UNIVERSITY OF CALIFORNIA, SAN DIEGO

Measuring Ideology, Dimensionality and Polarization in Politics

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Political Science

by

Yunkyu Sohn

Committee in charge:

Professor James Fowler, Chair
Professor Gary Jacobson
Professor Thad Kousser
Professor James Rauch
Professor Yiqing Xu

2017

The dissertation of Yunkyu Sohn is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2017

DEDICATION

To my father, mother and Rooni.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude to James Fowler for guiding me throughout my intellectual journey in the discipline of political science. I would also like to thank Gary Jacobson, Thad Kousser, James Rauch and Yiqing Xu for their guidance as committee members. Finally, I would like to acknowledge the value of inputs from Jong Hee Park, Kosuke Imai and his research group members, the organizers and participants of Princeton QSS colloquium and the members of HNG and UCSD methodology and American politics seminars.

Chapters 1, 2 and 3 are in preparation for submission for publication. Sohn, Yunkyu. The dissertation author was the sole researcher and author of this material.

VITA

2017	Ph. D. in Political Science, University of California, San Diego
2008	M. S. in Bio and Brain Engineering, Korea Advanced Institute of Science and Technology
2007	B. A. in Philosophy, Seoul National University
2007	B. S. in Physics, Seoul National University

PUBLICATIONS

Yunkyu Sohn and Jong Hee Park. 2017. Bayesian Approach to Multilayer Stochastic Blockmodel and Network Changepoint Detection. *Network Science* 5(2): 164-186.

Rooni Lee, Yunkyu Sohn and Wonjae Lee. 2017. Structuralizing the Fluxus Way of Life: the Social Network of Fluxus. *Leornardo* 50(1): 74-75.

Jaimie Y. Park, Yunkyu Sohn and Sue Moon. 2016. Power of Earned Advertising on Social Network Services: A Case Study of Friend Tagging on Facebook. *Proceedings of the 10th International Conference on Web and Social Media*. 299-308.

Yunkyu Sohn. 2014. Strategic Foundation of Computational Social Science. *Proceedings of the 23rd International Conference on World Wide Web*. 51-56.

Lorenzo Coviello, Yunkyu Sohn, Adam D. I. Kramer, Cameron Marlow, Massimo Franceschetti1, Nicholas A. Christakis and James H. Fowler. 2014. Detecting Emotional Contagion in Massive Social Networks. *PLoS ONE* 9(3): e90315.

Julia Lee, Yunkyu Sohn and James H. Fowler. 2013. Emotion Regulation as the Foundation of Political Attitudes. *PLoS ONE* 8(12): e83143.

Yunkyu Sohn, Mungkyu Choi, Yong-Yeol Ahn, Junho Lee and Jaeseung Jeong. 2011. Topological Cluster Analysis Reveals the Systemic Organization of *Caenorhabditis elegans* Connectome. *PLoS Computational Biology* 7(5): e1001139.

Yunkyu Sohn. 2006. Evolutionary Biology and Mental Functionalism. *Philosophical Forum* 34: 171-197.

ABSTRACT OF THE DISSERTATION

Measuring Ideology, Dimensionality and Polarization in Politics

by

Yunkyu Sohn

Doctor of Philosophy in Political Science

University of California, San Diego, 2017

Professor James Fowler, Chair

This dissertation introduces a set of new statistical methods for measuring foundational constructs in political science: ideology, dimensionality and polarization. Using the proposed methods, I offer novel findings on multidimensional ideological characteristics of American Congress. The final chapter provides the first complete description of ideological coalitions in Korean National Assembly from its birth to the present by introducing a new structural model of ideal point estimation for non-voting datasets.

Chapter 1

Identification and Estimation for Multidimensional Item Response Theory: An Analysis of Roll Call Votes in the United States Congress

Statistical estimation of legislators' ideological preferences is at the core of empirical political science. A long standing question in American legislative history is whether or not single dimensional preferences can explain legislators' voting behavior. Existing multidimensional item response theory models—originally developed in the educational testing literature and commonly used for the estimation of preferences—suffer from the lack of identifiability due to rotational and scale invariance. We propose a new identification strategy that constrains ability and item discrimination parameters to form orthonormal matrices. For estimation, we develop a Gibbs sampling algorithm using Matrix von Mises-Fisher distribution. The resulting estimates exhibit complete orthogonality among dimension-specific estimates and align on maximum variance (*i.e.* principal) directions. Our analysis of

the historical roll call voting records shows that the number of ideological dimensions in the House of Representatives is greater than previously believed throughout the 20th century.

1.1 Introduction

Statistical estimation of legislators' ideological preferences is at the core of empirical political science. Item response theory (IRT) is the canonical statistical model used for ideal point estimation. Originally developed for identifying test takers' ability and exam questions' difficulty from educational testing datasets (Bock & Lieberman 1970, Fox 2010), IRT and its variants have been used in all fields of political science, ranging from American politics, comparative politics to international relations (see Clinton (2012) for review of the applications). Following the pioneering models of ideal point estimation using maximum likelihood (Poole & Rosenthal 1991) and Bayesian approach (Jackman 2001, Clinton, Jackman & Rivers 2004), a variety of methods have been developed, including dynamic ideal point estimation (Poole & Rosenthal 2000, Martin & Quinn 2002), common scale ideal estimation of various political actors (Bailey 2007, Shor & McCarty 2011, Lo, Proksch & Gschwend 2014), text-based approaches (Slapin & Proksch 2008, Lowe, Benoit, Mikhaylov & Laver 2011, Clark & Lauderdale 2010), citizen-candidate estimation method using donation datasets (Bonica 2014) and variational approximation methods (Imai, Lo & Olmsted 2016).

A long standing question in American legislative history is whether or not single dimensional preferences can explain legislators' voting behavior in Congress. Despite the prevalent use of IRT in ideal point estimation, existing methods have largely focused on unidimensional ideal point estimation, leaving idealogical conflicts beyond the party-line dimension unexplained (Sin 2014, Bateman & Lapinski 2016). Informal tools that are not directly linked to the generative parameters of voting have been used to select the number of dimensions. For example, in-sample measures of fit, which always increases with the

number of dimensions, and subjective visual inspection on eigenvalue change of covariance matrices (Poole & Rosenthal 2000, McCarty, Poole & Rosenthal 2016) are used to justify low dimensionality. In fact, multidimensional ideal point estimation is necessary in the major cases of parliamentary voting analysis. When analyzing multi-coalitions literatures, for instance, Hix, Noury & Roland (2006) find that two dimensional ideological space of the European parliament only accounts for about 60% in the prediction of cut lines which divide supporters and objectors. By looking at bill subsets, Crespin & Rohde (2010) find that the number of ideological dimensions required for describing the patterns of voting on appropriation bills in the United States (U.S.) House may exceed three.

A fundamental statistical problem that comes with the use of IRT for multidimensional ideal point estimation is the lack of identification in existing approaches. Namely, multidimensional IRT (MIRT) and other probabilistic latent linear models in general (Tipping & Bishop 1999, Minka 2000, Murphy 2012) suffer from arbitrary scaling, rotation and additive aliasing of estimates (Jackman 2001, Rivers 2003, Bafumi, Gelman, Park & Kaplan 2005). Existing approaches of identification have imposed either an insufficient number of constraints or a constraint that requires researchers' discretion to select values for restriction. Although orthonormal constraints have been believed to be the most intuitive solution for identification (Tipping & Bishop 1999, Murphy 2012), implementing this idea into an actual estimation algorithm without the use of an ex-post processing has been difficult. Particularly for developing a Bayesian method, it has been difficult due to the lack of matrix distributions whose support satisfy the orthonormal property.

In order to resolve the identification problem, we introduce a new identification strategy that constrains both individual and item discrimination parameter matrices to form orthonormal matrices, satisfying the exact condition for identification. This strategy gives rise to dimension-specific estimates that are located along maximum variance directions with zero redundancy. Also, it does not require researchers' discretion for choosing identification

constraints. This is in a stark contrast with other existing approaches based on auxiliary standardization of dimension-specific estimates with an insufficient number of constraints for identification (Jackman 2009), triangular constraint with an arbitrary determination of the first dimensional values (Lopes & West 2004) and fixing $K+1$ legislators for K dimensional MIRT model (Clinton, Jackman & Rivers 2004, Rivers 2003), which is subject to an undesired rotation and scaling of estimates.

After introducing the identification strategy, we propose a hierarchical Gibbs sampling algorithm using Matrix von Mises-Fisher distribution (MvMF), which is a uniform matrix distribution defined over orthonormal matrix set of an arbitrary size (Chikuse 2012, Hoff 2009, Hoff 2007). The proposed method automatically incorporates the identification constraints into estimation. Hence *ex post* processing for identification or normalization is not necessary. Due to orthonormality, each pair of dimension-specific estimates are uncorrelated with each other (*i.e.* minimum redundancy property), and all of them hold a common scale (*i.e.* unit scale property). The resulting estimates achieve the minimum redundancy property among dimension-specific estimates due to orthogonality. This property also yields each dimensional estimates of the item parameter matrix to lie on the maximum variance (*i.e.* principal) directions, different from existing methods that produce arbitrary estimates lying on the K dimensional principal subspace. Simulation results demonstrate superior recovery performance of true parameters and dimension-specific weights of the proposed method compared to IDEAL (Albert 1992, Clinton, Jackman & Rivers 2004) and W-NOMINATE (Poole & Rosenthal 1991, Poole 2005).

We apply the proposed method to study multidimensional cleavage structure of the U.S. House of Representatives roll call voting dataset, which is one of the most widely studied datasets in American politics research. Due to the aforementioned problems of the existing IRT methods, full recovery of the historical cleavage structures has been difficult. The proposed methods provide multidimensional estimates of ideal points in the

U.S. House and quantitative evidence on the existence of intra-party factions in Republican and Democratic parties during the 20th century. In contrast to previous beliefs on the single dimensional dominance in the U.S. House, with an exception of the souther realignment period (Poole & Rosenthal 2000, McCarty, Poole & Rosenthal 2016), the proposed method identifies 3 dimensions that are associated with historical events, with substantial level of importance in explaining the patterns of voting in the roll call dataset.

1.2 Application: The United States House Roll Call

The U.S. House of representatives roll call dataset is an ideal dataset to understand legislators' behavior on the primary process of the U.S. Congress. Since Lowell (1902), this dataset has been a central workhorse for leveraging information for the impact of electoral and institutional factors on policy making (Bartels 1991, Cox & McCubbins 2005, Krehbiel 2010).

The most ubiquitous statistical measure used in the majority of studies utilizing the roll call dataset is ideal points, which are projected scores of legislators' voting patterns on a low dimensional ideological space. A typical voting matrix, consisting of hundreds of legislators and thousands of roll calls, is usually compressed into ideal points and bill-related parameters of 1 or 2 dimensions. This approximation draws on the psychological argument of Philip Converse, who states that human belief systems are configured by ideas and attitudes that are bound together by constraints and interdependence (Converse 1962). Since such constraints and interdependence diminish the set of feasible outcomes, the number of parameters necessary for describing voting behavior may be dramatically reduced compared to the size of the original dataset.

An imminent question for estimating the lower dimensional ideal points from the roll call dataset is the number of dimensions required to summarize the voting patterns. This is not merely a statistical question. In fact, dimensionality is a key to answering a well-

known debate on the patterns of voting in the Congress. Notably, dimensionality is highly correlated with the role of parties and factions on influencing individual vote choices (Cox & McCubbins 2005, Krehbiel 2010, McCarty 2001). In a two-party legislature, as in the U.S. Congress, when political parties play a significant role in maintaining cohesive voting blocs, cut points of the votes would be located around a single point, dividing each set of legislators on the left and the right by their party affiliation. In contrast, when there are a set of bills that are unexplained by party-line votes, researchers would need another dimension to represent the patterns of voting that cannot be fully summarized by the party-line dimension.

In this regard, dimensionality can be used to investigate the varying strength of the parties and the existence of extra-party voting coalitions over the history of the Congress. An interesting case is the rise and fall of intra-party factions within the Democratic and Republican parties over the 20th century. The existence of ideologically distinct factions with particular policy goals has been constantly suggested in the historical studies of the U.S. Congress (Bateman & Lapinski 2016, Sundquist 1983, Sin 2014). However, quantitative evidence on the existence of such factions, whose preference cuts across the primary party-line dimension, has been sparse with the exception of the civil rights era during the mid-20th century (McCarty, Poole & Rosenthal 2016, Poole & Rosenthal 2000).

The existing belief on the robustness of single or, at most, two dimensional space to describe the patterns of voting originates from the use of informal approaches for dimensionality selection (Poole & Rosenthal 2000, Poole 2005). For instance, the widely used in-sample measure of fit, average proportional reduction of error, always increases with the number of dimensions, and does not incorporate out-of-sample penalty adjustment to prevent over-fitting. Another popular method for dimensionality selection, detecting the elbow of the eigenvalue trend of legislators' voting profile covariance matrix, does not incorporate parameters in generative model-based approaches. Both approaches rely heavily on visual inspection or subjective judgment of the changes in the values over the number of

dimensions, making a principled dimensionality selection incapable.

In this article, we introduce a fully identified model of MIRT that can be used to estimate multidimensional ideal points of the members in the U.S. House. Compared to its Bayesian ancestor IDEAL (Albert 1992, Clinton, Jackman & Rivers 2004), the proposed method produces reliable samples with superior recovery rate. Also the proposed method offers an interpretable weight parameter as opposed to its deterministic counterpart W-NOMINATE (Poole & Rosenthal 1991, Poole, Lewis, Lo & Carroll 2008) whose weight estimate does not convey useful information. In particular, we focus on the analysis of the House roll call dataset ranging from the 46th to 113th sessions. The number of legislators, by including all individuals who ever participated in each session, spanned from 301 to 458 with the median value of 441, and the number of roll calls spanned from 72 to 1,865 with the median value of 309 (Poole 2016). While the Congress has been largely characterized as a two party legislature during this period, the proposed method identifies the second and the third dimensions that are associated with intra-party factions throughout the majority of period analyzed.

1.3 Multidimensional Item Response Theory Model

The goal of the original IRT model is to infer latent item discrimination parameter vector $\tilde{\boldsymbol{\beta}}_j \equiv \mathbf{w} \cdot \boldsymbol{\beta}_j \in \mathbb{R}^K$, latent item difficulty parameter vector $\boldsymbol{\alpha}_j \in \mathbb{R}^K$ and latent individual ability vector $\mathbf{x}_i \in \mathbb{R}^K \ \forall i, j$ in an educational testing context (Bock & Lieberman 1970, Reckase 2009, Fox 2010, De Ayala 2013). Here, dimension importance vector $\mathbf{w} \in \mathbb{R}^K$, whose element is arranged in a decreasing order (*i.e.* $w_1 > w_2 > \cdots > w_K$), is used to define $\tilde{\boldsymbol{\beta}}_j$ for the consistency of the expression with the newly proposed formulation. Suppose a binary response variable $Y_{ij} \in \{0, 1\}$ for individual $i \in \{1, \ldots, N\}$ and item $j \in \{1, \ldots, M\}$. The following model is the most well-known representation of multidimensional IRT, called

2 parameter normal ogive model:

$$\Pr(Y_{ij}=1|\theta_i,\alpha_i,\beta_j,\gamma_j) \;=\; \Phi\left(\sum_k w_k\beta_{jk}x_{ik}-\alpha_j\right)=\Phi\left(\sum_k \tilde{\beta}_{jk}x_{ik}-\alpha_j\right), \quad (1.1)$$

where $Y_{ij}=1$ if i gave a correct answer to item j, $Y_{ij}=0$ when incorrect, and $\Phi(\cdot)$ is the cumulative standard normal distribution function (Albert 1992). The model aims to obtain high α_j for difficult items, high $\mathbf{x}_i$ for individuals with superior performance, and high $\tilde{\boldsymbol{\beta}}_j$ for items with the correct answer rate approaching 50%, the highest level of variability.

In fact, Eq. 1.1 directly corresponds to the utility calculation in voting (Clinton, Jackman & Rivers 2004). Assume that each legislator achieves a maximum level of satisfaction when the spatial characteristic of an item (*i.e.* roll call) equals her ideal point, and her utility decreases with their distance. By assuming a quadratic loss (*i.e.* utility) function, for a legislator $i \in \{1,\ldots,N\}$, who makes a choice between legislative items $\mathbf{b}_j \in \mathbb{R}^{1\times K}$ (*i.e.* bill) and $\mathbf{s}_j \in \mathbb{R}^{1\times K}$ (*i.e.* status quo) for roll call $j \in \{1,\ldots,M\}$ with i's most preferred position $\mathbf{x}_i \in \mathbb{R}^{1\times K}$, the chance that i will choose $\mathbf{b}_j$ is determined by the following reasoning.

$$\begin{aligned}\Pr(i \text{ chooses } \mathbf{b}_j) &= \Pr\left(U_i(\mathbf{b}_j) > U_i(\mathbf{s}_j)\right)\\ &= \Pr\left(-\|\mathbf{x}_i-\mathbf{b}_j\|_{\mathbf{w}}^2+\|\mathbf{x}_i-\mathbf{s}_j\|_{\mathbf{w}}^2>\varepsilon_{ij}\right)\\ &= \Phi\left(\sum_k w_k\beta_{jk}x_{ik}-\alpha_j\right), \end{aligned} \quad (1.2)$$

where $\boldsymbol{\beta}_j \equiv 2(\mathbf{b}_j-\mathbf{s}_j)$, $\alpha_j \equiv \{\mathbf{w}\circ(\mathbf{b}_j+\mathbf{s}_j)\}\cdot(\mathbf{b}_j-\mathbf{s}_j)$ ($\circ$: entrywise product; $\cdot$: inner product), a squared weighted Frobenius norm of a vector $\mathbf{z}$ with a weight parameter vector $\mathbf{w}$, $\|\mathbf{z}\|_{\mathbf{w}}^2 \equiv \sum_k w_k z_k^2$ (*i.e.* squared Mahalanobis distance of $\mathbf{z}$ with the diagonal precision matrix $\mathrm{diag}(\mathbf{w})$). We use a cumulative normal function for $\Phi(\cdot)$ (*i.e.* 2 parameter normal ogive model)(Albert 1992, Clinton, Jackman & Rivers 2004), corresponding to the shape of the disturbance term (here, $\varepsilon_{ij} \sim \mathcal{N}(0,1/\phi)$).

One can rearrange Eq. 1.2 in matrix notations for the observed binary response

matrix $\mathbf{Y}$ and the unobserved probability measure matrix $\mathbf{Y}^*$:

$$\begin{aligned} \mathbf{Y}^* &= \Pr(\mathbf{X}\mathbf{W}\mathbf{B}^\top - \mathbf{A} > \mathbf{E}) \\ &= \Phi(\mathbf{X}\mathbf{W}\mathbf{B}^\top - \mathbf{A}). \end{aligned} \tag{1.3}$$

We define the augmented variable matrix $\mathbf{Z} \equiv \Phi^{-1}(\mathbf{Y}^*)$, $N \times K$ matrix $\mathbf{X} \equiv \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_N \end{pmatrix}$, $M \times K$ matrix $\mathbf{B} \equiv \begin{pmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_M \end{pmatrix}$, $K \times K$ diagonal matrix $\mathbf{W} \equiv \mathrm{diag}(\mathbf{w})$, $N \times M$ matrix $\mathbf{A} \equiv \mathbf{1}_{N\times 1}\boldsymbol{\alpha}$, where $\boldsymbol{\alpha} \equiv (\alpha_1, \cdots, \alpha_N)$ and $\mathbf{E}$ is an $N \times M$ independent and identically distributed Gaussian error matrix with variance $1/\phi$.[1]

With suitable constraints on the scales of the item discrimination parameter ($\mathbf{B}$) and the ability parameter ($\mathbf{X}$) and the correlation among their dimension-specific estimates, the dimension weight estimate $\mathbf{w}$ ($\mathbf{W}$) becomes an informative indicator that can be used to understand the importance of each dimension for explaining the overall structure of a roll call dataset. When $\mathbf{X}$ and $\mathbf{B}$ do not have a constant scale and their dimension-specific estimates are correlated, it becomes infeasible to make claims on the explanatory power of each dimension.

