{K^+}(\omega_i; \mu_g, \Sigma_g) \mathcal{N}_{K^+}(\mu_g; \mu_0, (\kappa_0\Sigma_g)^{-1}) IG(\Sigma_g; \alpha_0, \beta_0) d\Sigma_g d\mu_g d\omega_i \tag{3.13}$$

This quantity encapsulates the probability that i belongs in cluster g given all other observations that are currently in g.

This integral can be simplified by recognizing that the latter part of the integral is simply the posterior predictive distribution for standard conjugate multivariate normal analysis. Simplifying (3.12) gives:

$$P(g_i = g|g_{-i} \in g) \propto \int_{\omega_i} \mathcal{N}_p(x_i; \Lambda\omega_i - \alpha, \mathcal{I}_p)\mathcal{T}_{K^+;2\alpha_g}\left(\mu_g, \frac{\beta_g(\kappa_g+1)}{\alpha_g\kappa_g}\right) d\omega_i \tag{3.14}$$

Let $\omega_{-i,g}$ be the values of the latent variable for all observations in cluster g **not including** observation i and n_g be the number of observations currently in g **excluding** observation i. Define:

$$\bar{\omega}_{-i,g} = \frac{\sum \omega_{-i,g}}{n_g}$$

$$\mu_g = \frac{\kappa_0\mu_0 + n_g\bar{\omega}_{-i,g}}{\kappa_0 + n_g}$$

$$\kappa_g = \kappa_0 + n_g$$

$$\alpha_g = \alpha_0 + \frac{n_g}{2}$$

$$\beta_g = \beta_0 + \frac{1}{2}\sum_{j=1}^{n_g}(\omega_{-i,g,j} - \bar{\omega}_{-i,g})^2 + \frac{\kappa_0 n_g(\bar{\omega}_{-i,g} - \mu_0)^2}{2(\kappa_0 + n_g)}$$

Unfortunately, this integral has no analytical solution; numerical methods must be used. An accurate approach utilizes a Laplace approximation to the multivariate t-distribution. Using the quadrature approach, approximate this distribution:

$$\mathcal{T}_{K^+,2\alpha_g}\left(\omega_i; \mu_g, \frac{\beta_g(\kappa_g+1)}{\alpha_g\kappa_g}\right) \approx \mathcal{N}_{K^+}\left(\omega_i; \mu_g, \frac{\beta_g(\kappa_g+1)}{\left(\alpha_g + \frac{1}{2}\right)\kappa_g}\right) \tag{3.15}$$

This allows the integral to be solved analytically:

$$\int_{-\infty}^{\infty} \mathcal{N}_p(x_i; \Lambda\omega_i - \alpha, I_p)\mathcal{T}_{K^+,2\alpha_g}\left(\omega_i; \mu_g, \frac{\beta_g(\kappa_g+1)}{\alpha_g\kappa_g}\right) d\omega_i \approx \exp(q^* - q) \tag{3.16}$$

where

$$\Xi_g = \frac{\beta_g(\kappa_g+1)}{\left(\alpha_g + \frac{1}{2}\right)\kappa_g}$$

$$A = -\frac{1}{2}(\Lambda'\Lambda + \Xi_g^{-1})$$

$$b = (y_i + \alpha)'\Lambda + \mu_g'\Xi_g^{-1}$$

$$q^* = -\frac{1}{2}((y_i + \alpha)'(y_i + \alpha) + p\log(2\pi) + \mu_g'\Xi_g^{-1}\mu_g + K^+\log(2\pi) + \log(det(\Xi_g)))$$

$$q = \frac{1}{4}bA^{-1}b' - \frac{K^+}{2}\log(2\pi) - \frac{1}{2}\log\left(\det(-\frac{1}{2}A^{-1})\right)$$

Given the choice to use a CRP prior on g_i, a base distribution must also be specified from which new clusters are proposed. In the context of factor analysis, it makes sense to define the base distribution:

$$H \sim \mathcal{N}_{K^+}(0, I_{K^+}) \tag{3.17}$$

which is equivalent to the standard prior utilized in factor analysis procedures. This allows the probability that g_i belongs in a new cluster to be defined as:

$$P(g_i = g_{new}) \propto \int_{-\infty}^{\infty} \mathcal{N}_p(y_i; \Lambda\omega_i - \alpha, I_p)\mathcal{N}_k(\omega_i; 0, I_k)d\omega_i = \exp(\dot{q^*} - \dot{q}) \tag{3.18}$$

where

$$\dot{A} = -\frac{1}{2}(\Lambda'\Lambda + I_k)$$

$$\dot{b} = (y_i + \alpha)'\Lambda$$

$$\dot{q^*} = -\frac{1}{2}((y_i + \alpha)'(y_i + \alpha) + p\log(2\pi) + K^+\log(2\pi))$$

$$\dot{q} = \frac{1}{4}\dot{b}\dot{A}^{-1}\dot{b}' - \frac{K^+}{2}\log(2\pi) - \frac{1}{2}\log(\det(-\frac{1}{2}\dot{A}^{-1}))$$

Given these quantities, we can define the conditional posterior distribution for g_i. Due to the conjugacy of the DP prior and the categorical likelihood, the posterior will follow a Dirichlet distribution:

$$P(g_i|-) \sim Dirichlet(n_g P(g_i = g|g_{-i}) \ \forall \ g \in G, \alpha_{CRP} P(g_i = g_{new}) \tag{3.19}$$

The probability that g_i belongs to group g is a weighted average of the number of other observations in g and the probability that g_i is in the same cluster as the other like observations.

The differences between this approach and the standard continuous latent variable model arise in how each latent variable is taken into account. In the standard procedure, each ω_i is assumed to be *independent* of the other observations. In the clustered case, each ω_i is assumed to belong to a cluster and information about the location of the latent variable is shared between observations in the same cluster. In other words, latent variables are essentially estimated in groups rather than individually. Using the exchangeability properties of the Dirichlet process, this addresses the initial exchangeability problem by partitioning the set

of observations into conditionally exchangeable groups - observations are assumed to be exchangeable within the same cluster. This allows group properties to propagate through the latent variable rather than emerging as its own set of orthogonal latent manifolds.

This property can be expressed in another way - the latent variable associated with each individual is a function of both group and individual dynamics. Given the construction of the latent variable, there are two influences on the posterior distribution of each idea point: a group effect and an individual effect. Here, the group effect can be seen as a shared prior on the latent variable that is common across members of the same group. Then, the individual effect can be seen as a noise component that is separate from the group effect. Using the Dirichlet process approach allows the individual effect to be modeled as conditionally independent from the group effect, given the specification of the model. While it is not possible to directly estimate these two effects due to the rotational invariance that is inherent in latent variable models, post-processing procedures can be used to estimate the two effects given a specific rotation of the latent structural parameters.

A final consideration for the latent variable generating process is to examine the CRP hyperparameter, α_{CRP}. This parameter controls how likely it is for each observation to dictate a new cluster. When α_{CRP} is small, the *a prior* likelihood that an observation in cluster g finds a new unique cluster decreases. On the other hand, when αCRP is large, clusters will have lower overall membership and be less "stable" when estimated. Finding the appropriate value for α_{CRP} is key for estimation. Escobar and West (1995) shows that α_{CRP} follows a gamma distribution and a conditional posterior can be defined as:

$$P(\alpha_{CRP}|-) \sim Gamma(1 + C^{+}, 1 + \gamma_c log(n)) \tag{3.20}$$

where C^{+} is the number of active clusters and γ_c is Euler's constant.