As we will see shortly, the proposed method imposes orthonormality constraints on $\mathbf{X}$ and $\mathbf{B}$ to obtain the interpretability of the weight parameter. In contrast, the canonical Bayesian MIRT model (Albert 1992, Jackman 2001, Clinton, Jackman & Rivers 2004), IDEAL, does not have an explicit weight parameter in its formulation. Since IDEAL lacks orthogonal restriction across dimension-specific estimates, *ex-post* interpretation of dimension-specific estimates is also not feasible. While W-NOMINATE contains the weight

[1]For convenience, we respectively denote X_k (*i.e.* $\mathbf{X}_{[,k]}$) and B_k (*i.e.* $\mathbf{B}_{[,k]}$) as kth-dimension-specific estimates (*i.e.* column vectors) of $\mathbf{X}$ and $\mathbf{B}$.

parameter in its normal-normal formulation (Poole & Rosenthal 1991, Poole 2005), it becomes difficult to interpret the meaning of the dimension weight parameter due to its non-orthonormal formulation and non-standard rescaling process of the resulting estimates. In the latter section, we support this claim through synthetic examples.

1.4 Identification

A statistical model $f(\mathbf{y}|\boldsymbol{\theta})$ is identified when there is one-to-one correspondence between the set of parameters $\boldsymbol{\theta}$ and the probability distributions of observed data $\mathbf{y}$ (Koopmans 1949). In other words, for an identified model, $f(\mathbf{y}|\boldsymbol{\theta}) = f(\mathbf{y}|\boldsymbol{\theta}')\ \forall \mathbf{y}$ if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}'$. It has been well-known that the naive latent linear model including item response theory model, without appropriate constraints, is not identifiable (Tipping & Bishop 1999, Rivers 2003, Bafumi et al. 2005, Murphy 2012).

Previous literature provides specific invariance conditions for the likelihood function, or $\mathbf{XWB}^\top - \mathbf{A}$, that lead to arbitrary transformation of estimates (Jackman 2001, Bafumi et al. 2005). Namely, three types of transformations exist: rotation, addition and scaling.

1. **Rotation** Suppose an arbitrary orthonormal rotation matrix $\mathbf{R}$ of size $K \times K$. Then $\mathbf{XWB}^\top = (\mathbf{XR})\mathbf{W}(\mathbf{BR})^\top$ since $\mathbf{RR}^\top = \mathbf{I}$ by definition. Hence the one-to-one correspondence condition is not satisfied.

2. **Addition** Suppose an addition of the matrix of ones $\mathbf{1}_{N\times K}$ to $\mathbf{X}$, then $\mathbf{XWB}^\top - \mathbf{A} = (\mathbf{X} + \mathbf{1}_{N\times K})\mathbf{WB}^\top - (\mathbf{A} + \mathbf{1}_{N\times K}\mathbf{WB}^\top)$. The form of 2 parameter MIRT is preserved and the resulting value is the same, but some parameters exhibit different values.

3. **Scaling** Without a scale constraint, such as unit norm, $\mathbf{XWB}^\top - \mathbf{A} = (c\mathbf{X})\mathbf{W}(\mathbf{B}/c)^\top - \mathbf{A}$, dividing a parameter with an arbitrary constant c can result the same output with the multiplication of the other parameter in the factorized term.

A series of studies provide important results regarding local identification of latent linear models (Rothenberg 1971, Magnus & Neudecker 1988), where local identification indicates the identification of true parameters up to the sign reversal. On the basis of these results, Corrollary 2 of Rivers (2003) demonstrates a necessary condition ensuring the identification for the parameters of a MIRT model. The Corrollary states that there must be $R \geq K(K+1)$ number of constraints in order to determine unknown parameters in a set of simultaneous equations resulting from an IRT formulation.[2]

In ideal point estimation, the most well-known treatment satisfying this criterion is to fix the ideological locations of $K+1$ legislators in K-dimensional space, the so-called Kennedy-Helms restriction (Rivers 2003, Clinton, Jackman & Rivers 2004). While it is intuitive to apply this condition in unidimensional space, so as to fix the locations of a liberal legislator and a conservative legislator, however, it is not difficult to recognize the weakness of this treatment in $(K > 1)$-dimensional space. If we ignore the sign flip of ideal points, corresponding to the condition for local identification, there are ${}^{K+1}P_{K+1}/2$ combinations among the possible orderings of the $K+1$ legislators for each dimension, where ${}^{K+1}P_{K+1} = (K+1)!$ is $K+1$ permutation of $K+1$. Fixing the locations of these legislators *ex ante* is to arbitrarily choose a specific ordering among the potential ${}^{K+1}P_{K+1}/2$ orderings for each dimension (*e.g.* $12^3 = 1,728$ combinations in total when $K = 3$) without knowing the substantive meaning of each dimension. As a result, it is not feasible to apply this approach in most cases in which one does not have strong auxiliary evidence to do so.

In contrast, our approach satisfies the identification condition without imposing such arbitrary constraint. We impose orthonormality restrictions to both $\mathbf{X}$ and $\mathbf{B}$. Due to the introduction of the weight parameter $\mathbf{W}$, we define $\widetilde{\mathbf{B}} \equiv \mathbf{BW}$ to have the conventional bilinear formulation (*i.e* $\mathbf{Y}^* = \Phi(\mathbf{X}\widetilde{\mathbf{B}}^\top - \mathbf{A})$ for Eq. 1.3), ignoring the normality condition for $\mathbf{B}$. The zero mean condition for each dimensional estimate of $\mathbf{B}$ is obtained due to

[2]While the original corollary discusses a special case in which the constraints are given to legislator ideal point matrix $\mathbf{X}$, in the appendix, we show that the same result holds when the condition includes constraints on

the intercept regression (column centering) formulation of the Gibbs sampling algorithm introduced below. In fact, this condition leads to item centering, a well-known constraint in IRT literature (De Ayala 2013).

1. Number of constraints for individual parameters $\mathbf{X}$ for $k, l \in \{1, \ldots, K\}$
   - Unit Euclidean norm condition ($\sum_i X_{ik}^2 = 1$): K
   - Orthogonality condition ($\sum_{i,k \neq l} X_{ik} X_{il} = 0$): $K(K-1)/2$
2. Number of constraints for bill parameters $\mathbf{B}$ ($\widetilde{\mathbf{B}}$) for $k, l \in \{1, \ldots, K\}$
   - Orthogonality condition ($\sum_{i,k \neq l} B_{jk} B_{jl} = 0$): $K(K-1)/2$
   - Zero mean condition ($\sum_j B_{jk} = 0$): K

Theorem 1 *Constraining* $\mathbf{X}$ *and* $\mathbf{B}$ *to form orthonormal matrices and centering* $\mathbf{B}$ *locally identifies the multidimensional item response theory model.*

Proof is provided in the appendix. By summing up the four constraints, we obtain the number of constraints $R = K(K+1)$. The normality conditions for both parameters identify the scales of these parameters. The orthogonality condition resolves the rotational ambiguity of the estimates. Finally, the zero mean condition, or the item centering condition, resolves the additive aliasing.

In addition to the identification benefit, the orthonormal constraints enable an intuitive interpretation of the weight vector $\mathbf{w}$ as dimension weights. The squared element-wise Frobenius norm of the mean matrix $\|\mathbf{X}\mathbf{W}\mathbf{B}^\top\|^2 = \text{trace}((\mathbf{X}\mathbf{W}\mathbf{B}^\top)^\top \mathbf{X}\mathbf{W}\mathbf{B}^\top) = \sum_k w_k^2$ and the amount of variance explained by each dimension are fully represented by $\mathbf{w}$. This is in a direct accordance with the weight parameters in principal component analysis which yields a diagonal covariance structure among dimension-specific estimates.

---

both $\mathbf{X}$ and $\mathbf{B}$.

1.5 Bayesian Estimation

1.5.1 Sampling Orthonormal Matrix from Matrix von Mises-Fisher Distribution

An immediate difficulty, which arises from the application of the proposed identification strategy to parametric estimation models, is the necessity to sample orthonormal matrices for $\mathbf{X}$ and $\mathbf{B}$. This distribution is not derivable from standard distributions such as the multivariate normal. Each dimension-specific set of estimates must have a unit Euclidean norm and the inner product for every pair of dimension-specific estimates needs to be zero.

Matrix von Mises-Fisher Distribution (MvMF) is a matrix distribution which satisfies the orthonormal constraint (Chikuse 2012, Hoff 2007, Hoff 2009). Samples of MvMF lie on Stiefel manifold, where the set $V_K(\mathbb{R}^{N\times 1})$ consists of $N \times K$ matrices with column-wise orthonormality. Each (unit) column vector, which satisfies the unit Euclidean norm condition, is sampled from von Mises Fisher Distribution (vMF) whose P-hypersphere version (*i.e.* $\mathbb{R}^{P\times 1}$) is defined as

$$\text{vMF}(\boldsymbol{\mu}) \equiv c_P(\|\boldsymbol{\mu}\|)\exp\{\mathbf{x}^\top\boldsymbol{\mu}\}\begin{cases} c_P(\|\boldsymbol{\mu}\|) = (2\pi)^{P/2}\frac{\|\boldsymbol{\mu}\|^{P/2-1}}{I_{P/2-1}(\|\boldsymbol{\mu}\|)} & \mid \|\boldsymbol{\mu}\| > 0 \\ c_P(0) = \frac{\Gamma(P/2)}{2\pi^{P/2}} & \mid \|\boldsymbol{\mu}\| = 0, \end{cases} \tag{1.4}$$

where $c_P(\|\boldsymbol{\mu}\|)$ is a normalization constant corresponding to the mean direction $\boldsymbol{\mu} \in \mathbb{R}^{P\times 1}$ of the resulting vector sample and $\|\boldsymbol{\mu}\|$ equals the concentration parameter which is similar to the precision parameter in other distributions. MvMF is an orthogonal matrix extension of vMF. After simulating a column vector sample using vMF, one can apply orthogonal operators to obtain vectors for the rest of the columns by projecting a sample over the null space of the reference vector. For a more general class of MvMF, see Hoff (2009).

1.5.2 Gibbs Sampling

By incorporating the augmented variable $\mathbf{Z}$ defined in Eq. 1.3, the conditional distribution of $N \times M$ voting record matrix $\mathbf{Y}$ can be expressed as

$$p(\mathbf{Y}|\mathbf{X},\mathbf{W},\mathbf{B},\mathbf{A},\phi) = \int_{\mathbf{Y}=f(\mathbf{Z})} p(\mathbf{Z}|\mathbf{X},\mathbf{W},\mathbf{B},\mathbf{A},\phi)\,\mathrm{d}\mathbf{Z}. \tag{1.5}$$

Owing to the normality of the independent and identically distributed error matrix $\mathbf{E}$, the probability of $\mathbf{Z}$ conditioned on observed and unobserved parameters is

$$\begin{aligned} p(\mathbf{Z}|\mathbf{X},\mathbf{W},\mathbf{B},\mathbf{A},\phi) &= \left(\frac{\phi}{2\pi}\right)^{NM/2} \exp\left(-\frac{\phi}{2}\|\mathbf{Z}-\mathbf{X}\mathbf{W}\mathbf{B}^\top-\mathbf{A}\|^2\right) \\ &= \left(\frac{\phi}{2\pi}\right)^{NM/2} \exp\left(-\frac{\phi}{2}\|\mathbf{E}_{-k}-w_k\mathbf{X}_{[,k]}\mathbf{B}_{[,k]}^\top\|^2\right) \\ &= \left(\frac{\phi}{2\pi}\right)^{NM/2} \exp\left(-\frac{\phi}{2}\|\mathbf{E}_{-k}\|^2+\phi w_k\mathbf{X}_{[,k]}^T\mathbf{E}_{-k}\mathbf{B}_{[,k]}-\frac{\phi}{2}w_k^2\right) \end{aligned} \tag{1.6}$$

with the sampling distributions given by $\mathbf{X} \sim \mathcal{U}(V_K(\mathbb{R}^{N\times 1}))$ (*i.e.* uniform distribution defined over $V_K(\mathbb{R}^{N\times 1})$), $\mathbf{w} \sim \mathcal{N}(\mu, 1/\psi)$ and $\mathbf{B} \sim \mathcal{U}(V_K(\mathbb{R}^{M\times 1}))$. The rank truncated error matrix $\mathbf{E}_{-k} \equiv \mathbf{Z} - \mathbf{X}_{[,-k]}\mathbf{W}_{[-k,-k]}\mathbf{B}_{[,-k]}^\top - \mathbf{A}$, where $\mathbf{X}_{[,-k]}$ is $N \times K-1$ stack of column vectors of $\mathbf{X}$ with k-th column omitted.

Using conjugate distributions (the uniform vector distribution vMF($\mathbf{0}$): $\tilde{X}_k$ and $\tilde{B}_k$; Gamma distribution $\Gamma(\nu_0/2, \nu_0\sigma_0^2/2)$: precision (*i.e.* ϕ) of $\mathbf{E}$; $\mathcal{N}(\mu_0, v_0^2)$: mean (*i.e.* μ) of $\mathbf{w}$; $\Gamma(\eta_0/2, \eta_0\tau_0^2/2)$: precision (*i.e.* ψ) of $\mathbf{w}$), Gibbs sampling chain for the recovery of $\mathbf{X}$, $\mathbf{W}$, $\mathbf{B}$ and the hyperparameters ϕ, μ and ψ is as follows. See appendix for details of prior selection.

1. Sample $\mathbf{Z}$ from truncated normal distributions

- $$Z_{ij} \sim p(Z_{ij}|Y_{ij},\mathbf{X},\mathbf{W},\mathbf{B},\mathbf{A},\phi) = \begin{cases} \mathcal{N}(\mathbf{XWB}_{ij} - A_{ij}, 1/\phi)\mathbb{1}(Z_{ij} < 0) & \mid Y_{ij} = 0 \\ \mathcal{N}(\mathbf{XWB}_{ij} - A_{ij}, 1/\phi)\mathbb{1}(Z_{ij} \geq 0) & \mid Y_{ij} = 1 \\ \mathcal{N}(\mathbf{XWB}_{ij} - A_{ij}, 1/\phi) & \mid Y_{ij} = \text{NA} \end{cases}$$

2. Sample latent variables: For $k \in \{1,\ldots,K\}$

   - $(\mathbf{X}_{[,k]} \mid \mathbf{Z},\mathbf{X}_{[,-k]},\mathbf{W},\mathbf{B},\phi) \equiv \mathbf{N}^{\mathbf{X}}_{\{-k\}}\tilde{X}_k,$ where $\tilde{X}_k \sim \text{vMF}(\phi w_k \mathbf{N}^{\mathbf{X}\top}_{-k}\mathbf{E}_{-k}\mathbf{B}_{[,k]})$
   - $(\mathbf{B}_{[,k]} \mid \mathbf{Z},\mathbf{X}_{[,-k]},\mathbf{W},\mathbf{B},\phi) \equiv \mathbf{N}^{\mathbf{B}}_{\{-k\}}\tilde{B}_k,$ where $\tilde{B}_k \sim \text{vMF}(\phi w_k \mathbf{X}^{\top}_{[,k]}\mathbf{E}_{-k}\mathbf{N}^{\mathbf{B}}_{-k})$
   - $(w_k \mid \mathbf{Z},\mathbf{X},\mathbf{D}_{[-k,-k]},\mathbf{B},\phi,\mu,\psi) \sim \mathcal{N}[(\phi\mathbf{X}^{T}_{[,k]}\mathbf{E}_{-k}\mathbf{B}_{[,k]} + \mu\psi)/(\phi+\psi), 1/(\phi+\psi)]$
   - $w_k \leftarrow -w_k$ and $\mathbf{X}_{[,k]} \leftarrow -\mathbf{X}_{[,k]} \mid w_k < 0$
   - $(\boldsymbol{\alpha}|\mathbf{Z},\mathbf{X},\mathbf{W},\mathbf{B}) = \sum_i(\mathbf{Z} - \mathbf{XWB}^{\top})_{[i,]}/N$

3. Sample hierarchical parameters

   - $(\phi \mid \mathbf{Z},\mathbf{X},\mathbf{W},\mathbf{B}) \sim \Gamma[(\nu_0 + MN)/2, (\nu_0\sigma_0^2 + \|\mathbf{Z} - \mathbf{XWB}^{\top} - \mathbf{A}\|^2)/2]$
   - $(\mu \mid \mathbf{W},\psi) \sim \mathcal{N}(\psi\sum_k w_k + \mu_0/v_0^2)/(\psi K + 1/v_0^2), 1/(\psi K + 1/v_0^2)$
   - $(\psi \mid \mathbf{W},\mu) \sim \Gamma[(\eta_0 + K)/2, (\eta_0\tau_0^2 + \sum_k(w_k - \mu)^2)/2]$

$\mathbf{N}^{\mathbf{X}}_{\{-k\}}$ is $N \times N - (K-1)$ stack of basis vectors for the null space of the column stack $\mathbf{X}_{[,-k]}$. $\mathbf{N}^{\mathbf{B}}_{\{-k\}}$ is $M \times M - (K-1)$ stack of basis vectors for the null space of the column stack $\mathbf{B}_{[,-k]}$. $\mathbb{1}(\cdot)$ is an indicator function. Since each estimate is projected over a null space, the operation $\mathbf{N}^{\mathbf{X}}_{\{-k\}}\tilde{X}_k$ produces a column vector, which is orthogonal to $\mathbf{X}_{[,-k]}$ and has a unit norm. See appendix for the details of the estimation algorithm.

1.5.3 Properties of Estimates

We can investigate the properties of the proposed method with help of analytic properties of probabilistic PCA (PPCA), which has a similar formulation to MIRT (Minka 2000, Tipping & Bishop 1999). As shown in the appendix, we obtain the following results for the mean estimates of the item parameter matrix $\tilde{\mathbf{B}}$ that its columns lie on maximum variance (principal) directions of the covariance matrix $\mathbf{S} \equiv \text{Cov}(\mathbf{z}_i, \mathbf{z}_j) \forall i, j \in \{1, \ldots, N\}$. Such property is obtained by the orthogonality constraint given by the identification strategy. This (Lemma 2) is different from the rotational coordinates obtained by algorithms without orthogonal constraints, resulting in estimates that only span the maximum variance subspace (Lemma 1).

Proposition 1 *MIRT without orthogonal constraints produces item parameter matrix* $\mathbf{B}$ *(*$\tilde{\mathbf{B}}$*) whose posterior mode spans the principal subspace of the augmented variable covariance matrix* $\mathbf{S} \equiv \text{Cov}(\mathbf{z}_i, \mathbf{z}_j) \forall i, j \in \{1, \ldots, N\}$.

Proposition 2 *The proposed method produces item parameter matrix* $\mathbf{B}$ *(*$\tilde{\mathbf{B}}$*) whose posterior mode lies on the principal directions of the augmented variable covariance matrix* $\mathbf{S} \equiv \text{Cov}(\mathbf{z}_i, \mathbf{z}_j)$ $\forall i, j \in \{1, \ldots, N\}$.

Proofs are given in the appendix.

1.6 Simulation Evidence

Here, we present a series of simulation analysis to show the performance of the proposed method compared with two existing methods of MIRT, IDEAL and W-NOMINATE. After showing the proposed method's recovery performance on 3 dimensional datasets, we compare its mixing performance with IDEAL. It turns out that IDEAL outputs highly

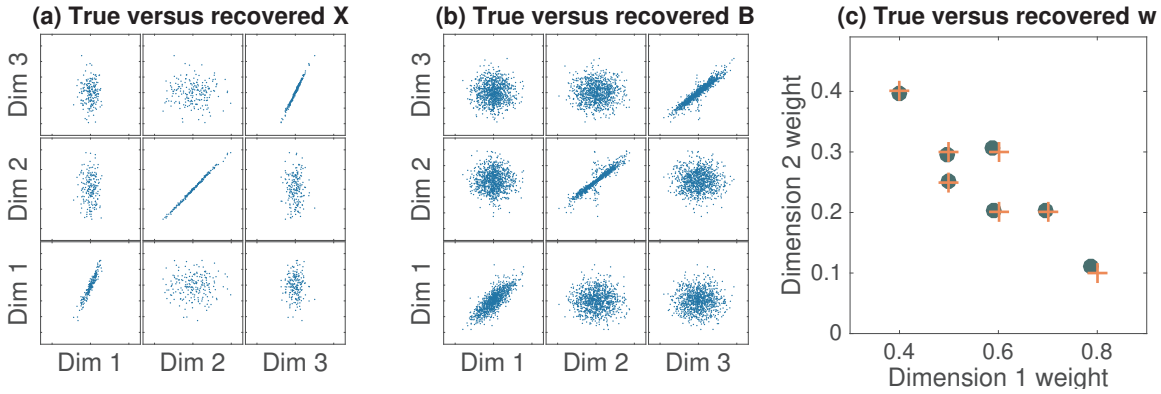

Figure 1.1: Synthetic examples. (**a**)(**b**) Matrix plot for mean estimates of $\mathbf{X}$ and $\mathbf{B}$ for a synthetic voting dataset with parameters: $N = 200$, $M = 1{,}000$, $\mathbf{w} = (0.5, 0.3, 0.2)$ and $\phi = 10^6$. Planted $\mathbf{X}$ and $\mathbf{B}$ are sampled from centered orthonormal matrices (*i.e.* $\forall k \in \{1, \ldots, K\}$, $\sum_i X_{ik} = 0$ and $\sum_i B_{ik} = 0$). (**c**) Seven independent examples with different true weight combination (+) and mean estimates of inferred weights (•). Only first two largest dimensional weights are depicted since all dimensional weights sum up to one. Each $\mathbf{w}$ is treated as a vector of linear fractions, so that $\mathbf{w} \leftarrow \mathbf{w} / \sum_k w_k$. Each axis of the ternary plot indicates each dimension weight for $K = 3$. All estimates are obtained from 100,000 Markov Chain Monte Carlo (MCMC) iterations with 10,000 burn-in trials by thinning out 1 over 100 outputs.

unreliable estimates when the number of data generating dimensions is three. Finally, the proposed method is compared with W-NOMINATE for the recovery of dimension weight parameter including the cases when the number of dimensions is misspecified.

1.6.1 Performance

We show the validity of the proposed approach through synthetic roll call voting datasets generated using the formulation given in Eq. 1.2. The series of examples depicted in Figure 1.1 demonstrate that the proposed method successfully recovers the true parameter tuple. Seven synthetic roll call datasets are generated. True parameters for 200×3 matrix $\mathbf{X}$ and $1{,}000 \times 3$ matrix $\mathbf{B}$ are sampled from uniform Stiefel manifolds of given sizes, and $\mathbf{w}$ values are set to shown values in Figure 1.1(c).

For all seven examples explored, the proposed methods successfully recovered the true parameter values. Pearson correlation coefficients between estimated mean $\mathbf{X}$ and $\mathbf{B}$ and

their true values exceed 0.8. For high weight dimensions larger than 0.5, all corresponding coefficients exceed 0.95. As shown in Figure 1.1(c), the proposed algorithm not only reproduces the individual and bill parameters, but also recovered the true weight parameter $\mathbf{w}$ even when one of the true dimensional weights was as small as 10% of the sum of the total weights.

1.6.2 Comparison to IDEAL

Albert (1992)'s normal ogive IRT sampler and its application to ideal point estimation, IDEAL (Clinton, Jackman & Rivers 2004), is the closest Gibbs sampler of the proposed method. While it has been used as a canonical model of Bayesian IRT, there are notable differences between IDEAL and the proposed method in addition to the standard bi-factor MIRT formulation of IDEAL.

First and most importantly, in IDEAL, both ability parameter and item parameter vectors are sampled over $i \in \{1,\dots,N\}$ and $j \in \{1,\dots,M\}$ respectively. In fact, this property yields an ineffective organization of its Gibbs sampler by allowing a large number of single site updates. In contrast, the proposed method samples over dimension $k \in \{1,\dots,K\}$. This setup, which enables orthogonalization among dimension-wise estimates, can be regarded as an effective blocking strategy for Gibbs sampling to reduce the number of single site updates. This difference creates a clear distinction between these two algorithms in terms of computational complexity. Whereas IDEAL requires the computation of $N+M$ loops in total for creating a single sample, the proposed method requires the computation of $2K$ loops, where $K << N, M$.

Moreover, this slow sequential updating of IDEAL, in conjunction with the lack of identification constraint, undermines the performance of the algorithm when $K > 2$. The top panels of Figure 1.2 compare the planted 3 dimensional ideal points and the mean estimates obtained from $10,000$ runs of sampling and $1,000$ burn-in trials, using the ideal

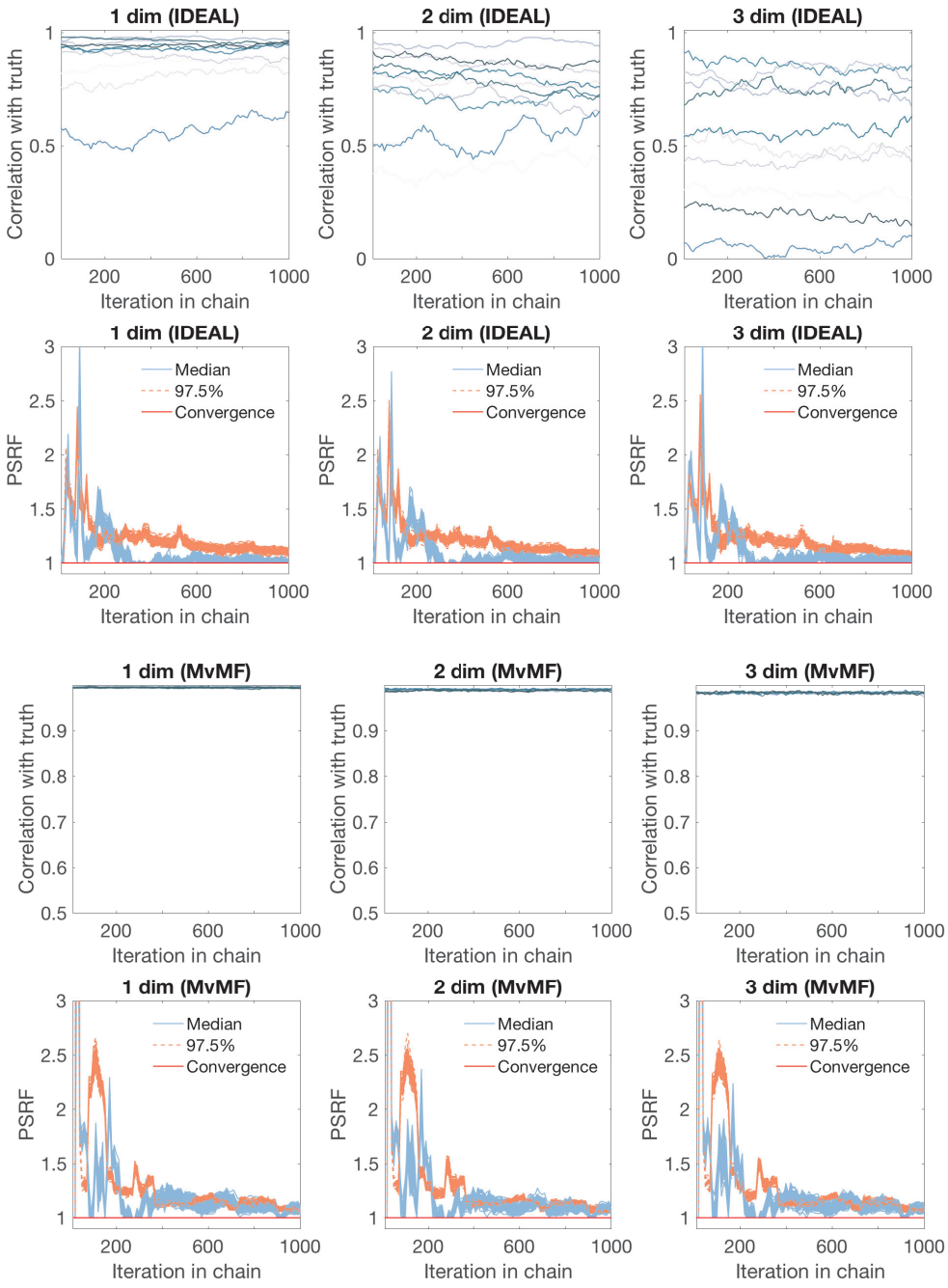

Figure 1.2: Convergence diagnostics. Pearson correlation of the sample draws with true values and convergence diagnostics of IDEAL and the proposed method. Same parameters are used as in Figure 1.1 for simulation. Estimates are obtained from $100,000$ runs of sampling and $1,000$ burn-in trials, and the graphs depict every hundredth samples. Top panel shows the correlation coefficients between each dimensional estimate of $\mathbf{X}$ and its corresponding ground truth values. Bottom panel shows the Gelman-Rubin's PSRF computed using 10 independent chains. Starting values for both methods were sampled from the standard multivariate normal distribution. Each plot contains $N = 100$ individual parameters for each output type.

function in pscl R package (Zeileis, Kleiber & Jackman 2008). We also present the outputs of the proposed method over the same number of iterations. Convergence diagnostics are measured through the Gelman-Rubin's potential scale reduction factors (PSRF) (Brooks & Gelman 1998), which approach 1 upon convergence. Despite the relatively short length of sampling chains, both IDEAL and MvMF converge. However, the resulting outcomes of IDEAL are highly inferior compared to the estimates of the proposed method. Pearson correlation of the estimates of IDEAL exhibit inconsistent values over the iterations, ranging from 0.3 to 1. The results become erroneous as the true dimension importance decreases from the first dimension to the third dimension. In sum, the converged estimates of IDEAL do not recover true ideal points when the number of dimensions is 3.

Finally, as it is trivial due to the lack of orthogonality constraints in IDEAL, the variance-covariances of multivariate normal densities associated with ability parameter and item parameter matrices are not diagonal. To be specific, for the sampling of a K dimensional vector for the ideal point of legislator i, the variance-covariance matrix equals $\mathbf{X}^\top\mathbf{X}+\mathbf{D}^{-1}$ where $\mathbf{D}$ is the prior distribution's covariance matrix which is diagonal. Also when sampling a K dimensional vector for item parameter vector of bill j, the variance-covariance matrix equals $\tilde{\mathbf{B}}^\top\tilde{\mathbf{B}}+\mathbf{V}^{-1}$ where $\mathbf{V}$ is the diagonal covariance of $\mathbf{B}$'s prior distribution (Jackman 2009). As shown, since both $\mathbf{X}^\top\mathbf{X}$ and $\tilde{\mathbf{B}}^\top\tilde{\mathbf{B}}$ are not diagonal due to the lack of orthogonality in $\mathbf{X}$ and $\tilde{\mathbf{B}}$, IDEAL and Albert (1992)'s sampler produce dimension-wise correlated samples.

1.6.3 Comparison to W-NOMINATE

Here, we compare the performance of the proposed method with W-NOMINATE, the most widely used multidimensional ideal point estimation method for roll call datasets. In contrast to IDEAL, which does not have an explicit weight paramete, W-NOMINATE contains a dimension-specific weight parameter $(\omega_k^2|k\in\{1,\ldots,K\})$. Since W-NOMINATE

incorporates a different voting model from Eq. 1.2, we first show the equivalence between the Normal-Normal formulation of W-NOMINATE and the Quadratic-Normal formulation in Eq. 1.2. In fact, the Quadratic-Normal formulation is the first-order approximation of the Normal-Normal formulation of W-NOMINATE (Armstrong, Bakker, Carroll, Hare, Poole & Rosenthal 2014).

$$
\begin{aligned}
\Pr(i \text{ chooses } \mathbf{b}_j) &= \Phi\left(U_i(\mathbf{b}_j) - U_i(\mathbf{s}_j)\right) \\
&= \Phi\left[\gamma \exp\left\{-\frac{1}{2}\sum_k \omega_k^2 (x_{ik} - b_{jk})^2\right\} - \gamma \exp\left\{-\frac{1}{2}\sum_k \omega_k^2 (x_{ik} - s_{jk})^2\right\}\right] \\
&= \Phi\left[\gamma \sum_{i=0}^{\infty} \frac{\left\{-\frac{1}{2}\sum_k \omega_k^2 (x_{ik} - b_{jk})^2\right\}^i}{i!} - \gamma \sum_{i=0}^{\infty} \frac{\left\{-\frac{1}{2}\sum_k \omega_k^2 (x_{ik} - s_{jk})^2\right\}^i}{i!}\right] \\
&\approx \Phi\left[-\gamma\frac{1}{2}\sum_k \omega_k^2 (x_{ik} - b_{jk})^2 + \gamma\frac{1}{2}\sum_k \omega_k^2 (x_{ik} - s_{jk})^2\right] \qquad (1.7)
\end{aligned}
$$

Accordingly, the weight parameter vector $\mathbf{w}$ in Eq. 1.2 corresponds to $(\omega_k^2 | k \in \{1, \ldots, K\})$ in W-NOMINATE.

In case when $\mathbf{X}$ and $\mathbf{B}$ have orthonormal constraints, we may have a direct comparison of the two vectors. Yet, when such adjustment is not made, it is infeasible. NOMINATE family assigns the unit-hypersphere constraint so that all values of $\mathbf{X}$ and $\mathbf{B}$ are adjusted to lie on $[-1, 1]$. This fixed min/max condition implies that the covariance matrices for $\mathbf{X}$ and $\mathbf{B}$ do not satisfy the unit-covariance (*i.e.* identity covariance) condition. As a result, the weight parameter in W-NOMINATE varies arbitrarily depending on the combination of the values in $\mathbf{X}$ and $\mathbf{B}$ estimates, and does not convey relevant information for the dimension weights assumed in the generative model (Eq. 1.7).

Figure 1.3 compares the estimates of the proposed method and W-NOMINATE to investigate what happens when the number of dimensions preset for modeling (K') is different from the true number of dimensions (K). Using the same voting matrix for both

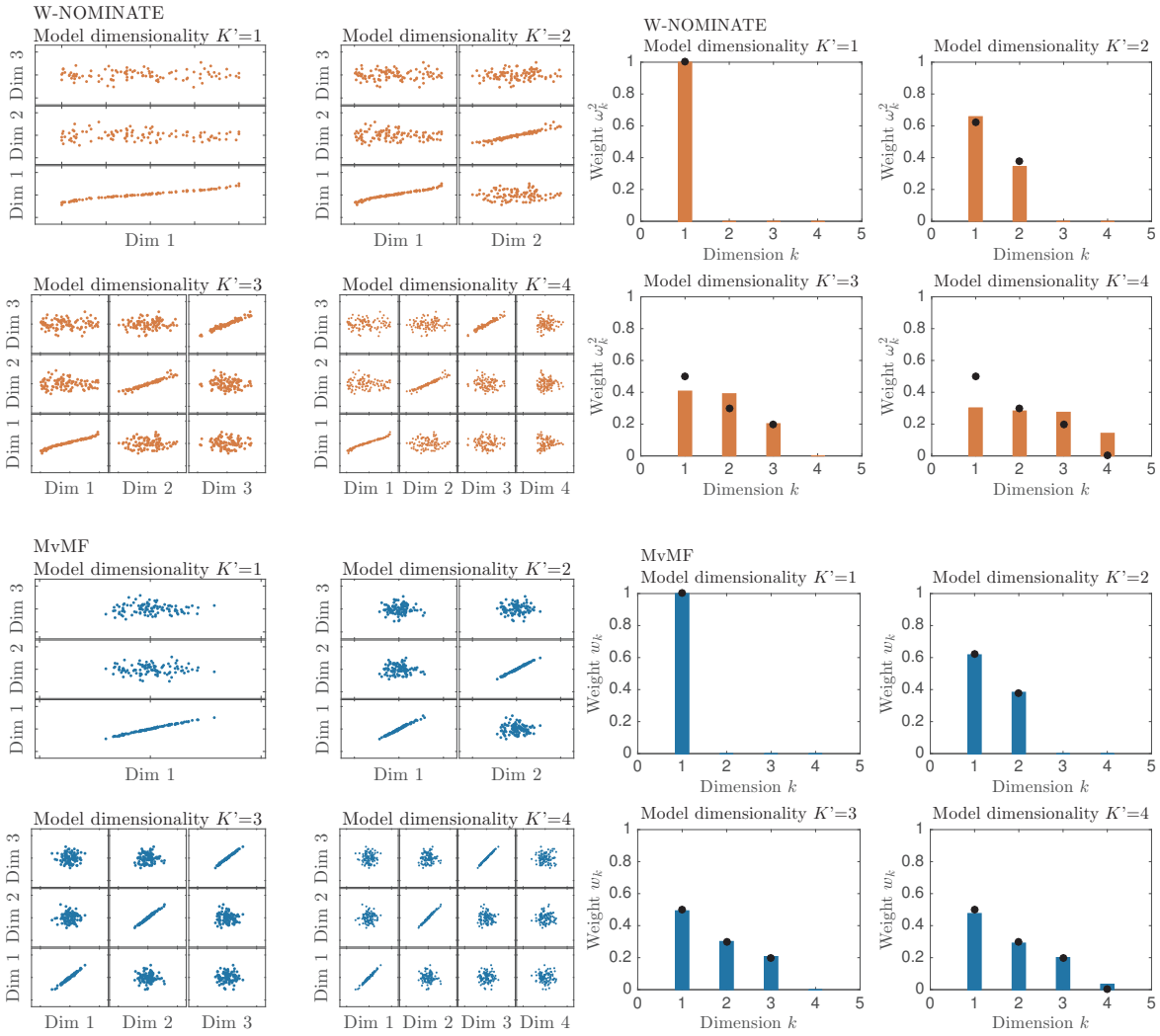

Figure 1.3: Comparison with W-NOMINATE. A synthetic example with model dimensionality (K') varying while the true number of dimensions $K = 3$. (**Left**) $\mathbf{X}$ outputs obtained using W-NOMINATE and the proposed method. (**Right**) $\mathbf{w}$ outputs obtained using W-NOMINATE and the proposed method. Planted parameters: $N = 200$, $M = 1,000$, $\mathbf{w} = (0.5, 0.3, 0.2)$ and $\phi = 10^6$. The weight output of W-NOMINATE is obtained using nomObject$weights function of W-NOMINATE R package, and then squared and normalized to have the sum of 1. For the proposed method, each $\mathbf{w}$ is treated as a vector of linear fraction, so that $\mathbf{w} \leftarrow \mathbf{w}/\sum_k w_k$. All estimates are obtained from 100,000 MCMC iterations with 10,000 burn-in trials and thinning out 1 over 100 outputs.

methods, we first applied wnominate function in wnominate R package with varying K'. In contrast to the mean estimates of the proposed method where the planted weight parameter is precisely recovered regardless of the model rank (K') specification, the nomObject$weights

output of wnominate R package (Poole et al. 2008) does not recover interval information of the weight estimates, making it infeasible to know the relative importance of each dimension in an intuitive scale. In particular, when the number of dimensions preset (K') varies, it does not preserve the ratio of the weight estimates. For example, the first and second dimensional weight ratio is close to 1:1 when $K' = 3$, whereas this ratio significantly varies when $K' = 2$. Moreover, the excess dimension for $K' = 4$ gets a significant amount of weight (0.14), which is comparable to other weight estimates.