4 A Clustered Latent Variable Approach to Roll Call Scaling

Utilizing the theory above, a roll call scaling procedure utilizing sparsity inducing beta process priors for measuring the dimensionality of the space and Dirichlet process priors on the latent variables is derived. Let $x_{i,j}$ be a an augmented representation of the binary variable $y_{i,j}$ 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 clustered beta process ideal point model (CBPCFA) as:

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

where R is a $p \times K$ binary matrix. $\odot$ is Hadamard multiplication, which is equivalent to elementwise multiplication in this situation.

This specification induces *spike and slab* priors on the matrix of loadings, Λ, and the latent variables, Ω. 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 matrices by allowing elements to take non-zero values if and only if a non-zero value adds something over fixing the value at zero.

This construction allows a full model definition:

$$\begin{aligned}
&P(x_i|-) \sim \mathcal{N}_p((R \odot \Lambda)\omega_i - \alpha, I_p) \\
&P(\omega_{i,k}|g_i) \sim \mathcal{N}_{K^+}(\mu_{g_i}, \Sigma_{g_i}) \\
&P(\lambda_{j,k}|r_{j,k}) \sim r_{j,k}\mathcal{N}_p(0, \gamma_{j,k}^{-1}) + (1 - r_{j,k})\delta_0 \\
&P(r_{j,k}) \sim IBP(\pi_{IBP}) \\
&P(g_i = g|\mu_g, \Sigma_g) \sim CRP(\alpha_{CRP}) \\
&P(\mu_g|\Sigma_g) \sim \mathcal{N}_{K^+}(\mu_0, (\kappa_0\Sigma_g)^{-1}) \\
&P(\Sigma_g) \sim IG(\Sigma_g; \alpha_0, \beta_0) \\
&P(\gamma_{j,k}) \sim Gamma(c, d)
\end{aligned}$$

where $i \in 1 \leq i \leq n$, $j \in 1 \leq j \leq p$, and $k \in 1 \leq k \leq K$. 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 basic factor analysis model. The manifest variables are decomposed into the loadings matrix, Λ, and the latent variables, Ω. The dimensions of the latent variables are assumed to be orthogonal. Marginally, $P(x_i + \alpha) \sim \mathcal{N}_K(0, \Lambda'\Lambda + I_K)$. The new additions, however, provide interesting properties for the factor analysis procedure. First, the addition of the infinite binary matrix, R, allows us to learn about the true number of latent dimensions within the manifest data. R represents whether feature $k \in 1 \leq k \leq K$ is a meaningful summary of the data. Similarly, the CRP prior on Ω allows us to sort the latent variables into meaningful groups, allowing simultaneous estimation of latent groupings with corresponding location and spread information. In contrast to simply clustering the data, this model grants the desirable properties of knowing latent groups while also having the same continuous measure properties that make factor analysis an attractive approach to estimating a latent space.

5 Estimation via Gibbs Sampling

Estimation using Gibbs sampling proceeds in the following way:

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

$$x_{i,j} \sim \begin{cases} \mathcal{TN}_{-\infty,0}(\lambda_j \omega_i - \alpha_j, 1) \text{ if } y_{i,j} = 0 \\ \mathcal{TN}_{0,\infty}(\lambda_j \omega_i - \alpha_j, 1) \text{ if } y_{i,j} = 1 \\ \mathcal{N}(\lambda_j \omega_i - \alpha_j, 1) \text{ if } y_{i,j} \text{ is missing} \end{cases} \tag{5.1}$$

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

$$t_{j,k} = \frac{P(r_{j,k} = 1|Y, -)}{P(r_{j,k} = 0|Y, -)} = \frac{P(Y|r_{j,k} = 1, -)}{P(Y|r_{j,k} = 0, -)} \tag{5.2}$$

$$\frac{P(Y|r_{j,k} = 1, -)}{P(Y|r_{j,k} = 0, -)} = \sqrt{\frac{\gamma_k}{\gamma}} \exp\left(\frac{1}{2}\gamma\mu^2\right) \tag{5.3}$$

   where $\gamma = \omega_k'\omega_k + \gamma_k$, $\mu = \frac{1}{\gamma}\omega_k'\hat{E}_j$, $\hat{E}_j = x_j - \lambda_j\Omega + \alpha$ setting $\lambda_{j,k} = 0$. Let

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

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

3. **Sample Ω.** For each $i \in 1 \leq i \leq n$, define Q_i as a $p \times K^+$ matrix where each column is q_i. Let μ_i be a K^+-column vector including the cluster means for g_i and Σ_i a $K^+ \times K^+$ diagonal covariance matrix with the variances for g_i. Sample ω_i from:

$$P(\omega_i|-) \sim \mathcal{N}_{K^+}(m_i, V_i) \tag{5.4}$$

   where:

$$V_i = \left(\Lambda'\Lambda + \Sigma_i^{-1}\right)^{-1}$$

$$m_i = V_i\left(\Lambda'(x_i + \alpha) + \mu_i'\Sigma_i^{-1}\right)$$

4. **Remove Inactive Features, Normalize Λ and Ω.** For each $k \in 1 \leq k \leq K^+$, if $r_{j,k} = 0 \; \forall \; 1 \leq j \leq p$ or $q_{i,k} = 0 \; \forall \; 1 \leq i \leq n$, remove k from the analysis. Recalculate K^+. Post-process Λ to normalize the variance. For each $j \in 1 \leq j \leq p$ and $k \in 1 \leq$

$k \leq K^+$ set $\lambda_{j,k}$:

$$\lambda_{j,k} = \frac{\lambda_{j,k}}{\sqrt{1 + \sum_{h=1}^{K^+} \lambda_{j,h}^2}} \tag{5.5}$$

Post process Ω to normalize location and variance. For each $k \in 1 \leq k \leq K^+$, set $\omega_{i,k}$:

$$\omega_k = \frac{\omega_k - \bar{\omega}_k}{sd(\omega_k)} \tag{5.6}$$

5. **Sample Latent Groups, G.** At each iteration, let G^+ be the current set of active clusters. Shuffle the order of observations. For each $i \in 1 \leq i \leq n$, find the probability that $\omega_i \in g \; \forall \; 1 \leq g \leq G^+$. For each existing cluster, calculate $P(g_i = g|\omega_{-i,g})$ where:

$$P(g_i = g|\omega_{-i,g}) \propto \int_{-\infty}^{\infty} \mathcal{N}_p(y_i; \Lambda\omega_i - \alpha, I_p)\mathcal{T}_{K^+,2\alpha_g}\left(\omega_i; \mu_g, \frac{\beta_g(\kappa_g + 1)}{\alpha_g \kappa_g}\right) d\omega_i \tag{5.7}$$

One method of solution is outlined in 3.2.

Determine the probability that observation i belongs in a new cluster:

$$P(g_i = g_{new}) \propto \int_{-\infty}^{\infty} \mathcal{N}_p(y_i; \Lambda\omega_i, I_p)\mathcal{N}_k(\omega_i; 0, I_k)d\omega_i = \exp(\dot{q^*} - \dot{q}) \tag{5.8}$$

This method is also outlined in 3.2.

Draw g_i from:

$$P(g_i|-) \sim Cat(n_h P(g_i = h|\omega_{-i,h}) \; \forall \; h \in 1 \leq h \leq G^+ \; , \; \alpha_{CRP} P(g_i = g_{new})) \tag{5.9}$$

If $g_i = g_{new}$, update G^+ and include the new cluster in proceeding iterations. If there are zero observations in an existing group, update G^+ and remove that cluster from the analysis.