In contrast, the proposed method successfully recovers the true parameters, exhibiting a very convenient property for interpreting the weight estimates even in the case of model misspecification. The proposed method discovers the ratio (*e.g.* $\mathbf{w} = (0.625, 0.375)$ for $K' = 2$) of the true dimension weights ($\mathbf{w} = (0.5, 0.3, 0.2)$) when $K' \leq K = 3$, and accurately gives the 4th dimension weight close to zero when $K' = 4$. In other words, the proposed method successfully retrieves the hidden dimensionality of the voting matrix for given K', best explaining the covariance structure when the model dimensionality (K') is lower than or equal to the true dimensionality. When $K' > K$, it assigns a negligible weight to the excess dimension. In contrast, the difficult-to-interpret weight parameter in W-NOMINATE produces arbitrary values, although it recovers the latent legislator ideology matrix $\mathbf{X}$ similar to the proposed method.

1.7 Application: The United States House of Representatives Roll Call Voting Records

We applied the proposed method to the roll call voting records of the U.S. House from the late 19 century to present (*i.e.* 46th-113th sessions). This period has been considered as the two party regime. We applied the proposed method by setting the number of dimensions as 3. Since session-specific measures are obtained, we matched contiguous

session dimensions by pairing the ones with the largest amount of correlation in incumbent ideology. Figure 1.4 illustrates the varying dimensionality in the House roll call voting records, which are identified using the proposed method. What is notable at first glance is the apparent multidimensionality of the ideological space across the entire period of the sessions analyzed. We observe that even the least significant dimension at minimum accounts for the 10% of the utility assumed in Eq. 1.2. This is in contrast with the previous studies, which state that the Congress has been constantly unidimensional with the exception of the second dimensional importance during the mid-20th century sessions (McCarty, Poole & Rosenthal 2016, Poole & Rosenthal 2000).

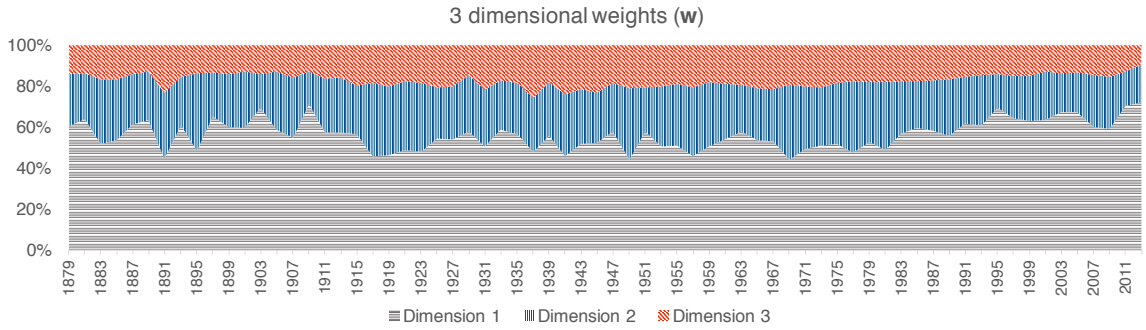

Figure 1.4: Dimensionality of U.S. House. Mean estimates of **w** obtained via 100,000 MCMC iterations with 10,000 burn-in trials and by thinning out 1 over 100 outputs. Each number on the x-axis indicates the starting year of the corresponding Congressional session. y-axis indicates the cumulative weight value. House dimensions between sessions are matched based on the correlation between incumbent dimension-specific ideology in contiguous sessions.

The first dimension, which exhibits its weight value above 50% during the majority of the analyzed period, largely corresponds to the dimension of conflict between Democrats and Republicans. When the mean party distance on each dimension is calculated, we see that the two party distance is mainly explained by the changes in the first dimensional estimates (Figure 1.5). The decline of the first-dimensional importance during the early 20th century coincides with the ending period of the strong leadership in the House and the beginning of the progressive era (late 1910s). After the end of the civil rights era around the early 1970s,

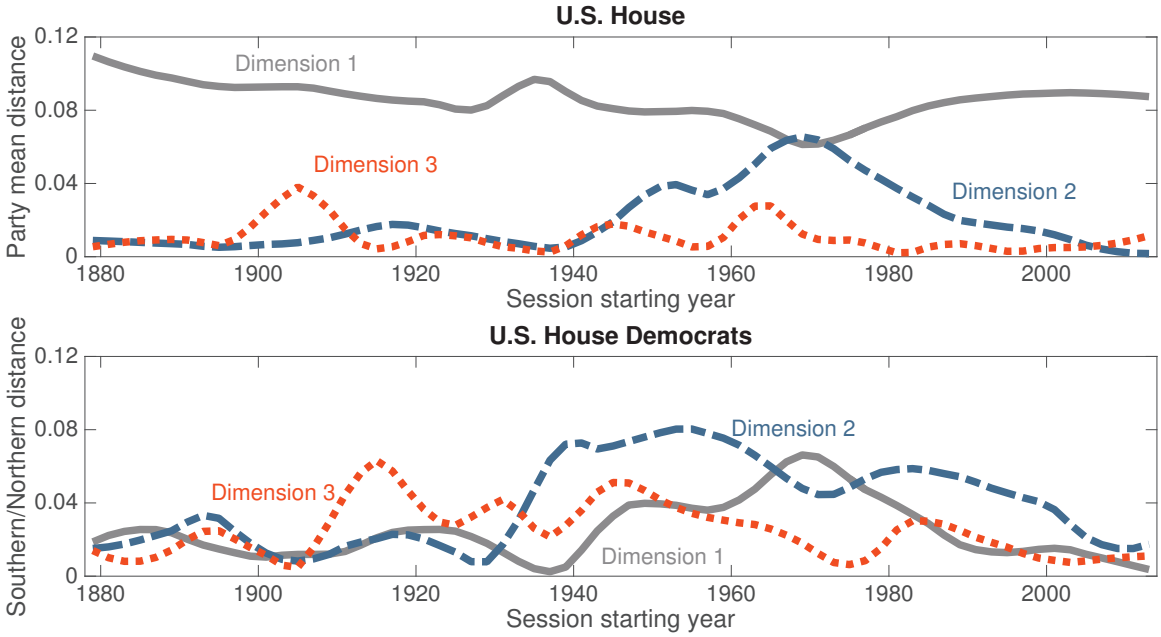

Figure 1.5: Mean Distance on each dimension. Party mean distance between Democrats and Republicans on each dimension, and mean distance between southern and northern Democrats. Southern Democrats are labeled by their membership in the Confederate States. Values are not adjusted by dimension weights. House dimensions between two adjacent sessions are matched based on the correlation between incumbent dimension-specific ideology. Unlike other methods, the unit norm condition sets dimension-wise estimates to have a common scale.

the weight value of the first dimension starts to increase.

The second dimensional estimates are largely correlated with the cleavage structure on civil right issues. The large amount of polarization between mean Democratic ideology and Republican ideology is broadly observed throughout the southern realignment period (from 1933 to 1971 (Sin 2014)) in which southern and northern democrats began to form ideologically distinct coalitions, eventually experiencing party switching and seat changes (Figure 1.5). Along with the distance between the factions within the Democratic party, the second dimensional importance has been magnified, hitting the maximum value of 38% in the 1960s. Due to the conflict over the civil rights issues, the second dimension has also been a primary battle ground between Democrats and Republicans during the late 1960s through the early 1970s.

An interesting period on which we need to focus for identifying the characteristic of the third dimension is the first three decades of the 20th century, specifically from 1897 to 1933. During this period, the central issue of conflict within each party was the government's role on industrialization (Shepsle 1978, Sin 2014). While conservative factions within each party asserted that interventions should be minimal, progressive Republicans and reform agrarian Democrats asserted progressive monetary and welfare policies of the Government for regulating the economy. Such conflict eventually led to the divide of the Republican party, and resulted in a short-lived party called the Progressive party. In Figure 1.6, the Progressive party, which mainly consists of defected legislators of the Republican party, locate its members at the edge of the third dimension. This is the only dimension that locates the median ideology of the progressive party members at the most extreme (progressive) side of the ideological axis.

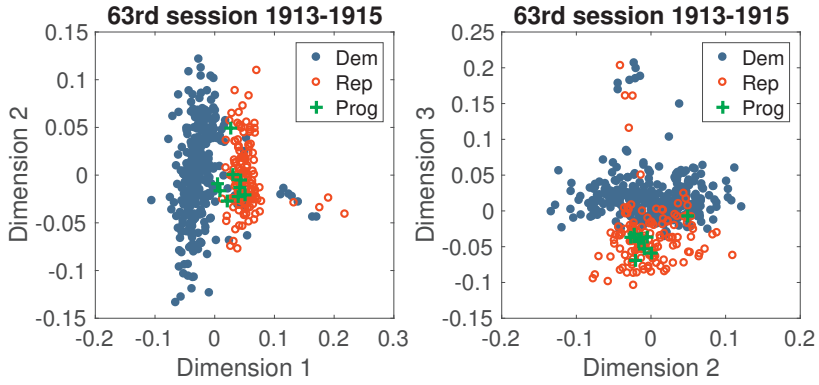

Figure 1.6: Progressivism era. Ideological spectrum of the short-lived Progressive Party in the 63rd session (1913-1915). Ideal points of Democrats are colored in filled blue and ideal points of Republicans are colored in blank red. Progressive party members are located in green cross-shaped markers.

One important perspective we can use to analyze the patterns of estimates in the multidimensional space is to see the relationship between different sets of items and legislators. For example, how much a specific bill or subsets of bills divide over the partisan cleavage structure?

We provide further analysis using relational representation of item and legislator

locations using a technique which we name **item2vec**. Figure 1.7 shows a schematic illustration of the idea. Analogous to **word2vec** in text analysis literature (Mikolov, Chen, Corrado & Dean 2013), the goal of **item2vec** is to find mean directions of items and legislators, as computed using the vector differences between a set of pairs among items and legislators, and their level of agreement. For a single pair locations, consisting of two locations $\mathbf{x}_i$ and $\mathbf{x}_j$, the vector difference $\vec{a}$ between these two locations are simply defined as $\vec{a} = \mathbf{x}_i - \mathbf{x}_j$.[3] In fact, the item discrimination parameter β_j for bill j is equivalent to the vector difference between the bill and status quo locations for item j.

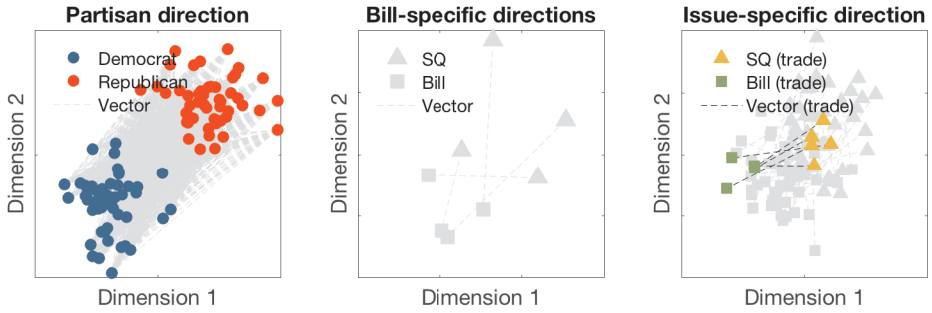

Figure 1.7: item2vec. [**Left**] Identification of partisan directions using all pairwise vectors between inter-party members of a Congressional session. [**Middle**] Identification of bill-specific directions by measuring the direction between a bill and its corresponding status quo (SQ) location. [**Right**] Among the entire set of bills and status quos in a session, we can identify issue-specific mean direction by subsetting and averaging the vectors of bill-stuatus quo pairs belonging to an issue category.

We can use vector differences to quantify how much votings over a single bill or a subset of bills are aligned with a particular direction of conflict (*e.g.* partisan cleavage). Formally, to measure the alignment between a bill direction $\vec{w}$ with a set of directions (or vector differences) A, we compute the following quantity.

$$p(\vec{w}, A) = \text{mean}_{\vec{a} \in A} |\cos(\vec{w}, \vec{a})|. \tag{1.8}$$

[3] $\vec{a}$, $\mathbf{x}_i$ and $\mathbf{x}_j$ are all vectors in the same ideological space. However, to make a distinction between vector differences and the estimates coming from the MIRT model, we use different notations to denote these objects.

$p(\vec{w}, A)$ represents the mean level of alignment between $\vec{w}$ and a set of vectors in A. $\cos(\vec{w}, \vec{a})$ is the inner product or cosine similarity between the two vectors. If these vectors are well aligned, the mean direction $p(\vec{w}, A)$ becomes 1 whereas it becomes 0 if they are orthogonal. Accordingly, it is a naturally normalized measure. For instance, the set of vector differences can be the set of pairwise vector differences of all pairs of legislators belonging to two different parties or the set of bill-status quo vector differences for all items belonging to a specific issue category. As illustrated in the left panel of Figure 1.7, if we measure $p(\vec{w}, A)$ for an item vector difference $\vec{w}$ and the partisan pairwise vector difference set A, $p(\vec{w}, A)$ will inform us how much the votes over the corresponding bill is divided over the party line.

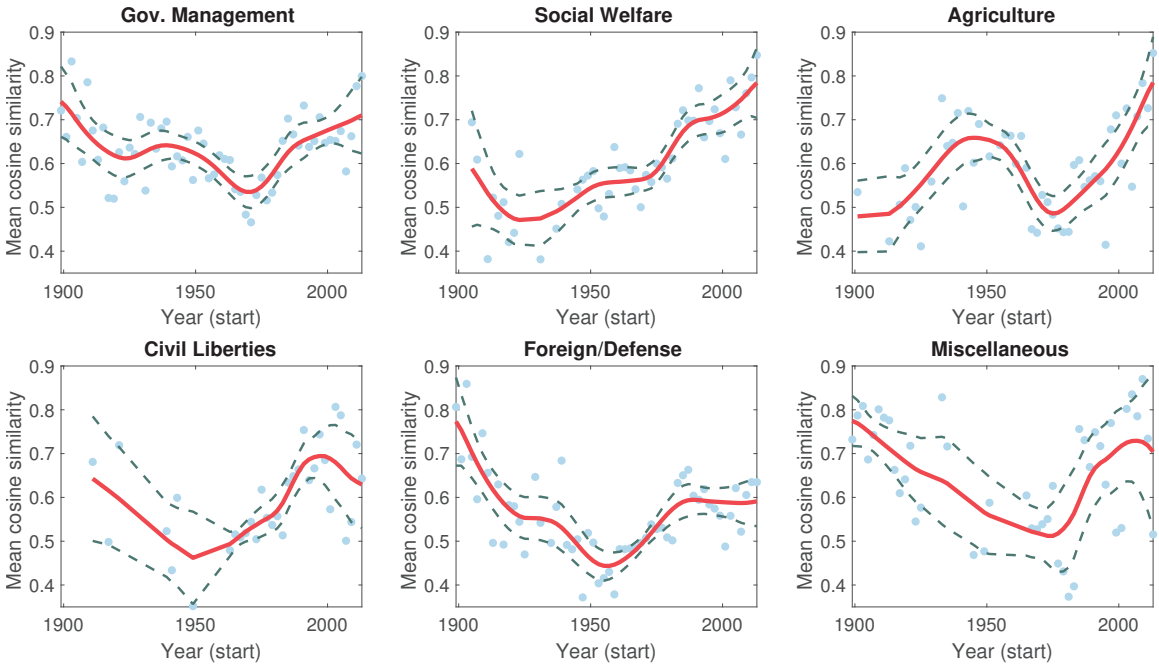

Figure 1.8: Issue dynamics. Mean vector alignment between bill issue subsets classified using Clausen issue category and partisan directions of each session. light blue dots indicate mean alignment scores, computed using **`item2vec`**, and red lines and green dashed lines indicate the trends smoothened using LOESS and the corresponding confidence intervals.

We apply **`item2vec`** for subsets of bills over the history of Congress analyzed. Figure 1.8 shows the results. For high mean similarity points, the corresponding issue related bills were mostly voted by the members' party line, indicating the issue was partisan issue in those periods. For low mean similarity points, the corresponding issue was likely to

be a valence issue so that inter-party voting coalition formation was frequent. For the period analyzed, we can verify that social welfare has been constantly progressed as a partisan issue. In contrast, Foreign and defense related bills do not show such pattern that the trend exhibits constant decrease over the first half of the 20th century and started to increases after the end of the second world war. This corresponds to the common belief on American legislative history that defense related issues reached a high level of consensus between the party members during the world war eras and began to follow the party line cleavage over the second half of the 20th century including the Vietnamese war period. On the other hand, civil liberty related bills follow the party realignment trend in American history so that their issue exhibits a high level party polarization over the civil rights era and shows a higher level of polarization in the course of party realignment due to the absorption of the issue into their party platforms.

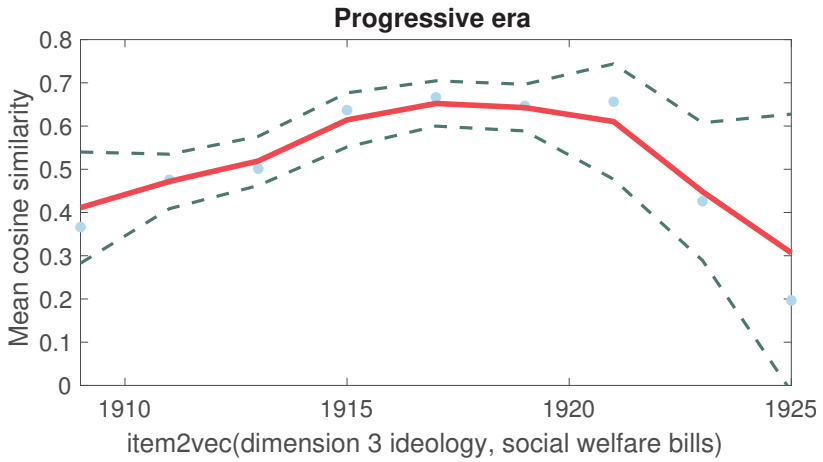

Figure 1.9: item2vec during progressive era. Mean vector alignment between the 3rd dimensional direction (*i.e.* $(0,0,1)$) and the pair-wise directions of bill-status quo pairs for social welfare related bills in each session over the progressive era. All dots and lines are drawn same as the previous figure.

Figure 1.9 appends the results shown in Figure 1.6. Each dot indicates the level of alignment between the third dimensional direction (*i.e.* $(0,0,1)$), exactly aligned with the third axis, and the set of item vector differences for social welfare related bills over the progressive era. The pattern follows the rise and fall of the alternative faction within

Republican party and reaches its maximum around the period when the Progressive party existed. The result demonstrates that the third dimensional cleavage was mainly due to the divide over welfare related issues as claimed in Figure 1.6 using the result of a session in which the Progressive party existed.

1.8 Conclusion

The proposed method is a Bayesian 2 parameter normal ogive MIRT model that constrains the multidimensional item discrimination and ability parameters to form orthonormal matrices. This treatment resolves the prevalent identification problem in 2 parameter IRT models. Also it produces normalized estimates along maximum variance directions as opposed to conventional methods that are subject to arbitrary rotation and scaling of their estimates. While it is generally applicable to all practices of IRT including ability and item difficulty parameter inference for educational testing datasets, we applied the proposed method to voting datasets. The proposed method provides fully Bayesian estimates of the multidimensional legislator ideal points, the bill parameters and the often overlooked dimension-weight parameter. Through simulation, we showed that the proposed method successfully recovers the true data generating parameters of the synthetic roll call matrices. When compared to its Bayesian ancestor IDEAL (Albert 1992, Clinton, Jackman & Rivers 2004), the proposed method almost perfectly recovered the true 3 dimensional parameters, whereas IDEAL was unable to recover true estimates, exhibiting declining accuracy as the dimension importance decreases. Furthermore, the proposed method produces an interpretable weight parameter, corresponding the actual dimension importance assumed in the structural model of voting. In contrast, W-NOMINATE's weight parameter contained arbitrary values that are uncorrelated with true values.

Empirical analysis of the U.S. House demonstrates the existence of 3 dimensional ideological space in the roll call votes throughout the 20th century. While the principal

dimension conveys cleavage structure, which largely corresponds to the inter-party conflicts among Democrats and Republicans, the second and the third dimensions illuminate distinct voting blocs within each party. By analyzing the patterns of clustering on these dimensions, we show evidence on the existence of intra-party factions over the major period of sessions analyzed. This finding is expected to encourage the use of multidimensional ideal point estimation methods and associated quantitative approaches in the study of American political development that has been largely carried out by utilizing qualitative research methodology.

The proposed method is expected to serve as a baseline model for future IRT developments. Namely, we can think of three extensions of the method. First, the proposed method can incorporate a linkage function to enable the common scale identification of multiple datasets originating from different sources through the use of orthogonal Procrustes solutions (Gower & Dijksterhuis 2004). In ideal point estimation context, this can be used to map the ideal points of political actors associated with the Senate, the House, the President and the public in the same space. Second, the method can incorporate a polychotomous link function to accommodate ordered outcomes by using a Gibbs sampler for ordered Probit regression (Rossi, Allenby & McCulloch 2005). Moreover, the MvMF sampler can be applied for the estimation of ideal points from non-voting datasets, such as networks (Barberá 2015) and texts (Lowe et al. 2011), not limited to the 2 parameter MIRT model. Finally, the proposed method can have another layer of hierarchy to incorporate covariates for modeling ideal points. The most well-known application of this idea is to create a dynamic model in which one of the covariates is the session number (Poole 2005).

1.9 Appendix

Theorem 1 *Constraining* **X** *and* **B** *to form orthonormal matrices and centering* **B** *locally identifies the multidimensional item response theory model.*

1.9.1 Proof

I show the validity of the propose identification strategy following the original notation of Magnus & Neudecker (1988) and Rothenberg (1971). To be specific, we show that the proposed identification strategy produces locally identified estimates regardless of prior specification. By locally identified, we denote that the estimates are identified up to sign reversal (Clinton, Jackman & Rivers 2004, Rivers 2003) and, by identification regardless of prior specification, we denote that the identification is not limited to the Bayesian sense of identification (San Martın & González 2010, Gelman, Carlin, Stern & Rubin 2014), which has various versions and is in a controversy, but satisfies more conservative and universal criteria of identification in non-Bayesian framework (Rothenberg 1971, Magnus & Neudecker 1988, Rivers 2003). As a result, the identification results even hold under the use of improper, non-informative or flat priors.

In the following, the parameter points $\boldsymbol{\theta}_0 \in \boldsymbol{\theta}$ and the parameter point set $\boldsymbol{\theta} \equiv \left(\mathbf{X}, \widetilde{\mathbf{B}}, \mathbf{A}\right)$, where $\mathbf{B}$ is redefined as $\widetilde{\mathbf{B}} \equiv \mathbf{WB}$. We now have a standard bilinear formulation of Eq. 1.3 with $\mathbf{Y}^* = \Phi(\mathbf{X}\widetilde{\mathbf{B}}^\top - \mathbf{A})$.

I start from the main result of Rothenberg (1971) that a statistical model is identified if the stacked Jacobian matrix of the parameter points ($\mathbf{J}(\boldsymbol{\theta}_0)$) and the constraint matrix ($d\mathbf{h}(\boldsymbol{\theta}_0)/d\boldsymbol{\theta}$) have the full rank ($NK + M + MK$) with the constraint condition $\mathbf{h}(\boldsymbol{\theta}_0) = \mathbf{0}$. The stacked matrix is defined as

$$\mathcal{J}(\boldsymbol{\theta}_0) = \begin{pmatrix} \mathbf{J}(\boldsymbol{\theta}_0) \\ \frac{d\mathbf{h}(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}} \end{pmatrix} \tag{1.9}$$

$$= \begin{pmatrix} \mathbf{J}^{\mathbf{X}}_{\mathbf{X}} & \mathbf{J}^{\widetilde{\mathbf{B}}}_{\mathbf{X}} & \mathbf{J}^{\mathbf{A}}_{\mathbf{X}} \\ \mathbf{J}^{\mathbf{X}}_{\widetilde{\mathbf{B}}} & \mathbf{J}^{\widetilde{\mathbf{B}}}_{\widetilde{\mathbf{B}}} & \mathbf{J}^{\mathbf{A}}_{\widetilde{\mathbf{B}}} \\ \mathbf{J}^{\mathbf{X}}_{\mathbf{A}} & \mathbf{J}^{\widetilde{\mathbf{B}}}_{\mathbf{A}} & \mathbf{J}^{\mathbf{A}}_{\mathbf{A}} \\ \frac{d\mathbf{h}(\boldsymbol{\theta}_0)^\top}{d\mathbf{X}^\top} & \frac{d\mathbf{h}(\boldsymbol{\theta}_0)^\top}{d\widetilde{\mathbf{B}}^\top} & \frac{d\mathbf{h}(\boldsymbol{\theta}_0)^\top}{d\mathbf{A}^\top} \end{pmatrix}, \tag{1.10}$$

where $\mathbf{J}_{\Gamma}^{\Psi}$ is the component of the Jacobian matrix $\mathbf{J}(\boldsymbol{\theta}_0)$ with Γ portion of $\boldsymbol{\theta}_0$ partially differentiated by $\Psi \subset \boldsymbol{\theta}$. It is known that the bilinear formulation of the 2 parameter IRT model gives rank $NK+M+MK-K(K+1)$ to $\mathbf{J}(\boldsymbol{\theta}_0)$ (Rivers 2003), and we need at least $K(K+1)$ for the rank of the rows of $\mathcal{J}(\boldsymbol{\theta}_0)$ corresponding to $\frac{d\mathbf{h}(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}$.

The orthonormality restrictions for $\mathbf{X}$ and $\mathbf{B}$ can be written as

$$\mathbf{h}(\boldsymbol{\theta}_0) \quad = \quad \begin{pmatrix} \mathbf{X}^{\top}\mathbf{1}_{N\times 1} \\ \mathbf{vec}(\mathbf{X}^{\top}\mathbf{X}-N\mathbf{I}_K) \\ \widetilde{\mathbf{vec}}(\widetilde{\mathbf{B}}^{\top}\widetilde{\mathbf{B}}) \end{pmatrix}, \tag{1.11}$$

where the off-diagonal vector operator $\widetilde{\mathbf{vec}}$ converts a $K\times K$ symmetric matrix into a $K(K-1)$-dimensional vector consisting of off-diagonal elements of the input matrix. The first component indicates the column centering or zero intercept condition for $\mathbf{X}\widetilde{\mathbf{B}}^{\top}$, leading to zero mean constraint for each column of $\mathbf{X}$ (Jolliffe 2002). This column-centering constraint for $\mathbf{X}$ is naturally achieved due to the intercept matrix $\mathbf{A}$. The second component of $\mathbf{h}(\boldsymbol{\theta}_0)$ denotes the orthonormality condition for $\mathbf{X}$, and the last component indicates the orthogonality condition for $\widetilde{\mathbf{B}}$.

Calculating the Jacobian of this vector of restrictions, we obtain

$$\frac{d\mathbf{h}(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}^{\top} \quad = \quad \begin{pmatrix} \mathbf{I}_{K^2+K} & \mathbf{0} \\ \mathbf{0} & \mathbf{O}_K \end{pmatrix} \begin{pmatrix} \mathbf{I}_K & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{K^2}+\mathbf{K}_{KK} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{K^2}+\mathbf{K}_{KK} \end{pmatrix} \left(\begin{pmatrix} \mathbf{1} \\ \mathbf{X} \\ \widetilde{\mathbf{B}} \end{pmatrix}^{\top} \otimes \mathbf{I}_K \right) \tag{1.12}$$

where a $K^2\times K^2$ commutation matrix $\mathbf{K}_{KK}$ is a commutation matrix satisfying $\mathbf{K}_{KK}\mathbf{vec}(\mathbf{A}) = \mathbf{vec}(\mathbf{A}^{\top})$ for a $K\times K$ matrix $\mathbf{A}$, and the $K(K-1)\times K^2$ diagonal matrix $\mathbf{O}_K$ is the off-diagonal matrix operator which contracts a $K^2\times 1$ input vector (*i.e.* the vectorized input matrix) into a $K(K-1)\times 1$ output vector whose elements are the off-diagonal elements of the (vectorized) input matrix. Finally, the rank condition matrix (Magnus & Neudecker 1988, Rivers 2003)

becomes

$$\frac{d\mathbf{h}(\boldsymbol{\theta}_0)}{d\boldsymbol{\theta}}^{\top}\left(\begin{pmatrix}\mathbf{1}\\ \mathbf{X}\\ \widetilde{\mathbf{B}}\end{pmatrix}\otimes\mathbf{I}_K\right) = \begin{pmatrix}\mathbf{I}_{K^2+K} & \mathbf{0}\\ \mathbf{0} & \mathbf{O}_K\end{pmatrix}\begin{pmatrix}\mathbf{I}_K & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{K^2}+\mathbf{K}_{KK} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{K^2}+\mathbf{K}_{KK}\end{pmatrix}, \tag{1.13}$$

and, due to the rank of the commutation matrix (Magnus & Neudecker 1988), the rank of this matrix is

$$\begin{aligned}&\operatorname{rank}\begin{pmatrix}\mathbf{I}_{K^2+K} & \mathbf{0}\\ \mathbf{0} & \mathbf{O}_K\end{pmatrix}\begin{pmatrix}\mathbf{I}_K & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{K^2}+\mathbf{K}_{KK} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{K^2}+\mathbf{K}_{KK}\end{pmatrix}\\ &= K+K(K+1)/2+K(K+1)/2-K = K(K+1).\end{aligned} \tag{1.14}$$

Proposition 1 *MIRT without orthogonal constraints produces item parameter matrix* $\mathbf{B}$ *(*$\tilde{\mathbf{B}}$*) whose posterior mode spans the principal subspace of the augmented variable covariance matrix* $\mathbf{S} \equiv \mathrm{Cov}(\mathbf{z}_i, \mathbf{z}_j) \forall i, j \in \{1, \ldots, N\}$.

1.9.2 Proof

For notational convenience, let us first define the augmented variable column vectors $\mathbf{z}_i \equiv \mathbf{Z}_{[i,]}^{\top} \in \mathbb{R}^M \ \forall i \in \{1, \ldots, N\}$ and $\mathbf{x}_i \equiv \mathbf{X}_{[i,]} \in \mathbb{R}^K \ \forall i \in \{1, \ldots, N\}$.[4] For each individual, the augmented response vector can be written as,

$$\mathbf{z}_i = \tilde{\mathbf{B}}\mathbf{x}_i^{\top} + \boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \tag{1.15}$$

where the item parameter matrix $\tilde{\mathbf{B}} \in \mathbb{R}^{M \times K}$, offset column vector $\boldsymbol{\alpha} = -\sum_i \mathbf{z}_i / N \in \mathbb{R}^M$ and Gaussian disturbance vector $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ with $M \times M$ isotropic diagonal covariance

[4] The following proof is analogous to main results of PPCA (Tipping & Bishop 1999) and Bayesian PCA (Minka 2000).

matrix $\sigma^2\mathbf{I}$. Eq. 1.15 can be rewritten in the following multivariate normal distribution:

$$\Pr(\mathbf{z}_\mathrm{i}|\mathbf{x}_\mathrm{i}) \sim \mathcal{N}(\tilde{\mathbf{B}}\mathbf{x}_\mathrm{i}^\top + \boldsymbol{\alpha}, \sigma^2\mathbf{I}). \tag{1.16}$$

We can calculate the marginal distribution of $\mathbf{z}_i$ given the uniformly distributed $\Pr(\mathbf{x}_i) = \exp(-\mathbf{x}_i^\top \mathbf{x}_i)$ (Hoff 2007), and it is written as:

$$\begin{aligned} \Pr(\mathbf{z}_\mathrm{i}) &= \int \Pr(\mathbf{z}_\mathrm{i}|\mathbf{x}_\mathrm{i})\Pr(\mathbf{x}_\mathrm{i})\mathrm{d}\mathbf{x}_\mathrm{i} \\ &= \mathcal{N}(\boldsymbol{\alpha}, \mathbf{C}), \end{aligned} \tag{1.17}$$

where the covariance matrix $\mathbf{C} \equiv \tilde{\mathbf{B}}\tilde{\mathbf{B}}^\top + \sigma^2\mathbf{I}$.

The log likelihood of parameters for all N individuals can be written as:

$$\mathcal{L} = -\frac{N}{2}\{M\ln(2\pi) + \ln|\mathbf{C}| + \mathrm{tr}[\mathbf{C}^{-1}\mathbf{S}]\}, \tag{1.18}$$

where the augmented variable covariance matrix $\mathbf{S} \equiv \frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i - \boldsymbol{\alpha})(\mathbf{z}_i - \boldsymbol{\alpha})^\top$. By following the procedures of matrix differentiation (Magnus & Neudecker 1988), the stationary points of this function satisfies

$$\frac{\partial \mathcal{L}}{\partial \tilde{\mathbf{B}}} = N(\mathbf{C}^{-1}\mathbf{S}\mathbf{C}^{-1}\tilde{\mathbf{B}} - \mathbf{C}^{-1}\tilde{\mathbf{B}}) = \mathbf{0}. \tag{1.19}$$

This equation reduces to $\mathbf{S}\mathbf{C}^{-1}\tilde{\mathbf{B}} = \tilde{\mathbf{B}}$. Let us ignore trivial solutions $\tilde{\mathbf{B}} = \mathbf{0}$ and $\mathbf{C} = \mathbf{S}$, and reparameterize the remaining solution as $\tilde{\mathbf{B}} = \mathbf{U}\mathbf{L}\mathbf{V}^\mathbf{T}$ via singular value decomposition where $\mathbf{U}$ is the orthonormal matrix of left eigenvectors, $\mathbf{L}$ is the diagonal matrix of eigenvalues and $\mathbf{V}$ is the orthonormal matrix of right eigenvectors. By plugging in this expression to $\mathbf{S}\mathbf{C}^{-1}\tilde{\mathbf{B}} = \tilde{\mathbf{B}}$, we obtain $\mathbf{S}\mathbf{U}\mathbf{L} = \mathbf{U}(\sigma^2\mathbf{I} + \mathbf{L}^2)\mathbf{L}$. For K-dimensional model, this expression is

reduced to

$$\tilde{\mathbf{B}} = \mathbf{U}_K(\mathbf{L}_K - \sigma^2\mathbf{I})^{1/2}\mathbf{R}, \tag{1.20}$$

where $\mathbf{R}$ is an arbitrary $K \times K$ orthonormal matrix for rotation.

Under the uniform prior used in the proposed algorithm, because $(\mathbf{L}_K - \sigma^2\mathbf{I})^{1/2}$ and $\mathbf{R}$ are $K \times K$ matrices of full rank, we can confirm that the mean estimates of $\tilde{\mathbf{B}}$, $\text{span}(\tilde{\mathbf{B}}) = \text{span}(\mathbf{U}_K)$ where $\text{span}(\mathbf{U}_K)$ is K-dimensional principal subspace of the covariance matrix $\mathbf{S}$.

Proposition 2 *The proposed method produces item parameter matrix* $\mathbf{B}$ *(*$\tilde{\mathbf{B}}$*) whose posterior mode lies on the principal directions of the augmented variable covariance matrix* $\mathbf{S} \equiv \text{Cov}(\mathbf{z}_i, \mathbf{z}_j)\ \forall i, j \in \{1, \ldots, N\}$.

1.9.3 Proof

By using Proposition 1, We can further make a progress for the property of the proposed method. From Eq. 1.20, we can easily obtain the fact that the first dimensional mean estimate stretches along the principal axis of $\mathbf{S}$ since the orthonormal rotation matrix for $K = 1$ is uniquely determined as the constant $\mathbf{R} = 1$. Since the estimation process of the proposed method samples dimension-specific estimates sequentially, we can assure that the first dimensional estimate spans the first principal subspace of $\mathbf{S}$. Because the first dimensional direction is determined, the remaining $K - 1$ directions are uniquely determined due to the orthogonality restriction, so that $\mathbf{R} = \mathbf{I}_{K \times K}$.

1.9.4 Prior Specification

For Bayesian estimation of the parameters, $(\mathbf{X}, \mathbf{B}, \mathbf{w})$, we need to specify their priors and priors for sampling their hyperparmeters, (μ, ψ, ϕ). Orthonormal matrices $\mathbf{X}$ and

$\mathbf{B}$ have uniform priors on the Stiefel manifolds of size $V_K(\mathbb{R}^{N\times 1})$ and $V_K(\mathbb{R}^{M\times 1})$ respectively. Due to Eq. 1.6, we aware that the conditional distribution of k-th column of $\mathbf{X}$ or $\mathbf{B}$ given the rest of the columns follows a vMF distribution, with $\boldsymbol{\mu} = \phi w_k \mathbf{N}_{-k}^{\mathbf{X}\top}\mathbf{E}_{-k}\mathbf{B}_{[,k]}$ or $\boldsymbol{\mu} = \phi w_k \mathbf{X}_{[,k]}^{\top}\mathbf{E}_{-k}\mathbf{N}_{-k}^{\mathbf{B}}$, multiplied by a basis for the null space of a matrix stacked by the rest of the columns. Since uniform vMF distribution (*i.e.* $vMF(\mathbf{0})$) is conjugate to any vMF distribution of given length, we can incorporate $vMF(\mathbf{0})$ as conjugate prior of each column. The weight parameter, $\mathbf{w} \sim \mathcal{N}(\mu, 1/\psi)$, is sampled with a prior $\mathcal{N}(\mu_0, v_0^2)$ for μ and a prior $\Gamma(\eta_0/2, \eta_0\tau_0^2/2)$ for precision parameter ψ. Finally, a prior for precision ϕ is given as $\Gamma(\nu_0/2, \nu_0\sigma_0^2/2)$. For the six hyperparameters, we adopt the constant values given by Hoff (2007) for probabilistic singular value decomposition, so that $\eta_0 = 2$, $\nu_0 = 2$, and other values are computed using empirical Bayesian estimates, computed from the observed dataset itself. To be specific, using the initial values of $(\mathbf{X}, \mathbf{B}, \mathbf{w})$, we first compute dimension-specific components of (μ_0) as the mean value for the cumulative sum of the elements of $\mathbf{w}$ divided by its rank, so that $\mu_0 = \frac{1}{N+1}\sum_{k=0}^{N}\sum_{j=1}^{k} w_j/k$. Similarly, $\sigma_0^2 = \frac{1}{N+1}\sum_{k=0}^{N}\|\mathbf{Y} - \mathbf{X}_{[,1:k]}\mathbf{W}_{[1:k,1:k]}\mathbf{B}_{[,1:k]}^{\top}\|^2/NM$, $v_0^2 = \frac{1}{N+1}\sum_{k=0}^{N}(\sum_{j=1}^{k} w_j/k - \frac{1}{N+1}\sum_{k'=0}^{N}\sum_{j'=1}^{k'} w_{j'}/k')^2$ and $\tau_0^2 = \frac{1}{N+1}\sum_{k=0}^{N}(\sum_{j=1}^{k} w_j - \frac{1}{N+1}\sum_{j=0}^{N} w_j)^2/k$. To keep the notation simple, for the null rank case, in which the subscript k equals 0, the corresponding value is treated as 0 (*e.g.* $\sum_{j=1}^{0} w_j = 0$), or, if divided by its subscript, 1 (*e.g.* $\sum_{j=1}^{0} w_j/0 = 1$). When $M < N$, N in the hyperparameter equations can be replaced by M.