6. **Sample α_{CRP}.** Sample α_{CRP} from:

$$P(\alpha_{CRP}|-) \sim Gamma(G^+, \gamma_c \log(n)) \tag{5.10}$$

where γ_c is Euler's number.

7. **Sample μ_g and Σ_g for all clusters.** For each cluster $g \in 1 \leq g \leq G^+$, sample the mean, μ_g, and covariance, Σ_g, for each Gaussian cluster. Σ_g is assumed to be a diagonal

covariance matrix. Let ω_g be the collection of latent variables that are currently placed in cluster g. For each $k \in 1 \leq k \leq K^+$, draw:

$$\mu_{c,k} \sim \mathcal{T}_{2\alpha_g}\left(\bar{\omega}_{g,k}, \frac{\beta_g}{\alpha_g \kappa_g}\right) \tag{5.11}$$

$$\Sigma_{g,k} \sim InverseGamma(\alpha_g, \beta_g) \tag{5.12}$$

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

$$P(\gamma_k|-) \sim Gamma\left(c + \frac{m_k}{2}, d + \sum_{j=1}^{p} \hat{\lambda}_{j,k}^2\right) \tag{5.13}$$

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

9. **Sample d.** Sample d from:

$$P(d|-) \sim Gamma\left(c_0 + cK^+, d_0 + \sum_{k=1}^{K^+} \gamma_k\right) \tag{5.14}$$

5.1 Posterior Inference

A major problem in the standard ideal point specification is that the estimates for the structural parameters are not uniquely identified without further constraints. We can obtain an identical Ω by multiplying Λ by an orthonormal matrix, M, such that $MM' = I$. 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 Λ 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.

CBPCFA completely avoids the issue of rotational invariance due to its infinite specification. In order for a factor analytic decomposition to be uniquely identified, there must be $\frac{k(k-1)}{2}$ constraints in the loadings matrix. Recall that CBPCFA assumes that $k = \infty$ due to the beta process priors placed on the loadings matrix. In turn, the sparsity inducing spike and slab priors place structural zeroes on any elements of inactive dimensions. Given that only a finite number of dimensions are active at any given point in the model, there also exist infinite inactive dimensions and, thus, infinite structural zeroes. Therefore, there are

always uniquely identified estimates of the structural parameters in CBPCFA.

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 (Doshi et al., 2009).

Inference on the structural parameters of CBPCFA can be done as normal. The posterior distribution of the factor scores, Ω, 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, Λ, which load highly on them; this can be used to determine which covariates influence each latent dimension. Similarly, R can be used to see which covariates have a non-zero contribution to each dimension. This is in direct contrast to standard factor analysis procedures which require that all questions load on all dimensions.

A key quantity of interest using this model is the set of group assignments, G. As mentioned previously, inference on g_i is difficult - when estimated via MCMC, the model is only identifiable up to a permutation of the cluster labels. From iteration to iteration in the Gibbs sampler, there is no guarantee that the cluster labels are significant. However, the membership of the clusters is consistent. Though there are numerous solutions to the label switching problem when the number of clusters is known and fixed, the fact that the number of clusters is stochastic proves problematic for these approaches. Thus, inference about group membership should proceed examining the probability that two or more observations share the same cluster. Using this information, many interesting techniques can be used to establish cluster membership.

6 Simulation Results

To assess the quality of CBPCFA, estimates of the structural parameters are compared against known quantities using a synthetically created data set. For the known data set, $n = 1000$, $p = 500$, and $k = 10$. Each of the manifest variables is designed to be binary. The data set is generated with known Λ, Ω, and idiosyncratic error variances. Λ is also designed with specific structural zeroes, thus a known R is also simulated. To assess the quality of the clustering procedure embedded within this method, Ω was generated to have 4 distinct clusters. These clusters were designed to be distinct, but with some overlap around zero. 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

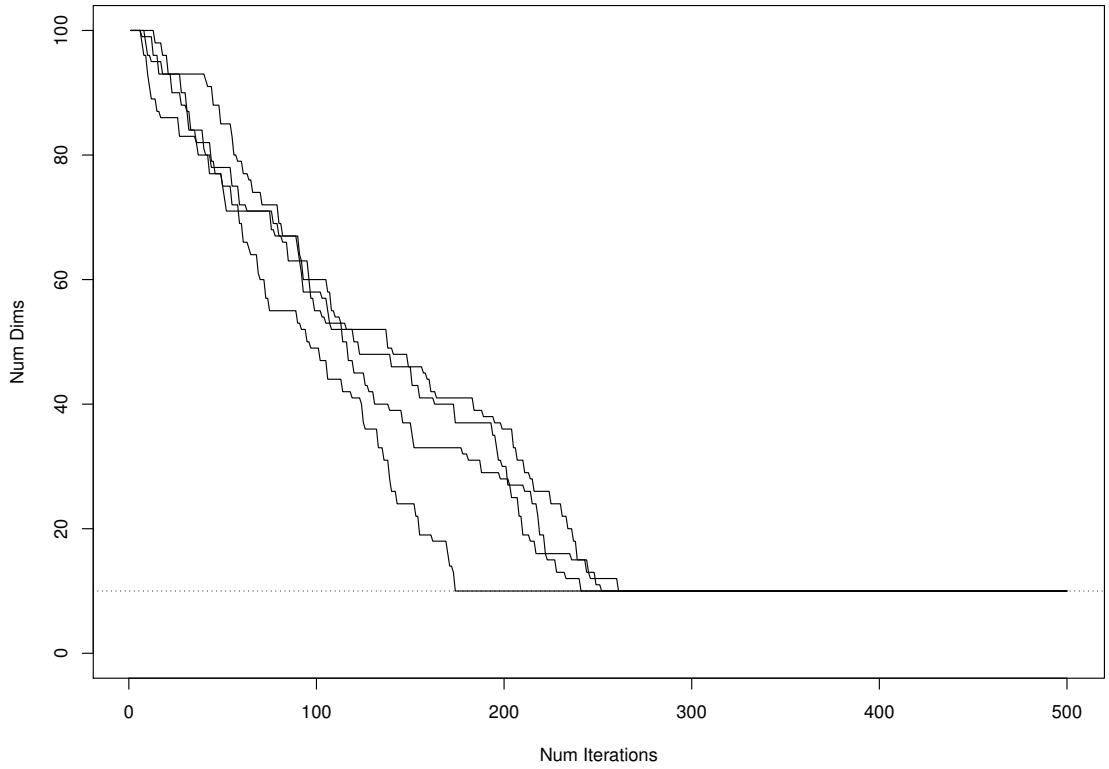

Figure 1: Number of Dimensions over Iterations

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 CBPCFA converges to the appropriate number of dimensions given sufficient n and p.

Since the number of dimensions is correctly identified, R can also be examined to determine how accurately the model discovers the structural zeroes. With relation to Λ, structural zeroes correspond to covariates ($j \in 1 \leq j \leq p$) that contribute only noise variance on a specific latent dimension. Figure 2 shows the true R against the median of the posterior draws for R. In this case, R was recovered with approximately 94% accuracy. This performance is quite good and shows that CBPCFA also does a great job at locating structural zeroes in the loadings matrix.

With regards to the clustering component, the model estimated between four and six clusters with the modal number of clusters at four. More than 95% of the density of the resulting posterior was at four clusters. In order to establish cluster membership, a dyadic network table was computed using the MCMC samples and the probability that $g_i = g_j \; \forall \; 1 \leq i, j \leq n$ was recorded. This was then analyzed as a weighted undirected graph and optimal non-overlapping communities were imputed using a Spinglass algorithm (Orman et al., 2011).