1.9.5 Computational Implementation

The proposed Gibbs sampling algorithm is implemented in MATLAB. For fast computation, we utilize Wood (1994)'s algorithm for sampling vMF distribution and Botev (2016)'s algorithm for sampling multivariate truncated normal distribution. Starting values for $(\mathbf{X}, \mathbf{B}, \mathbf{w})$ are set as the principal component outcomes of the double centered matrix of $\mathbf{YY}^{\top}$ and $\mathbf{A}$ is set as residual values. Setting random draws as starting values does not affect

the performance of the algorithm.

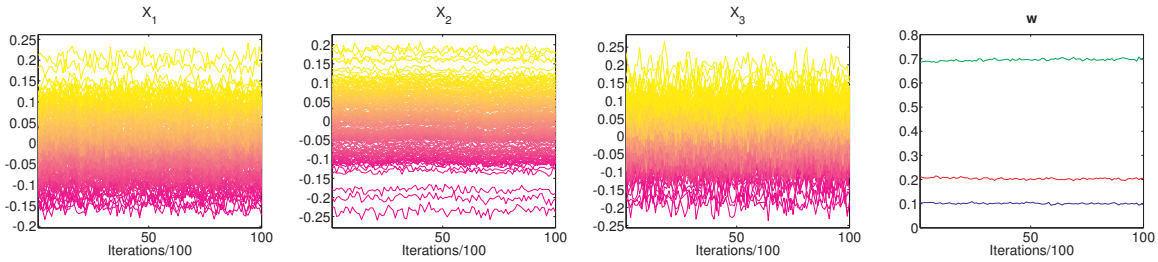

Figure 1.10: Mixing patterns. Convergence plots of estimates for a synthetic example with planted parameters: $N = 200$, $M = 1,000$, $\mathbf{w} = (0.7, 0.2, 0.1)$ and $\phi = 10^6$. The first three subplots indicate the dimension-specific outcomes of $\mathbf{X}$ over Gibbs sampling iterations, and the right most subplot indicates the dynamics of $\mathbf{w}$ over iterations.

Because the scale of utility is irrelevant (Train 2009), the precision (*i.e.* ϕ) of $\mathbf{E}$ affects the resulting estimates of $\mathbf{w}$ and $\mathbf{A}$ only up to constant multiplication. For an output $\hat{\phi}$, $\hat{\mathbf{A}}$'s corresponding $\hat{\mathbf{w}} = (\hat{\phi}/\phi)\mathbf{w}$ and $\hat{\mathbf{A}} = (\hat{\phi}/\phi)\mathbf{A}$. While setting ϕ as an arbitrary constant value does not affect the estimation results in theory, we find that incorporating ϕ estimation stage dramatically reduces the computational cost required for the convergence of other parameters.

Figure 1.10 shows the converging trends of the parameters over iterations. As illustrated, all estimates exhibit stable resonance around mean values. For all examples analyzed in the study, sampling outputs did not reject the null hypothesis in Geweke's test (Geweke 1991).

Chapter 1 is currently in preparation for submission for publication of the material. Sohn, Yunkyu. The dissertation author was the sole researcher and author of this material.

Chapter 2

Automatic Dimensionality Selection Method for Latent Variable Models: Application to Voting Datasets

Despite the prevalent use of latent variable models (*e.g.* factor models, principal component analysis, item response theory models) in political science and other fields of social sciences, there has been lack of approaches for identifying the number of dimensions underlying the generating process of data matrices. Here, I propose a changepoint detection method for the automatic discovery of noise-free spectral bands in the eigenspectra of covariance matrices. I first prove the benefit of using eigenvalue ratio statistic over eigenvalues, and propose an automatic discovery algorithm. The proposed method corrects flaws in the well-known automatic discovery method based on profile likelihood comparison, and replaces the conventional elbow-detection visual inspection scheme. I demonstrate the performance of the method using historical roll call votes of American Congress.

2.1 Problem Statement

Latent variable models (*e.g.* factor models, principal component analysis, item response theory models) have been a canonical approach for measurement in political science. Latent variable models have been used in many fields of political science, to name a few, including legislative (Poole 2005), electoral (Jackman & Vavreck 2010), institutional (Treier & Jackman 2008) and human rights studies (Fariss 2014). The outputs of these methods, lower dimensional latent variables, recover key aggregate information underlying the data generating process of high dimensional datasets.

An essential decision a researcher needs to make when using latent variable models is to determine the number of dimensions of the recovered latent traits, often expecting to identify the number of dimensions underlying the generative process of data matrices. However, there has been lack of methods for dimensionality selection. In fact, this problem prevails when using latent variable models across all disciplines of data science, not limited to political and social science research (Murphy 2012). Among many applications of latent variable models, I focus on identifying the number of dimensions for ideal point estimation of voting datasets over roll calls in parliaments.

Here, I propose a changepoint detection method for the automatic discovery of noise-free spectral bands in the eigenspectra of covariance matrices. I first prove the benefit of using eigenvalue ratio statistic over eigenvalues, and propose an automatic discovery algorithm. The proposed method corrects flaws in the well-known automatic discovery method based on profile likelihood comparison (Zhu & Ghodsi 2006), and replaces the conventional elbow-detection visual inspection scheme (Poole 2005). I demonstrate the performance of the method using historical roll call votes of American Congress. The proposed method is expected to provide a fast-check for guessing the true dimensionality of latent variables in a high dimensional dataset during pre-analysis stages.

2.2 Theoretical Framework

Before introducing the statistical method, I present the theoretical framework explicating why the eigenvalue ratio statistic produces desirable results for dimensionality selection. While the following discussion uses the vote correlation matrix example, the majority of results can be applied to any type of latent linear variable models in general. The following reasoning heavily relies on the results on the spectral properties of modular matrices by Bai & Silverstein (2010), Nadakuditi & Newman (2012) and Peixoto (2013).

Let us define the vote correlation matrix on bill t, $\tilde{\mathbf{V}}_t$ and $\mathbf{V} \equiv \sum_t \tilde{\mathbf{V}}_t$. Each vote correlation matrix $\tilde{\mathbf{V}}_t$ is defined as:

$$\tilde{V}_{ijt} = \tilde{V}_{jit} = \begin{cases} 1, & \text{if i and j vote same on } t \\ 0, & \text{otherwise.} \end{cases}$$

In the standard item response theory formulation, the probability of legislators i and j vote same on bill t is given as $\mathbf{Pr}(V_{ijt} = 1) = \Phi(\sum_k w_k \beta_{it} x_{it} - \alpha_t)\Phi(\sum_k w_k \beta_{jt} x_{jt} - \alpha_t) + (1 - \Phi(\sum_k w_k \beta_{it} x_{it} - \alpha_t))(1 - \Phi(\sum_k w_k \beta_{jt} x_{jt} - \alpha_t))$ where $\Phi(\cdot)$ is the standard cumulative normal density function evaluated $\cdot$. All other parameters are as defined in Chapter 1.

Instead of the double-centered agreement matrix which is used for dimensionality selection in voting profiles (Poole 2005), I study a standard linear operator matrix for quadratic cost representation called normalized graph Laplacian. Normalized graph Laplacian is a mathematically well-motivated structure of the operator which is proven to give the best K dimensional approximation of a data matrix among other representations of data covariance matrices (Chung 1997, Ng, Jordan & Weiss 2002, Rohe, Chatterjee & Yu 2011). Unlike the double-centered agreement matrix, normalized graph Laplacian has a symmetric bulk noise spectrum (*i.e.* Wigner semicircle distribution) that is tractable, different from the skewed noisy spectrum (*i.e.* Marcenko-Pastur Quarter semicircle distribution) of the double-centered

agreement matrix (Rao & Edelman 2008).

The normalized graph Laplacian for the aggregate vote correlation matrix $\mathbf{V}$ is defined as

$$\boldsymbol{\mathcal{L}} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{V}\mathbf{D}^{-1/2} \tag{2.1}$$

where $\mathbf{D} = \text{Diag}\sum_i V_{ij}$(Von Luxburg 2007). It is known that the eigenvalues and eigenvectors of this matrix correspond to the cost function and solutions minimizing the normalized graph cut (Von Luxburg 2007) which is used to partition the input correlation matrix or adjacency matrix $\mathbf{V}$ to modular entities.

Recently, theoretical progress has been made on the eigenvalue spectra of general classes of random matrices including normalized graph Laplacian ($\boldsymbol{\mathcal{L}}$) (Bai & Silverstein 2010, Nadakuditi & Newman 2013, Peixoto 2013). A normalized graph Laplacian generated from the vote correlation matrix $\mathbf{V}$ can be decomposed into two parts, each consisting of the fluctuation term ($\boldsymbol{\mathcal{X}}$) due to stochasticity in the data generating process and the mean term ($\langle\mathbf{M}\rangle$) representing the data generating parameters.

$$\boldsymbol{\mathcal{L}} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{V}\mathbf{D}^{-1/2} = \mathbf{I} - (\underbrace{\boldsymbol{\mathcal{X}}}_{\text{fluctuation}} + \underbrace{\langle\mathbf{M}\rangle}_{\text{mean}}) \tag{2.2}$$

In the case of vote similarity matrix, each element of $\boldsymbol{\mathcal{L}}$ is the normalized version of $\mathbf{Pr}(V_{ijt} = 1)$. Since $\mathbf{I}$ shifts all eigenvalues by +1, the eigenvalue spectrum of $\boldsymbol{\mathcal{L}}$ is same as $\mathbf{D}^{-1/2}\mathbf{V}\mathbf{D}^{-1/2}$ with -1 multiplied and +1 added.

This general representation can be understood more intuitively if we assume a truncated data generating matrix $\mathbf{M}_B$ of size $B \times B$. That is, now we assume that there are B unique voting probability profiles and each of N total legislators are assigned to the one of the B voting probability profiles. Note that, due to the stochasticity in the data generating process, two legislators having identical *voting probability profiles* can exhibit distinct *realized voting profiles*. Following this representation, $\mathbf{b}$ ($b_i \in \{1,\ldots,B\}$) is a block (*e.g.*

voting coalition) affiliation vector of size n, and the off-diagonal mean matrix $\langle M_{ij} \rangle = [\mathbf{M}_B]_{rs}$ with affiliation indexes for i and j respectively ($r = b_i$ $s = b_j$).

Due to Peixoto (2013), we get exact expressions for each element of the mean matrix and its corresponding variance for large n. That is, $[\mathbf{M}_B]_{rs} = e_{rs}/\sqrt{n_r e_r n_s e_s}$ and $\sigma^2_{rs} \simeq e_{rs}/e_r e_s$ where e_{rs} is the expected number of 1s the on r-s similarity compartment of $\mathbf{M}_B$ and e_r is the total expected number of 1s (*i.e.* agreements) for legislators belonging to group r.

On the basis of this result, we can calculate the eigenspectrum of $\boldsymbol{\mathcal{L}}$ (or $\mathbf{D}^{-1/2}\mathbf{V}\mathbf{D}^{-1/2}$) given $\langle \mathbf{M} \rangle$ and $\mathbf{b}$. Such progress can be made by decomposing the determinant of $\mathbf{D}^{-1/2}\mathbf{V}\mathbf{D}^{-1/2}$ as:

$$\det[z\mathbf{I} - (\boldsymbol{\mathcal{X}} + \langle \mathbf{M} \rangle)] = \underbrace{\det[z\mathbf{I} - \boldsymbol{\mathcal{X}}]}_{random} \underbrace{\det[\mathbf{I} - (z\mathbf{I} - \boldsymbol{\mathcal{X}})^{-1}\langle \mathbf{M} \rangle]}_{DGP}. \tag{2.3}$$

This decomposition strikingly simplifies the spectrum of $\mathbf{D}^{-1/2}\mathbf{V}\mathbf{D}^{-1/2}$ as the superposition of the random eigenvalues and data generating process (DGP) eingevalues.

Figure 2.1 illustrates this intuition. The goal of spectral decomposition methods in general is to separate the bulk noise spectrum coming from the random noise part (*i.e.* the first term of the right hand side) and relevant eigenvalues coming from the DGP part (*i.e.* the second term of the right hand side). In the figure, the relevant eigenvalues correspond to the eigenvalues of the DGP relevant matrix $\langle \mathbf{M} \rangle$ and the random eigenvalues correspond to the rest of eigenvalues in $\mathbf{V}$, coming from $\boldsymbol{\mathcal{X}}$. Among the two operators, normalized graph Laplacian $\mathcal{L}(\cdot)$ exhibits superior performance as opposed to unnormalized graph Laplacian $\mathbf{L}(\cdot)$. That is, whereas the majority of the relevant eigenvalues in $\mathbf{L}(\mathbf{V})$ are absorbed into the eigenvalues of the bulk noise part, all random eigenvalues in $\mathcal{L}(\mathbf{V})$ are concentrated around the trivial eigenvalue 1. To recap, our goal is to count the number of eigenvalues that attribute to DGP of the vote correlation matrix using a good operator such as the normalized graph Laplacian.

Further analytic approximation can be made using complex analysis. After applying

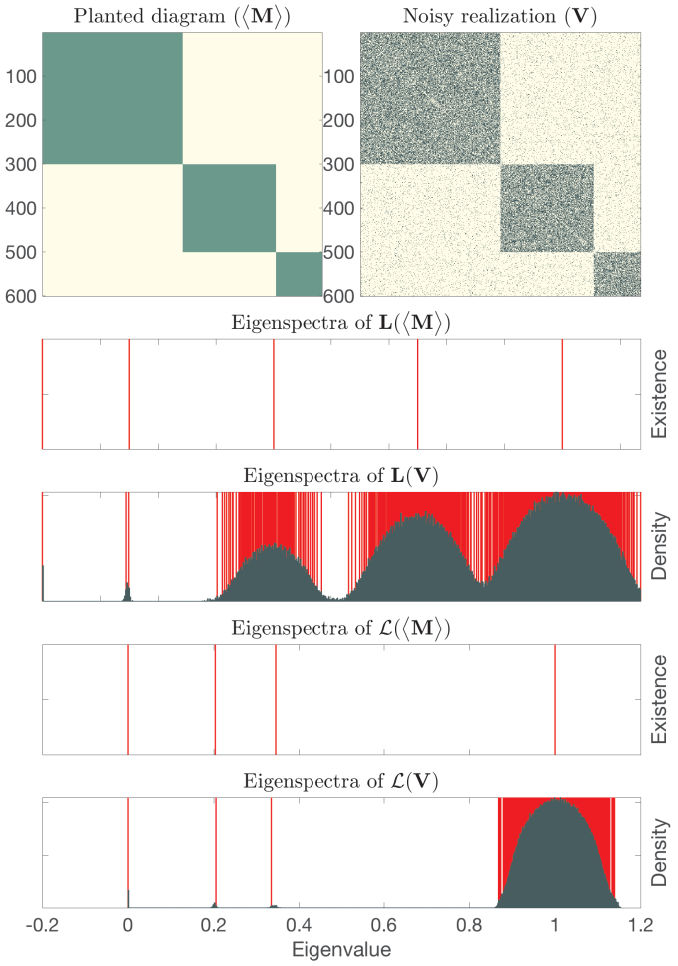

Figure 2.1: Random matrix theory intuition. Schematic illustration of the idea using three voting coalition example consisting of 600 voters. Vote similarity planted diagram ($\langle \mathbf{M} \rangle$) and realized vote correlation matrix ($\mathbf{V}$) where dark dots indicate the existence of positive correlation and light dots indicate the absence of such correlation. $\mathcal{L}(\cdot)$ is the normalized graph Laplacian operator whereas $\mathbf{L}(\cdot)$ is the unnormalized graph Lapalcian operator. Each eigenspectra subplots indicate the distribution and density of eigenvalues. For the eigenspectra of the realized vote correlation matrix ($\mathbf{V}$), in order to obtain the density values, I simulated 100 realizations of the vote correlation matrix with same data generating process, and averaged their occurrence.

the Stielties transform to $\boldsymbol{\mathcal{X}}$, the spectrum of the noisy matrix equals

$$\rho(z) = -\frac{1}{N\pi}\text{Im}\text{Tr}\langle (z\mathbf{I} - \boldsymbol{\mathcal{X}})^{-1} \rangle = -\frac{1}{N\pi}\sum_r n_r \text{Im} t_r(z) = -\frac{1}{N\pi}\sum_r n_r \text{Im}\frac{1}{z - \sum_s \sigma_{rs}^2 n_s t_s(z)} \tag{2.4}$$

with $t_r(z) \equiv \langle [(z\mathbf{I} - \boldsymbol{\mathcal{X}})^{-1}]_{ii} \rangle = \frac{1}{z - \sum_s \sigma_{rs}^2 n_s t_s(z)}$ for $i \in r$ (Bai & Silverstein 2010, Nadakuditi & Newman 2012, Peixoto 2013).

Here I show the relationship between the solutions ($z*$) of $\det[\mathbf{I} - (z\mathbf{I} - \boldsymbol{\mathcal{X}})^{-1}\langle \mathbf{M} \rangle] = 0$ and the eigenvalues of $\mathbf{M}_B\mathbf{N}$ where $\mathbf{N}$ is the diagonal group size matrix of size B so that $\forall r \in \{1, \ldots, B\}, N_{rr} = n_r$. Let us first show that the solutions of $\det[\mathbf{I} - (z\mathbf{I} - \boldsymbol{\mathcal{X}})^{-1}\langle \mathbf{M} \rangle] = 0$ approximate to eigenvalues of the $B \times B$ matrix $\mathbf{T}(z)\mathbf{M}_B\mathbf{N}$, since

$$\det[\mathbf{I} - \langle (z\mathbf{I} - \boldsymbol{\mathcal{X}})^{-1} \rangle \langle \mathbf{M} \rangle] = \det[\mathbf{I} - \mathbf{T}(z)\langle \mathbf{M} \rangle] = \det[\mathbf{I}_B - \mathbf{T}(z)\mathbf{M}_B\mathbf{N}] = 0, \tag{2.5}$$

where $\mathbf{T}(z)$ is a $B \times B$ diagonal matrix containing the values of $t_r(z)$ and the second equality comes from the fact that Gaussian elimination of $\mathbf{T}(z)\langle \mathbf{M} \rangle$ for duplicate rows. We can now apply the approximation $t_r(z) \simeq 1/z$ for detached eigenvalues from the bulk eigenvalues coming from the noisy matrix $\boldsymbol{\mathcal{X}}$ (Nadakuditi & Newman 2012, Peixoto 2013), which leads to

$$\mathbf{T}(z)^{-1}\mathbf{I}_B - \mathbf{M}_B\mathbf{N} \simeq z\mathbf{I}_B - \mathbf{M}_B\mathbf{N}. \tag{2.6}$$

Finally, we obtain

$$z|\det[\langle (z\mathbf{I} - \boldsymbol{\mathcal{X}})^{-1} \rangle \langle \mathbf{M} \rangle] = 0 \simeq z|\det[z\mathbf{I}_B - \mathbf{M}_B\mathbf{N}] = 0 \tag{2.7}$$

2.3 Proposed Method

To recap, the goal is to devise a method which separates the eigenvalues coming from the disturbance term of the stochastic realization and the relevant eigenvalues due to the coalitional structure underlying DGP. As illustrated in Figure 2.1, the normalized graph Laplacian is a very nice operator that can be used to discriminate noisy eigenvalues and relevant eigenvalues. Using the theoretical framework introduced so far, I introduce an automatic discovery algorithm which counts the number of the relevant eigenvalues.