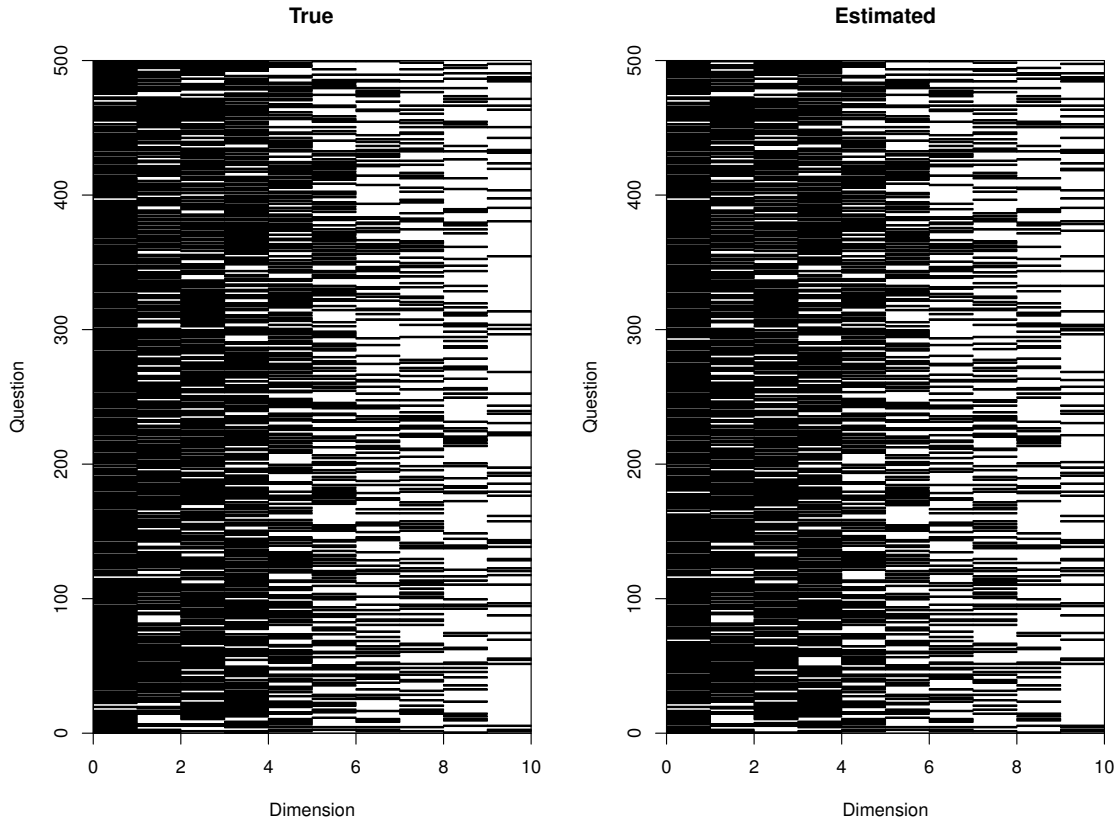

Figure 2: True R vs. Estimated R

	1	2	3	4
1	0	1	0	283
2	231	0	1	3
3	1	226	0	2
4	0	3	245	0

Table 1: Crosstab for True Cluster Assignments vs. Estimated Cluster Assignments

Figure 3 shows the network imputed by the adjacency matrix. Here, we see that this algorithm does a fine job of grouping the correct observations into the same clusters. While there are some observations that fall into the wrong cluster, this is to be expected as the clusters were designed to have some noise and crossover. Table 1 shows a cross-tabulation of the true cluster assignments and the imputed values. This table shows that, overall, the algorithm performs well at determining the number of clusters and the individual cluster memberships. These simulation results are encouraging and show the power of the CRP priors on sorting the latent variables into appropriate groups.

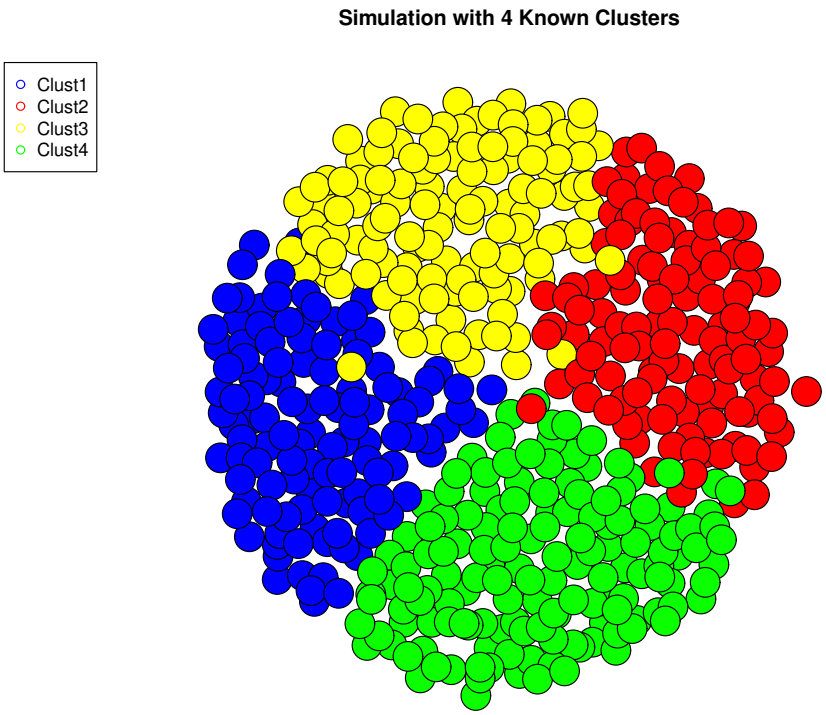

Figure 3: Cluster Membership As Dyadic Network

7 Application to Roll Call Scaling

Using the CBPCFA model, the roll call record from the 114th U.S. House is examined. The 114th session of the U.S. Congress was perhaps best known for its unparalleled obstructionism from both parties. Though the Republicans held a majority in both chambers, they did not have enough votes in the Senate to have a supermajority. In turn, while a record number of bills were proposed and voted upon, the passage rate was minimal due to strong levels of partisan voting. The DW-NOMINATE scaling procedure found this session of Congress to be one of the most polarized of all time and strongly suggested that the first dimension is more than enough for understanding the underlying latent patterns of voting.

In order to analyze the votes, the CBPCFA methodology was used. Four chains of a MCMC procedure were used with a ten-thousand iteration burn-in and 500 iterations captured while thinned to capture every 10 iterations. Convergence diagnostics were examined for the chains of the structural parameters and no issues with convergence were found.