2.3.1 Eigenvalue Ratio Statistic

In practice, because we do not know the data generating matrices $\mathbf{M}_B$ and $\mathbf{N}$ when observing a realized voting matrix, it is infeasible to directly utilize the analytic results introduced above for developing empirical statistical methods. A breakthrough can be made by focusing on the interval characteristics and separability of the eigenvalues of $\boldsymbol{\mathcal{L}}(\mathbf{V})$ consisting of eigenvalues originating from the noisy matrix $\boldsymbol{\mathcal{X}}$ and the data generating matrices $\mathbf{M}_B$ and $\mathbf{N}$.

To be short, we can use the fact that eigenvalue bulk of $\boldsymbol{\mathcal{X}}$, $\rho(z)$, is bounded by a confined range of which edges are solutions $(z*)$ for $\rho(z) = 0$ and $\det[\mathbf{I}_B - \mathbf{J}(z)] = 0$ where the Jacobian $J_{rs}(z) = \partial t_r / \partial t_s$ (Nadakuditi & Newman 2012, Peixoto 2013). Since the solutions of this set of equations, $z \in \{z_L, z_R\}$, where z_L is the left boundary solution and z_R is the right boundary solution, equal the edges of $\rho(z)$, and we do know that $z_R - z_L$ is finite, for large N (and assuming the matrix $\mathbf{V}$ is full rank, thus $|\{z|z \in \rho(z)\}| = N - \text{rank}(\mathbf{M}_B)$):

1. eigenvalue interval statistic: $\Delta z \approx \frac{z_R - z_L}{N-B} | z \in \rho(z) \to 0$
2. eigenvalue ratio statistic: $z_{i+1}/z_i | z \in \rho(z) \to 1$.

Now if we focus on eigenvalues of $\boldsymbol{\mathcal{L}}$, or equivalently $\mathbf{D}^{-1/2}\mathbf{V}\mathbf{D}^{-1/2}$, in the range of $z \notin \rho(z)$, we can easily see that $\Delta z \neq 0$ and $z_{i+1}/z_i \neq 1$. Thus, by using this fact, we can expect that the *regime change* of the both statistics occurs on the boundary of $z \in \rho(z)$ and $z \notin \rho(z)$. For the very special case of homogeneous block structure with exactly same diagonal elements for all groups of $\mathbf{M}_B$ and off-diagonal elements for all groups of $\mathbf{M}_B$ and homogeneous group size, $\rho(z) = \sqrt{4N\sum_s \sigma_{rs}^2/B - (z-1)^2}/(2\pi N \sum_s \sigma_{rs}^2/B)$, corresponding to an exact semi-circle distribution (Peixoto 2013). Hence we do know the exact finite bounds of the *continuous* bulk spectrum and that both statistics exhibit discrete transitions in their values around the boundary of the bulk and DGP eigenvalues. As we will see later, I find that the ratio statistic gives superb results than the well-known automatic method (Zhu

& Ghodsi 2006) based on the interval statistic for finite sized voting matrices.

2.3.2 Changepoint Detection for Dimensionality Selection

On the basis of the properties of the eigenvalue interval and ratio statistics, we can think of ways to detect the boundary where z changes from $z \in \rho(z)$ to $z \notin \rho(z)$. By using this idea, we can count the number of eigenvalues belonging to $z \notin \rho(z)$, thus the dimensionality of the utility embedded in the data generating multidimensional utility function (introduced in Chapter 1).

Since now we do know that the ratio statistic for $z_{i+1}/z_i | z \in \rho(z)$ converges to 1, unlike standard changepoint detection algorithm (Chib 1998), which is agnostic about the mean and variance parameters of one of the regimes, we can fix the ratio statistic for the *bulk regime* have mean 1 with a very small variance. Also the number of regimes equals 2, since we do know that the eigenvalues for $\boldsymbol{\mathcal{L}}$ exist in the range of $[0, 1+\varepsilon]$ due to the trivial characteristic of voting matrices (*i.e.* higher intra-coalition voting similarity propensity than inter-coalition voting similarity), where $1+\varepsilon$ is the right edge of the bulk eigenvalue spectrum. Accordingly, all relevant eigenvalues lie in the range of $(0, 1-\varepsilon)$.

In order to compare the performance of the new formalism based on eigenvalue ratio statistic with well-known automatic detection method based on eigenvalue interval (Zhu & Ghodsi 2006), I use profile likelihood based changepoint detection algorithm (Murphy & Van der Vaart 2000, Murphy 2012). Different from the proper likelihood approach, profile likelihood utilizes pooled variance so that the nuisance parameter variance is fixed.

Following the procedure introduced in Zhu & Ghodsi (2006), the goal of inference is to partition the eigenvalues or eigenvalue statistics into two groups (Λ_1 and Λ_2). According to the discussion so far, these two groups respectively correspond to the relevant eigenvalue set and the random eigenvalue set. Due to the incorporation of profile likelihood, the variance parameter is fixed for both of the sets as σ^2. The Gaussian mixture log likelihood function

is defined as

$$l(q) = \sum_{i \in \Lambda_1} \log \phi\left(\frac{d_i - \mu_1}{\sigma}\right) + \sum_{i \in \Lambda_2} \log \phi\left(\frac{d_i - \mu_2}{\sigma}\right), \tag{2.8}$$

where q is the number of consecutive eigenvalue statistics $\{d_i | 1 \leq i \leq N\}$ (*e.g.* eigenvalues, eigenvalues ratio statistics) belonging to Λ_1 and $\phi(\cdot)$ is the standard Gaussian density at $\cdot$. The goal of this procedure is to find q which maximizes $l(q)$ with $\hat{\mu}_1 = \frac{\sum_{i \in \Lambda_1} d_i}{q}$ and $\hat{\mu}_2 = \frac{\sum_{i \in \Lambda_2} d_i}{N-q}$. The pooled variance is estimated as

$$\hat{\sigma}^2 = \frac{(q-1)s_1^2 + (N-q-1)s_2^2}{N-2}, \tag{2.9}$$

where s_1^2 and s_2^2 correspond to the sample variance of the partitions respectively. This method is designed to detect a boundary which gives the best fit of the bimodal Gaussian mixture distribution.

For eigenvalue ratio statistic version, I set the mean value of the second set as $\mu_2 = 1$ since we do know the converging limit. In addition, since the variance of ratio statistics belonging to the second set should be very small relative to the pooled variance or the variance of the first set ratio statistics, I set the variance parameter of the second set to be fixed as $\hat{\sigma}^2/100$. Accordingly, the mixture log likelihood for the ratio statistic changepoint equals

$$l(q) = \sum_{i \in \Lambda_1} \log \phi\left(\frac{d_i - \mu_1}{\sigma}\right) + \sum_{i \in \Lambda_2} \log \phi\left(\frac{d_i - 1}{\sigma/10}\right). \tag{2.10}$$

Since we are using eigenvalue ratio statistic instead of eigenvalues, the size of the input truncates by 1, so that $\{d_i | 1 \leq i \leq N-1\}$.

2.4 Simulation Evidence

Figure 2.2 compares the performance of the proposed eigenvalue ratio statistic approach with the naive eigenvalue changepoint detection method of Zhu & Ghodsi (2006).

The figure shows two simulation cases with a low noise variance (top) and a high noise variance (bottom) for the quadratic voting model introduced in Chapter 1. The true dimensionality used to generate the vote matrix was 3. While both cases do have eigenvectors recovering the true planted ideal points, the coresponding eigenvalues are difficult to be discriminated from the noise eigenvalue bulk in the difficult voting case. This problem affects the result of the naive eigenvalue changepoint detection algorithm of Zhu & Ghodsi (2006). Yet, when using the eigenvalue ratio statistic, we can see visually the clear distinction between the noise bulk and the DGP eigenvalue ratio values. Accordingly, the proposed automatic detection approach provides correct results.

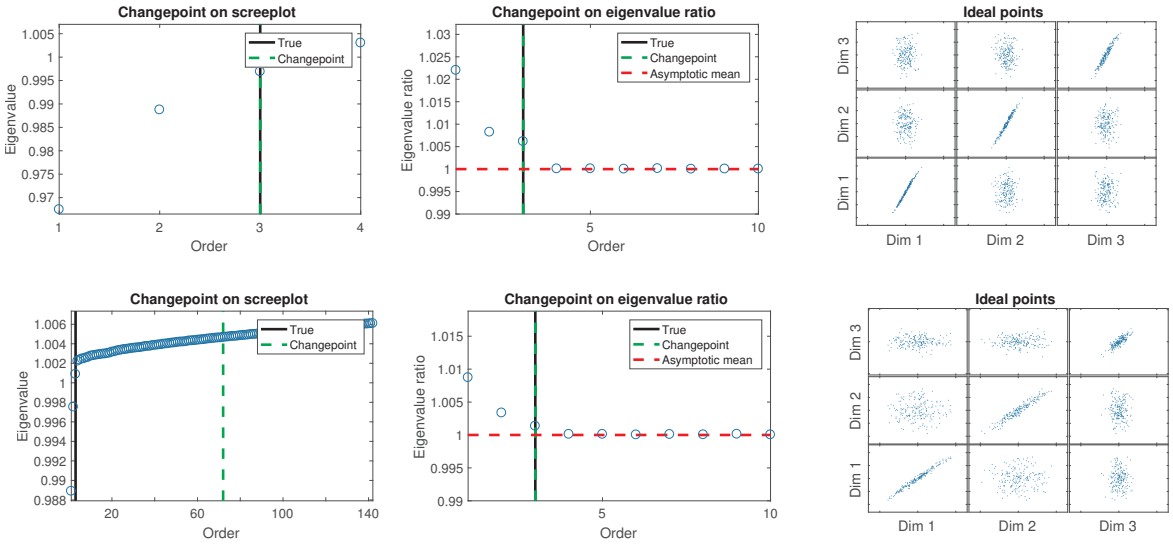

Figure 2.2: Simulation examples. [**Top**] Easy (less noisy) voting case with small voting error relative to the amount of utility difference. [**Bottom**] Difficult (noisy) voting case with large voting error relative to the amount of utility difference. The variance of the noisy case was 4 times of the variance of the less noisy case. Two simulation examples have same parameter setup, with 200 legislators and 2,000 bills with 3 dimensional underlying space and dimension importance: 0.5:0.3:0:2, except for the noise amount. The naive eigenvalue changepoint detection approach ('Changepoint on screeplot') proposed by Zhu & Godshi (2006) cannot detect the true dimensionality 3 whereas the proposed eigenvalue ratio changepoint detection method ('Changepoint on eigenvalue ratio') detects the correct value 3. All eigenvectors show significant levels of correlation with the truth ('Ideal points').

Figure 2.3 demonstrates results over a wide range of disturbance level. As shown

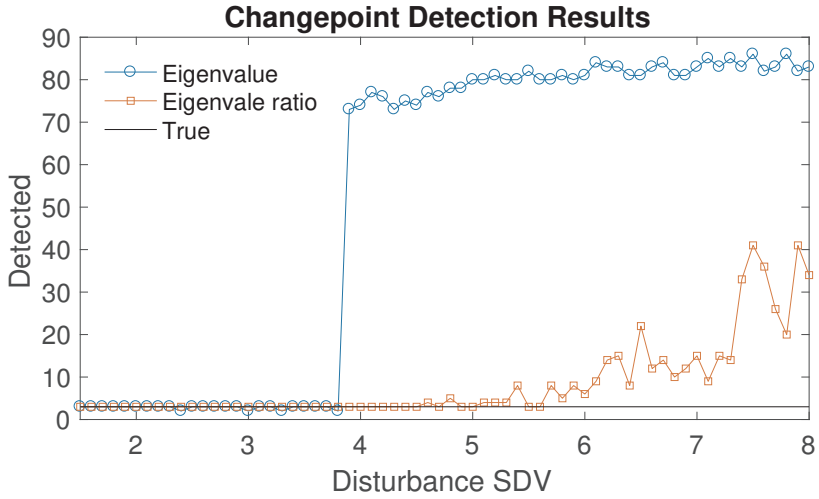

Figure 2.3: Changepoints with varying disturbance. Changepoint detection results for the same DGP parameters as in Figure 2.2 except noise variance. X-axis denotes the Gaussian noise standard deviation level. Each value should read with 10^{-3} multiplied. See Chapter 1 for details of the DGP model. 'Eigenvalue' denotes results obtained using the naive eigenvalue profile likelihood comparison method whereas 'Eigenvalue ratio' indicates the results obtained using eigenvalue ratio inputs. Eigenvalue ratio method recovers the true dimensionality (*i.e.* 3) for the majority of disturbance levels.

over the varying level of variance, the true dimensionality 3 is obtained for the range of disturbance levels when using eigenvalue ratio whereas the naive original approach ('Eigenvalue') of Zhu & Ghodsi (2006) never gets the true value when the standard deviation approaches 4×10^3. Unlike the original algorithm, which gives extremely high value in the majority of disturbance levels analyzed, the proposed ratio based algorithm gives correct values for the most of the disturbance levels below 5×10^3, and it gradually increases afterwards. The latter is in contrast with the result of the original algorithm which outputs unreasonably high values around these levels.

2.5 Application to United States House of Representatives Roll Call Voting Records

I apply the proposed method to the roll call voting records of the United States House ranging from the late 19 century to the present. I apply the changepoint detection algorithm on the eigenvalue ratio statistics of normalized graph Laplacians of the vote correlation matrices, and find that the number of detected dimensions has been greater than 2 during a significant number of sessions (Figure 2.4).

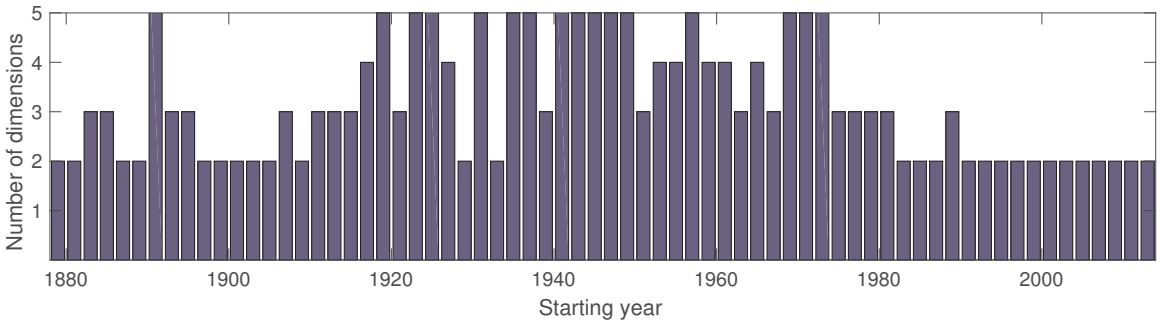

Figure 2.4: U.S. Houe Dimensionality. Number of dimensions estimated using the proposed changepoint detection algorithm applied to eigenvalue ratio statistics of normalized graph Laplacians of roll call voting correlation matrices from the 46th to the 113th session of United State House of representatives. Y-axis represents the estimated number of dimensions computed using the eigenvalue ratio statistic changepoint detection method and X-axis shows the starting year of each session.

Readers should note that one drawback of the current version of the proposed approach is that it does not allow us to differentiate between the cases where there is a single estimated dimension and two estimated dimensions. That is, due to the construction of the likelihood function that a single data point set is not allowed to have variance, the minimum possible value of the maximum likelihood solution for $l(q)$ is 2. Ongoing research aims to resolve this problem by incorporating Bayesian model selection procedure for comparing the 2-regime changepoint model marginal likelihood with null-changepoint (*i.e.* single regime) model marginal likelihood using the methods detailed in the chapter 7 of Gelman et al. (2014).

In addition, the users of the proposed method should also note that this approach provides a fast check of the underlying dimensionality since it is a truncated deterministic version of the probabilistic statistical model. Hence while this method can be used as a suggestive guideline for knowing the approximate true dimensionality of the underlying model, rigorous statistical judgment should be given carefully using DGP-based models such as the one introduced in Chapter 1. Thus it is recommended to set the input dimensionality values of the statistical model in a range including the inferred dimensionality obtained using the proposed method.

2.6 Conclusion

The changepoint detection algorithm for dimensionality selection replaces the conventional visual elbow detection approach for eigenvalues of the voting agreement matrix, which is vulnerable to subjective judgment. The new approach makes the selection process automatic, not relying on researchers' discretion on selecting the number of dimensions. Empirical analysis suggests evidence on high dimensionality of the US House roll call voting dataset over the late 19th century to the present. The proposed method is expected to serve in the pre-analysis stage when researchers do not have any idea about the true dimensionality of the latent variables in a high dimensional dataset. The resulting estimate can be used as a reference point for input dimensionality when conducting the analysis with a rigorous parametric model reflecting realistic DGP.

Chapter 2 is currently in preparation for submission for publication of the material. Sohn, Yunkyu. The dissertation author was the sole researcher and author of this material.

Chapter 3

A Unified Framework for Analyzing Aggregate and Issue-specific Preference from Non-voting Datasets: Coalitional Item Response Theory Model

Position-taking motivated individuals, who are uncertain about the ideological leaning of other individuals, will form coalitions based on the expectation for ideological characteristics of potential coalitions. I use this fact to build an ideological signaling model of coalition formation in which individuals (e.g. legislators, interest groups) pursue stable biding coalitions best serving their ideological interest. The binding coalitions, core, lead to an exact derivation of a specific ideal point inference method. I present three types of solutions for the obtained statistical model to recover spatial preference of actors from empirical coalition datasets: 1) exact solution 2) fully Bayesian solution, and 3) variational Bayesian solution. All solutions offer multi-dimensional estimates. The fully Bayesian solution provides probabilistic estimates, and the variational Bayesian solution specifies a

truncation threshold for selecting the number of dimensions. I apply the method to newly collected bill proposal coalition data covering the entire history of Korean National Assembly, from which individual ideology extraction is otherwise impossible due to the absence of recorded voting institution until the very recent sessions. Compared to NOMINATE on roll call voting data, the proposed method controls for government opposition and seniority bias. As the first to uncover the historical legislative landscape of Korea, the result provides a rare depiction of the evolutionary path of a parliament under dynamic political environments (i.e. trusteed, authoritarian and democratic; parliamentary and presidential) from its birth to the present. Finally, I extend the basic ideal point estimation framework to accommodate issue inference of bills using bill text information, and provide a framework for understanding the issue trend and issue-specific ideology from non-voting datasets. All data used for analyzing Korean National Assembly will be available as `Korean Legislation and Election Database`. All ideology and issue estimation algorithms will be available as `Coalitional Item Response Theory` programming package.