Using the finite beta process approach, the model estimated that six dimensions are needed to estimate the latent space. Figure 4 shows the rate of convergence to the final estimate over the course of the MCMC procedure. This approach converges in on the final answer quickly. Though the other parameters take longer to converge, the number of dimensions needed

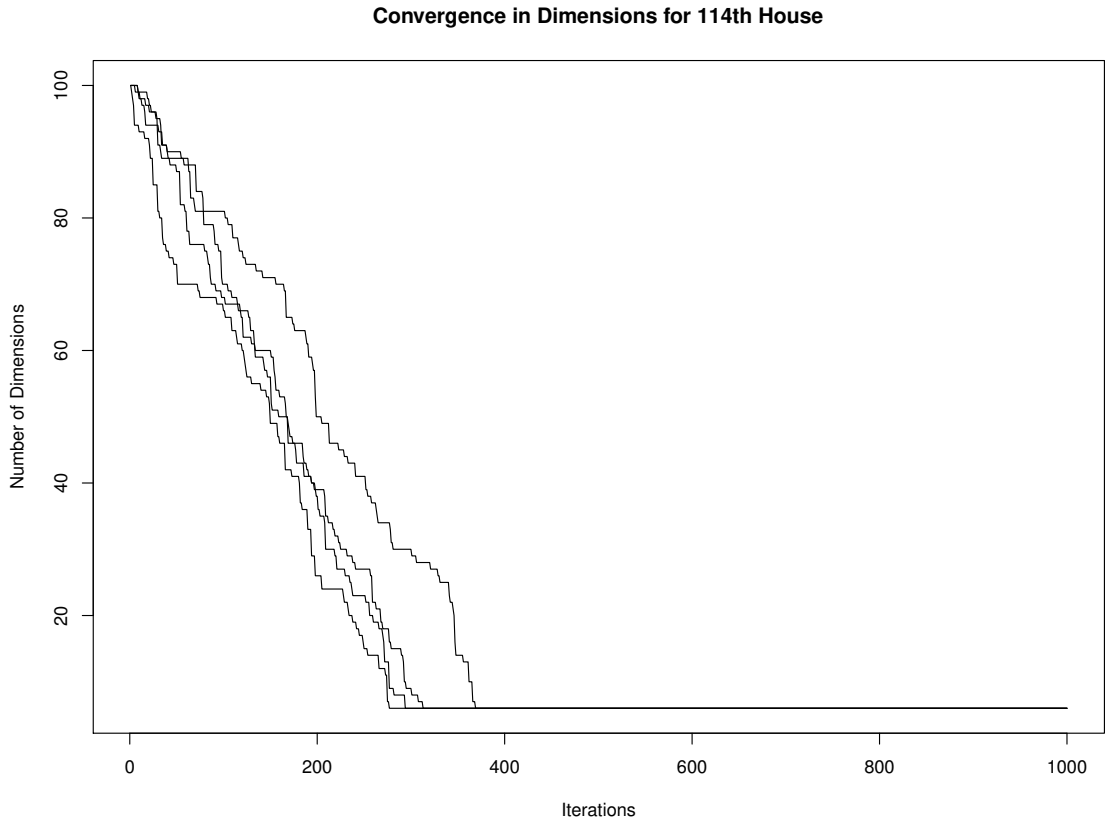

Figure 4: Number of Dimensions Estimated by CBPCFA for the 114th House

seems to converge rather quickly. 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 their respective factor loadings. The most important words were extracted and used to figure out which kinds of bills each dimension covers.

Each of the six dimensions correspond with a specific issue area. Using the above approach, the six dimensions are found to be an economic dimension, a defense dimension, a rural/infrastructure issues dimension, a procedural dimension, a social dimension, and a budget dimension. Examining the roll call record and examining where significant differences between and within parties, these dimensions make sense. For example, this session of the U.S. House had tremendous difficulties passing the required budget measures needed to keep the government operating. This difficulty was cited as one of the main reasons that John Boehner stepped down as Speaker of the House and, subsequently, left Congress. While there was significant disagreement between parties, there was also disagreement within the Republican party that did not seem to correspond with other bill types. These sources of variation lead to an important orthogonal dimension in the policy space. Likewise, it has long been known that Representatives from rural districts vote differently on bills related to

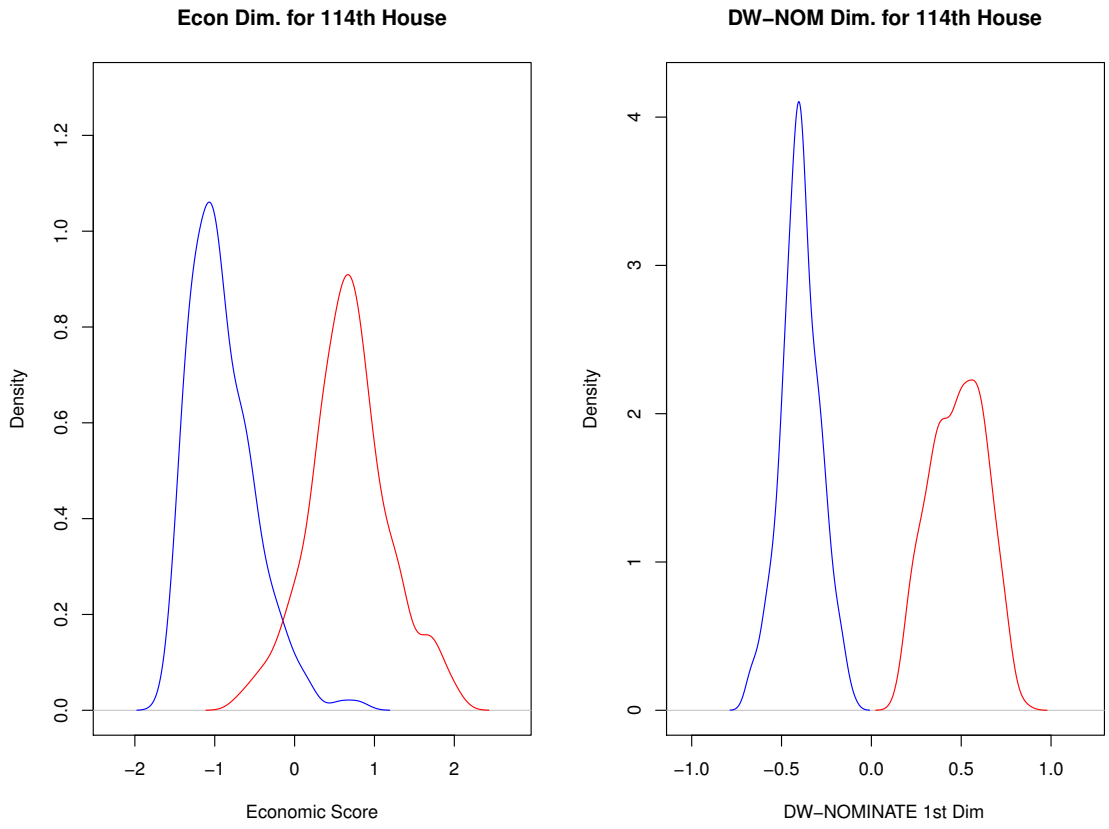

Figure 5: Comparison of First Dimension for CBPCFA and NOMINATE

infrastructure, agriculture, etc. All in all, this set of issue dimensions makes sense.

An important comparison to make is to the number of dimensions found by the NOMINATE procedure and by the beta-process procedure without clustering in McAlister (2018). Using DW-NOMINATE, one dimension is deemed to be necessary for the ideal point space for roll calls in the 114th House. Because of strong party teamsmanship, the amount of variance explained by the first NOMINATE dimension is quite high ($\approx$ 92% of the total variance). A similar amount of variance is explained by the first dimension of the beta process IRT procedure, but the procedure finds that there are four other dimensions that, in addition to the first, best explain the variance within the data. While the first dimension explains quite a bit of the unexplained variance, dimensions related to procedural votes, defense spending, rural/infrastructure issues, and social issues emerge as important in explaining voting behavior beyond party level voting. This is even more apparent in CBPCFA, which models the underlying group manifolds as separate effects and then models leftover variance as separate dimensions. Here, the same dimensions as BPIRT are present plus a dimension related to the budget. The economic dimension in CPBCFA is a combination of economic issues and group voting that is not accounted for in the group effects. This points to the fact that the clustering used by CBPCFA is capturing most of the group variance. However, there are still effects that are not captured that can be addressed by richer clus-

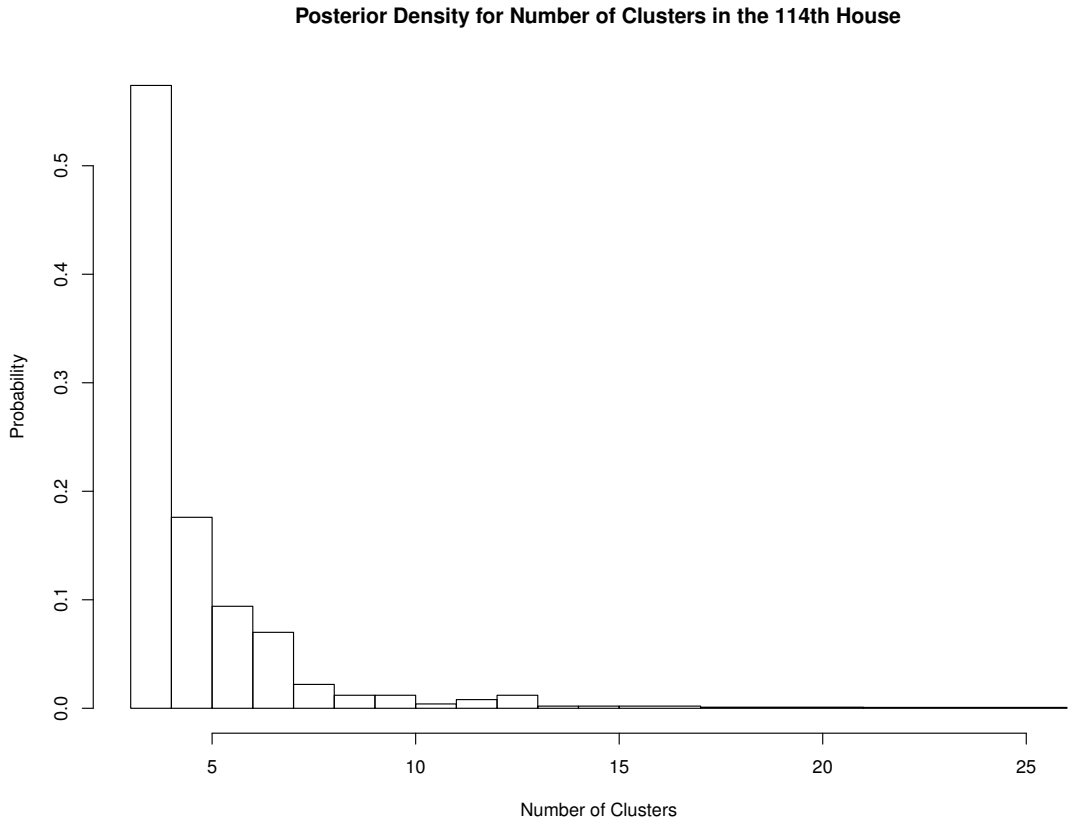

Figure 6: Number of Clusters Estimated by Dirichlet Process Priors using CBPCFA

tering routines, such as hierarchical clustering methods or by allowing for covariance among clusters in the CRP specification.

The first dimension corresponds to an omnibus economic dimension. While approximately 90% of bills load on this dimension, those with high absolute loadings correspond heavily with standard notions of domestic economic decision making in Congress (i.e. tax reform, program spend- ing, other distributive bills, etc.). While this dimension correlates with the first dimension from the DW-NOMINATE procedure (Correlation = .74), there is a key difference in how the scores within parties are distributed. Figure 5 shows the density of latent scores on the first dimension for the CBPCFA method and the density of the first dimension of DW- NOMINATE broken out by party. The difference can be seen at the center of latent space: CBPCFA shows that there is crossover between parties while DW-NOMINATE says that there is none. This finding can be explicitly linked to controlling for the group dynamics in the roll call scaling procedure, as explained below.

One of the main features of this new approach is that it takes into account group dynamics. Latent groupings within the data are estimated using the CRP priors discussed previously and the latent scores are estimated taking these groupings into consideration. To see this

effect, recall that the mean of the conditional posterior for ω_i is:

$$E[\omega_i|-] = V_i(\Lambda'(x_i + \alpha) + \mu_i'\Sigma_i^{-1}) \tag{7.1}$$

The expected value of ω_i is made up of two parts: the mean of the distribution implied by cluster g and the implied effect of certain votes on bills. When the clusters have low variance, there is little effect from the bills on the latent variable. In turn, this effects the mean of $\lambda_{j,k}$, driving it to 0. Thus, when the implied clusters are distinct and have low variance, $|\lambda_{j,k}|$ close to zero plays little role in the estimation process. However, when $|\lambda_{j,k}| >> 0$, the effect of the loading still affects the latent variable. Thus, values of the latent variable are a weighted average between the amount of variance that can be explained by being a member of a group and the effect of voting one way or another on a bill.

This effects implies a different interpretation of the loadings matrix when using CBPCFA. In standard roll call scaling techniques, large values of λ_k correspond to votes on bills which are highly discriminative for the values of the ideal points on that dimension. For example, a bill related to womens' rights yields votes that are highly determinative of the ideal point on a social dimension. These values are then used to determine the meaning of the latent dimension. CBPCFA creates a slightly different interpretation of the loadings - because Λ links to both the group effects and individual effects, high values of Λ correspond to bills which provide highly discriminative votes **beyond** the group effect. Thus, votes that occur strictly along group lines (mostly party lines throughout the history of the U.S. Congress), have high loadings while party line votes have lower loadings.

A model of the latent space that conditions informativeness of a vote on group dynamics is desirable - how much information does a straight party line vote really give about the latent preferences of an individual member of Congress? In the situation where most voting follows party lines and departures from the norm are random, this model reduces to a one-dimensional, party driven model. However, even in a polarized session like the 114th, there is enough structured departure from the cluster voting that multiple dimensions are estimated. In contrast, DW-NOMINATE and its cousins treat each vote as equally informative, *a priori*. This leads to extracting the group dynamics as the most important explainer of covariance when party line voting is frequent and results in the one dimensional conjecture, regardless of whether bipartisan votes have meaningful structure. Given recent work which posits that the role of the party largely determines a MCs vote, the results from this model shows that it is important to control for this kind of behavior when scaling roll call votes.

This effect can be explicitly linked back to the crossover of parties seen in Figure 5. Recall that the ideal point estimated by CBPCFA can be decomposed to two additive parts:

$$\omega_i = \xi_i + \tau_i \tag{7.2}$$

where τ_i is the group effect for voter i. Specific values of τ_i are correlated with group membership and, in turn, the errors for each individual vote. In the context of roll call scaling, this effect is omitted when not taken into account and leads to an omitted variable

	Democrats	Republicans
Cluster 1	0	1
Cluster 2	0	230
Cluster 3	188	9
Cluster 4	2	10
Cluster 5	0	1

Table 2: Cluster Assignment by Party for 114th U.S. House

	Economic	Defense	Rural	Procedural	Social	Budget
C2(Republicans)	.67	-.32	-.27	.14	.40	-.42
C3(Democrats)	-.79	.32	.41	0	-.62	.35
C4(Abstainers)	-.09	1.14	-1.79	-2.61	2.25	2.06

Table 3: Cluster Means For Each Dimension

	Economic	Defense	Rural	Procedural	Social	Budget
C2(Republicans)	.19	.40	.22	.07	.16	.19
C3(Democrats)	.48	.10	.28	.14	.28	.18
C4(Abstainers)	.42	1.72	2.89	1.55	1.25	2.09

Table 4: Cluster Variances For Each Dimension

bias for the estimate of the ideal point. This expectation of this bias can be quantified as:

$$Bias[\xi_i] = \frac{cov(\xi_i, \tau_i)}{var(\xi_i)} E[\tau_i] \tag{7.3}$$

Aldrich et al. (2014) shows that these two terms are positively correlated and the proofs above also show that this is the case. For this reason, we should expect that results for ξ_i from the standard model are biased away from zero. By properly accounting for this group effect, the ideal points estimated by CBPCFA are unbiased conditional on the model. This effect also causes the number of dimensions to be properly measured, with the CBPCFA model showing strong evidence that there is more than one dimension.

Along with the continuous measurement model that controls for group membership, CBPCFA also allows for inference about voting groups that exist within the Congress. Figure 6 shows the number of clusters imputed by the CBPCFA model. The 95% high posterior density interval for the number of clusters is between 3 and 6, with the majority of the density at 3 clusters. This is corroborated using the community detection approach for the adjacency table, which indicates that there are 5 clusters. Table 2 shows the cluster membership imputed by this model broken out by party. While most cluster membership follows party lines, there are two interesting observations. First, CBPCFA suggests that there are nine members of the Republican party in this session that align more with Democrats than standard Republican voting. Second, there is a third cluster with two Democrats and 10 Republicans that is distinct from the two main clusters.

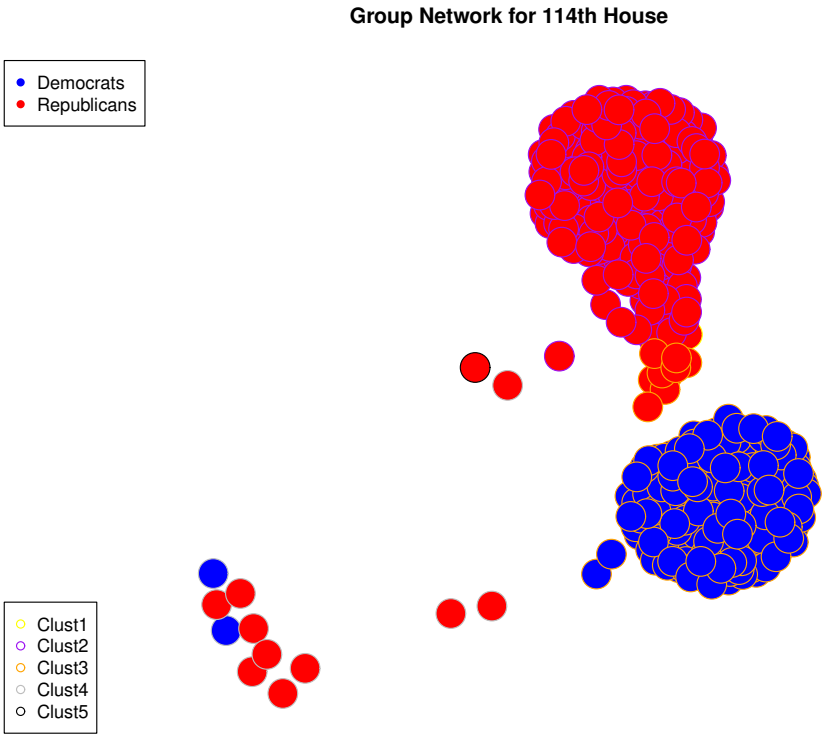

Figure 7: Implied Cluster and Vote Network from 114th U.S. House

For each of the clusters, means and variances are computed for each important dimension in the latent space. Table 3 and Table 4 show the posterior expectations for cluster means and variances, respectively.[1]

Cluster 4 has large variances on all dimensions except the first and a zero mean on the first. This implies that membership in fourth cluster is more complex than a systematic difference in voting compared to the two main clusters. Examining the raw roll calls reveals that this cluster includes 12 MCs who had a large proportion of abstentions, including resigning Speaker of the House John Boehner, Presidential candidate Paul Ryan, and numerous seats that were elected in special elections. Though Cluster 4 is not substantively interesting, it is worth noting that the clustering procedure separated these observations from the rest with no prior indication that anything was different.

Also of interest is to examine which Republicans fell into the Democrat cluster. Figure 7 shows the undirected network implied by the dyadic cluster relationships. While the Democrats are tightly connected and share very common voting records, Republicans have a little more variance and show some portion of members voting more closely to Democrats

[1]Cluster 1 and Cluster 5 only have one member each. These observations correspond to two MCs that have a large portion of missing values in their vote record. Overall, these two points are not substantively interesting.

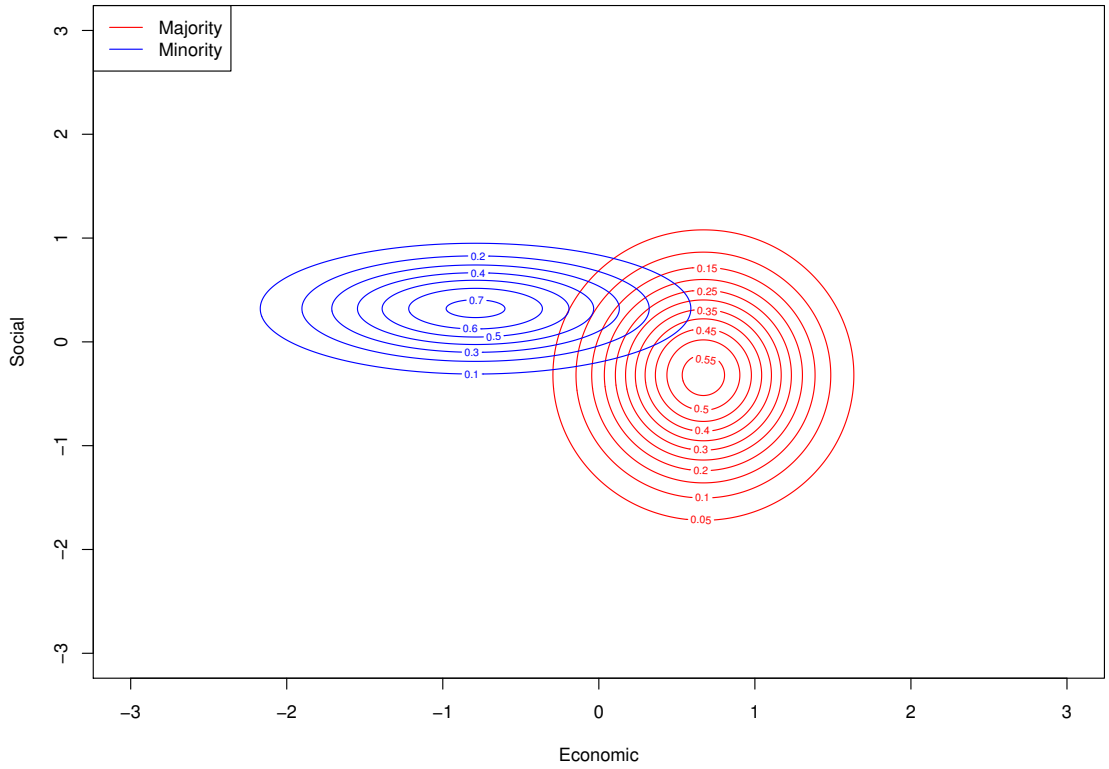

Figure 8: Distribution of Group Effects for the 114th U.S. House

than the median member of the Republican party. Of the more moderate voters, there is some portion that still gets lumped in with the Republican party. On the other hand, there is a threshold and these moderate Republicans crossover to the Democratic cluster. Examining the vote records of those across the threshold, there is still allegiance to the Republican party on straight party line votes, but on other bills there is a tendency to vote with the Democrats. Republican members of Congress that cluster with the Democrats come from typically more moderate districts, so this result is not surprising. However, using NOMINATE methodology, this conclusion would not be reached. This result shows one of the many advantages of using a clustered model like CBPCFA.

Using the measures estimated by CBPCFA, the implied group effects can be estimated. Using a post-processing procedure, τ_i is computed for the two main clusters. Figure 8 shows the density of the implied party effects for the economic and social dimensions. The Republican party was in the majority during this session and exhibited similar variances on both dimensions. However, the minority Democrats exhibit different behavior - a wider distribution of group effect on the economic dimension and a smaller variance on the social dimension. This outcome seems to point to issue level party control, in line with Aldrich and Rohde (2000) and Cox and McCubbins (2007). Unlike NOMINATE and other standard roll call scaling techniques, CBPCFA allows for examination of group effects within the ideal

points. This technique is particularly promising with regards to testing theories of party control in Congress. When previous studies have used NOMINATE scores to test these theories, the assumptions of the NOMINATE model are misaligned with the theories being tested. CBPCFA provides a flexible tool for testing these theories, allowing party control to be learned from the model rather than assuming, *a priori*, which theory is true. This flexible approach is a key reason that CBPCFA is a meaningful addition to the literature on legislative behavior.

One interpretation of the implied group effects is that they correspond to a probability density of approving of a policy at a specific location within the policy space. This to say that the group effects provide a way to think of a spatial voting game where there are only G players rather than N players. This would correspond to a game where there exist G strong groups that are cohesive voting blocs. In social choice theory, this game has a strong link to the *uncovered set*, the heart of the spatial model, and the core of the voting game (Miller, 1980; Cox, 1987; Schofield, 1999). Given the spatial model at the core of the roll call voting model, the group effects correspond to the probabilities that the groups will approve an alternative, given a location for the status quo. While these quantities are not identifiable under the IRT specification of the voting model, high probability areas of approval can be found using the estimated group effect densities - what is the location in the voting space that maximizes the probability that all groups will approve of an alternative? Formally:

$$\omega^* = \underset{\omega^* \in \mathbb{R}^K}{\arg\max} \prod_{h=1}^{G} P(\omega^*; \mu_g, \Sigma_h) \tag{7.4}$$

Under the assumption that the alternatives are only proposed on the economic and social dimensions, Figure 8 shows this region in the intersection of the two density plots. Along with further work that attempts to place the alternative and status quo associated with a vote decision in the same ideal point space, CBPCFA provides a way to empirically test concepts related to social choice theory.

8 Conclusion

Understanding the impact of groups on voting behavior within legislatures is an important concept in the literature. Previous approaches have had to rely on measures that come from sources related to the actual votes, themselves. This paper derives a new approach to estimating group level effects within roll call data that utilizes a flexible nonparamteric approach to model estimation that allows for both estimation of groups within the data and the resulting dimensions of the ideal point space. CBPCFA sheds many of the restrictive assumptions that are needed for identification in other scaling procedures and presents a method with less restrictive assumptions that allow for proper testing of theories related to group effects within voting bodies. I apply this approach to the 114th session of the U.S. House and show the rich inference that can be made about roll call data using this model.

This model is ripe for extension. From a statistical perspective, the clustering model used is a simple one where each person belongs exclusively to one cluster. In reality, groups and voting are hierarchical in nature and there are many varying influences which influence the vote of a member of Congress. One approach that can be used to capture these more complex effects is hierarchical clustering, which allows each observation to belong to a hierarchy of groups. There are multiple approaches that are promising, but work on Bayesian rose trees provides an approach that does not require binary leaves within the hierarchy (Blundell et al., 2012). This approach promises to provide hierarchies within the U.S. Congress which are richer than simply party membership.

This model can also be applied across different periods of the U.S. Congress to understand the changes in the role of groups over the course of Congress. Numerous theories have posited that Congress is polarizing over time and this results in less bipartisan compromise in passed legislation (McCarty et al., 2016; Harbridge, 2015). Given that this method can extract group level effects and potentially models differences in party control on different dimensions of votes within the U.S. Congress, there are many interesting substantive conclusions which can be examined. On a similar note, allowing this method to search for geospatial dependencies can answer question related to distributive spending in the U.S. Congress. One drawback of this approach is that it requires a relatively large number of voters and bills in order to make meaningful inferences about the data. Advances in stick-breaking constructions of the beta process and Dirichlet process provide methods for exact inference which promise to get around the large sample requirements for these methods to work.

References

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

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

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

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

Bafumi, J. and Herron, M. C. (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., Dunson, D. B., 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.

Blackwell, D. and MacQueen, J. B. (1973). Ferguson distributions via pólya urn schemes. *The annals of statistics*, pages 353–355.

Blundell, C., Teh, Y. W., and Heller, K. A. (2012). Bayesian rose trees. *arXiv preprint arXiv:1203.3468*.

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

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

Cox, G. W. (1987). The uncovered set and the core. *American Journal of Political Science*, pages 408–422.

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

Cox, G. W. and McCubbins, M. D. (2007). *Legislative leviathan: Party government in the House*. Cambridge University Press.

Cox, G. W. and Poole, K. T. (2002). On measuring partisanship in roll-call voting: The us house of representatives, 1877-1999. *American Journal of Political Science*, pages 477–489.

De Finetti, B. (1992). Foresight: Its logical laws, its subjective sources. In *Breakthroughs in statistics*, pages 134–174. Springer.

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

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

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the american statistical association*, 90(430):577–588.

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

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

Gruhl, J., Erosheva, E. A., and Crane, P. K. (2013). A semiparametric approach to mixed outcome latent variable models: Estimating the association between cognition and regional brain volumes. *The Annals of Applied Statistics*, pages 2361–2383.

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

Heckman, J. J. and Snyder Jr, J. M. (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.

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.

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

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

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

Krehbiel, K. (2010). *Pivotal politics: A theory of US lawmaking*. University of Chicago Press.

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

McAlister, K. (2018). *Disagreement and Dimensionality: A Beta Process Approach to Roll Call Scaling for the U.S. Congress*.

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

Miller, N. R. (1980). A new solution set for tournaments and majority voting: Further graph-theoretical approaches to the theory of voting. *American Journal of Political Science*, pages 68–96.

Orman, G. K., Labatut, V., and Cherifi, H. (2011). On accuracy of community structure discovery algorithms. *arXiv preprint arXiv:1112.4134*.

Pitman, J. et al. (2002). Combinatorial stochastic processes.

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

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

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

Rai, P. and Daumé, H. (2009). The infinite hierarchical factor regression model. In *Advances in Neural Information Processing Systems*, pages 1321–1328.

Rasmussen, C. E. (2000). The infinite gaussian mixture model. In *Advances in neural information processing systems*, pages 554–560.

Rohde, D. W. (2010). *Parties and leaders in the postreform House*. University of Chicago Press.

Schofield, N. (1999). The heart and the uncovered set. *Mathematical Utility Theory*, pages 79–113.

Spirling, A. and Quinn, K. (2010). Identifying intraparty voting blocs in the uk house of commons. *Journal of the American Statistical Association*, 105(490):447–457.