3.1 Introduction

3.1.1 Theoretical Motivation

Coalition formation is an defining element of political practice (Riker 1962). Individuals assemble groups and assert coordinated actions due to the economy of scale. This creates the needs for collective action and produce collective goods. Even self-interested decisions often lead to such outcomes (Axelrod 1981, Nowak 2006, Ostrom 1990, Putnam, Leonardi & Nanetti 1994). Then, why not always observe grand coalitions with everyone involved? An opposing effect of coalition formation is largely due to preference diversity among the group members (Axelrod 1967, Laver & Shepsle 1996, Laver & Schofield 1998, Demange 2005). Conflict of interest hampers a full exploitation of coordination and motivates the breaking

of society into small self-sufficient groups. The following examples show how this takes place in reality.

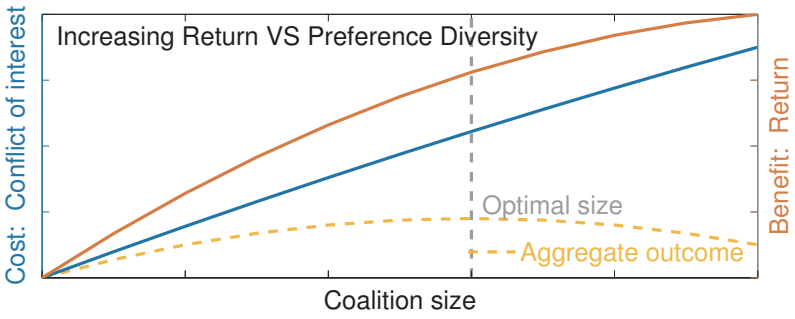

Figure 3.1: The Trade-off of Increasing Return and Preference Heterogeneity. Increasing return by group size yields benefit whereas preference diversity incurred by getting along with too many individuals causes conflict of interest. As a result there can be an optimal coalition where the trade-off between the two conflicting effects yields maximum payoff.

Case 1 Party Coalition Riker (1962) originally formulated the incentive for forming coalitional government in parliamentary democracies being proportional to the number of seats each party can get in the cabinet. This reasoning directs to minimum winning coalition idea in which its members maintain majority status with the minimum possible number of legislators. However one crucial reason this does not happen in reality is due to ideological heterogeneity among parties (Strøm & Nyblade 2007). That is, at the cost of disorganizing winning coalitions, a leading party would abandon other parties as coalition partners due to potential conflict of interest in policy making and misrepresentation of their ideology toward the public.

Case 2 Big-City Bills in American State Legislatures Gamm & Kousser (2013) find that the proportion of bills that pass state legislatures decreases with city size. One of the critical barriers for the success of big-city bills was large legislative delegation. They find when there are multiple legislators to represent a single city, internal divide among the legislators prevents producing a coherent and focused voice, thereby increasing the chance of failure.

Case 3 Inter-Firm Tariff Competition Product differentiation is a key to determining the size of lobbying coalitions in tariff determination in United States. Kim (2017) finds if

an industry has a higher level of product differentiation (e.g. electronic products), it is more likely to split the size of coalitions for reducing tariff rates. In other words, if an industry experiences conflict of interest in tariff determination due to product differentiation, it becomes rare to observe large lobbying coalitions.

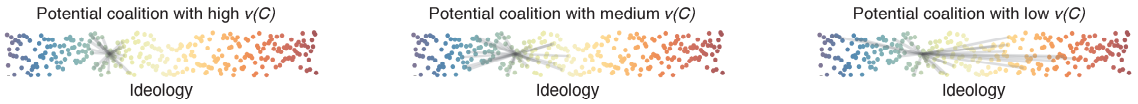

Figure 3.2: Utility of coalition formation. The utility $v(C)$ of forming a narrow coalition vs a broad coalition (C) under the position taking framework. Grey lines indicate coalitional partnership. Color gradation indicates ideological diversity in which dark blue corresponds to very liberal and dark red indicates very conservative. For position-taking motivated individuals, forming coalitions with like-minded individuals will provide a high level of utility (left) whereas forming coalitions with different-minded individauls will provide a low level of utility (right).

In recognition of the strategic nature of coalition formation, we can expect that realized coalitions convey important information on the latent preference of individuals (Figure 3.2). To best serve the narrow and specific incentive of individuals, self-interested decision makers will form groups, at the same time satisfying the size of group required for producing collective goods. For example, voting coalitions in parliaments can be formed for passing a bill but its members would not have an excessive level of ideological conflict with each other.

3.1.2 Statistical Method

In line with this understanding, political science community has developed a field of measurement theory which aims to discover latent ideological preference of decision makers from voting coalition datasets (Brady 1989, Clinton, Jackman & Rivers 2004, Poole 2005). These models build upon spatial models of choice in which decision makers with similar preference choose similar items. In particular, the most successful application of this idea

was to apply the method on parliamentary voting records where individuals face binary agenda comparison decisions with a utility assessment scheme. This low-dimensional projection approach has shaped a canonical framework for conducting empirical political science research, and has been successfully applied to voting records of multiple institutional bodies (Poole & Rosenthal 2000, Martin & Quinn 2002, Dewan & Spirling 2011).

However, in practice, voting is a comparatively rare and restricted opportunity for legislators to appeal their ideology to extra-legislative audience, especially among minority party members. Other means, such as bill proposal (Fowler 2006, Desposato, Kearney & Crisp 2011, Alemán, Calvo, Jones & Kaplan 2009), formal and informal opinion expression (Gentzkow & Shapiro 2010, Kim, Londregan & Ratkovic 2015), and physical violence (Kim 2011), are effective practices for position-taking that are less affected by known biases in voting (e.g. party effect and strategic opposition).

Coalition formation is a general and pervasive theoretical concept which puts these different types of behavior together. However, a well-motivated ideal point estimation method, especially from rigorous coalitional game theoretic foundation, has been missing. Although several non-parametric methods can be used to serve the same purpose, behavioral rationale underlying the use of these methods has been weak.

In this paper, I use the aforementioned strategic nature of coalition formation, which motivates to form coalitions with minimum ideological conflict and sufficiently large coalition size, to propose an ideological signaling model of coalition formation. The coalitional stable solution of this model, core, leads to an exact derivation of a specific factor-analytic type model for ideal point estimation.

I present three types of solutions for the obtained statistical model to recover spatial preference of actors from empirical coalition datasets: 1) exact solution, 2) fully Bayesian solution and 3) variational Bayesian solution. All solutions offer multi-dimensional estimates. The fully Bayesian solution provides probabilistic estimates. The variational Bayesian

solution specifies a truncation threshold for selecting the number of dimensions.

The method is applied to newly collected bill proposal coalition data covering the entire history of Korean National Assembly, from which individual ideology extraction is otherwise impossible due to the absence of recorded voting institution until the very recent sessions. Compared to NOMINATE on roll call data, the proposed method controls for government opposition and seniority biases. As the first to uncover the historical legislative landscape of Korea, the result provides a rare depiction of the evolutionary path of a parliament under dynamic political environments (i.e. trusteed, authoritarian and democratic; parliamentary and presidential) from its birth to the present.

I extend the basic estimation framework to infer issue-specific ideology through bill-issue matching. The proposed estimation method is flexible enough to apply it to issue-specific subsets. Through case analysis, I also show that the proposed method can be used to promote our understanding of legislative and party politics under various political conditions. Finally, I discuss the possibility of using the proposed framework to analyze ideological characteristics of political actors from various types of non-voting datasets.

3.2 An Informational Rationale for Coalition Formation

3.2.1 The Game

I consider a game G^b for a pair (N, v) and event $b \in B$ (*e.g.* proposing bill b among the set of all bills B), where the set of agents is $N = \{1, \ldots, n\}$ and $v : 2^n \to \mathbb{R}$ is a characteristic function, denoting the collective payoff of forming a coalition. This function assigns a real number (value) $v(C)$ to a coalition $C \subseteq N$. I assume that one can summarize all information given prior to the coalition formation for event b as s^b. The shape of utility curve is characterized by a threshold value $\underline{n}$. For instance, in the case of bill proposal coalition formation, s^b can be sponsor ideology, and $\underline{n}$ can be a minimum size barrier for

each legislative proposal. We can conceive s^b as taking the role of formateur party in party coalition formation example, who has an agenda power for shaping the coalition and cannot be excluded in the event b. This formulation yields a conditional representation of the characteristic function as $v(C|s^b, \underline{n})$.

As I will detail in the following, the structure of utility is better modeled as non-transferable with only a single possible *fair* division option for $v(C|s^b, \underline{n})$ into $\mathbf{u}^C$, where each element in $\mathbf{u}^C$ is utility share of $i \in C$. Hence individual utility vector becomes $\mathbf{u}^C = \langle u_1, \ldots, u_{|C|} \rangle$ and the collective payoff becomes $v(C|s^b, \underline{n}) = u^C \equiv \sum_{i \in C} u_i$.

3.2.2 Characteristic Function

Now, I detail the structure of the characteristic function v for realized coalitions among the individuals N. Suppose a position-taking-motivated decision maker (e.g. legislator) i who seeks to collaborate with $|C| - 1$ other individuals. By position-taking motivated individual, I denote an individual who cares more about signaling one's ideology toward his audience (e.g. via bill proposal to one's district electorates) rather than actual policy implementation (e.g. as a consequence of bill passage in the legislation application) (Mayhew 1974).

Let i's median audience preference (e.g. district preference) x_i. i's payoff becomes a function of i's ideology captured by the electorates $\hat{x}_i$ and the actual audience median x_i. Constituents' perception of i's ideology involves Gaussian uncertainty component and the voters infer i's ideology by integrating a prior distribution $\mathcal{N}(x_i, \sigma_0^2)$ and ideological signals of i's collaborators. In other words, I assume i's constituents have an ambiguous idea about i's ideology which spreads around their district median with standard deviation σ_0. i's audiences infer i's ideology using the prior information as well as the ideological characteristics of the members of the coalition i is participating. Note that these other members ideological locations are simultaneously estimated using the same procedure

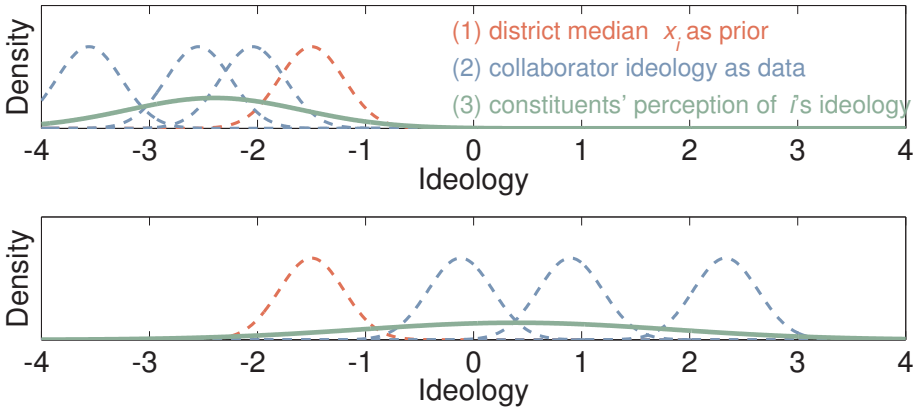

Figure 3.3: Informational contents of coalition formation. Constituents' perception of their district incumbent i's ideology. Given district median ideology as prior, constituents gather i's collaborators' expected ideological characteristics to infer i's ideology. Top and bottom panels denote two cases in which collaborator ideological preferences differ substantially. There always exists the risk of being misrepresented by forming coalitions. All lines indicate probabilistic density functions.

by the electorates. In other words, the electorates do not have fixed knowledge on their locations.

Figure 3.3 illustrates two distinct examples of ideological signaling as the results of legislative coalitions. The perceived ideology of legislator i in the top and the bottom panels differ substantially. In the top, i forms a coalition with like-minded individuals, and the consequential ideological signal is relatively close to i's district median. On the contrary, in the bottom, i collaborates with a number of legislators with relatively diverse and heterogeneous ideological characteristics. The realized signal of this coalition is diffusive whose mean largely deviates from i's district preference.

We can formulate the above conception exactly using the following likelihood function. The realized, maximum a posteriori estimate of i's ideology (i.e. i's ideology inferred by the voters) is equivalent to

$$\hat{x_i} = \underset{x}{\operatorname{argmax}} \left\{ \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2\sigma_0^2}(x_i - x)^2} \prod_{j \in C \setminus i} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_j - x)^2} \right\}, \tag{3.1}$$

where σ_0 and σ respectively denote standard deviations for i and i's collaborators' prior ideological signals. Calculating $\hat{x_i}$ by solving equation 3.1, we obtain

$$\hat{x_i} = \hat{\sigma_i^2}\left(\frac{x_i}{\sigma_0^2} + \frac{|C\backslash i|\overline{X}_{C\backslash i}}{\sigma^2}\right), \tag{3.2}$$

where $C\backslash i$ indicates the coalition set C with the member i removed and the maximum a posteriori estimate of perceived ideological variance of i is equivalent to

$$\hat{\sigma_i^2} = 1/\left(\frac{1}{\sigma_0^2} + \frac{|C\backslash i|}{\sigma^2}\right). \tag{3.3}$$

Assuming (risk-averse) quadratic utility, i's utility becomes a function of district ideology x_i and ideology of i perceived by its constituents $\hat{x_i}$: $U_i = -(x_i - \hat{x_i})^2$. The expected utility becomes

$$\begin{aligned} eu_i &= E(u_i|\{x_j|j \in C\}, s^b, \underline{n}) \\ &= -\underbrace{(x_i - \hat{x_i})^2}_{\text{difference with district median}} - \overbrace{\hat{\sigma_i^2}}^{\text{perceptual variance}} \\ &= -\left\{x_i - \frac{1}{\frac{1}{\sigma_0^2} + \frac{|C\backslash i|}{\sigma^2}}\left(\frac{x_i}{\sigma_0^2} + \frac{|C\backslash i|\overline{x}_C}{\sigma^2}\right)\right\}^2 - \frac{1}{|C|} \end{aligned} \tag{3.4}$$

following Marschak & Radner (1972) and Enelow & Hinich (1981) which provide expected utility calculation for uncertain payoff realization. The utility is realized if and only if the coalition size meets a threshold $|C| \geq \underline{n}$ and $s^b \in C$. Assuming unit variance for both $\sigma^2 = 1$ and $\sigma_0^2 = 1$, we can trim equation 3.4 as

$$\begin{aligned} eu_i &= -\left(x_i - \frac{x_i + |C\backslash i|\overline{x_{C_i}}}{|C\backslash i| + 1}\right)^2 - \frac{1}{|C\backslash i| + 1} \\ &= -(x_i - \overline{x}_C)^2 - 1/|C|. \end{aligned} \tag{3.5}$$

I additionally incorporate an incentive component of legislators to avoid increasing the number of cosponsors above a threshold number ($\underline{n}$), the size of coalition which satisfies the minimum requirement for producing collective goods (*e.g.* size barrier for legislative proposal). This makes the utility of cosponsorship inversely proportional to the number of cosponsors with a discounting factor $\gamma < 1$. By incorporating this factor, the aggregate utility for coalition C becomes

$$\begin{aligned} v(C|s^b, \underline{n}) &\equiv \sum_{k \in C} \frac{eu_k}{|C|^\gamma} \\ &= -(Var_C + 1)|C|^{1-\gamma}. \end{aligned} \tag{3.6}$$

with

$$Var_C \equiv \frac{\sum_{k \in C}(x_k - \bar{x}_C)^2}{|C|}$$

where $|C|$ is the size of coalition C. In words, legislative activity yields credit-claiming benefit and the magnitude of the benefit decreases if many others also participate the activity (Arnold 1992, Grimmer, Messing & Westwood 2012, Hix & Jun 2009). In economic terminology, γ corresponds to the degree of rivalry of the good, where $\gamma = 1$ is the case of non-subtractable club good and $\gamma < 1$ is the case of subtractable private good. As mentioned in the introduction, this assumption directly corresponds to the logic behind minimum winning coalition in which increasing the size of coalition reduces the number of office seats for each participant party.

Figure 3.4 illustrates the varying incentive of forming coalitions by their size and intra-coalition ideological heterogeneity. As the above reasoning directs, maximum collective utility is obtained when the coalition size hits the threshold size $\underline{n}$ and intra-coalition preference variance becomes minimum. We arrive at the following propositions.

Proposition 3 ***Affinity Principle*** *For a fixed coalition size, the variance of intra-coalition preference is negatively correlated with the collective payoff:* $\frac{\partial v(C|s^b, \underline{n})}{\partial Var_C} = -|C|^{1-\gamma} < 0$.

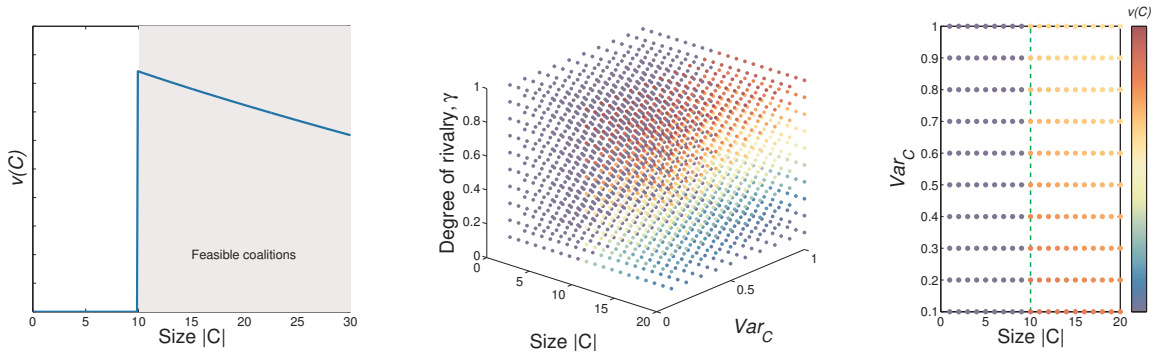

Figure 3.4: Collective payoff and coalition formation. Exemplary graphs showing the behavior of collective payoff (characteristic function) over key parameters for $\gamma = 0.5$. (Left) An example characteristic function with the size threshold 10 when intra-coalition variance is fixed. (Right) The graph displays a slice of the 3-dimensional graph on the Middle panel for $\gamma = 0.5$. As shown in the colormap, dark red indicates high values whereas dark blue indicates low values.

Proposition 4 ***Exclusiveness Principle*** *For a fixed variance of intra-coalition preference, the collective payoff is negatively correlated with the size of coalition:* $\frac{\partial v(C|s^b,\underline{n})}{\partial |C|} = (\gamma - 1)(Var_C + 1)|C|^{-\gamma} < 0$ *since* $\gamma < 1$.

3.2.3 Identification of Stable Binding Agreement

Here I propose a solution concept, core, of the game. To be specific, the realized coalition structure needs to be stable. That is, any member of coalition C (*i.e.* $i \in C$) should not have an incentive to deviate from $C \in CS$, where CS is the set of coalitional membership among N (e.g. When a single coalition C with $|C| \geq \underline{n}$ is realized, CS consists of C and singleton coalitions of $n - |C|$ individuals with size 1. In CS, C is the only relevant coalition which reached the size barrier.), and CS is weakly preferred to other possible coalition structures for $j \in N \backslash C$ (i.e. C should not be joined by individuals outside C). The corresponding solution concept, core $\mathcal{C}(G)$, is the set of all outcomes $(CS, \mathbf{u})$ satisfying $u(C) \geq v(C)$ for every subset of $C \subseteq N$.

Definition 3.2.1 $\mathcal{C}(G^b)$ *is* ***core*** *of* $G^b = (N, v)$ *if and only if the outcome* $(CS^b, \mathbf{u})$ *satisfies* $u(C) \geq v(C)$ *for every combination* $C \subseteq N$, *where* CS^b *is the set of realized coalitions in* G^b

where $C^b \equiv \{C^b | v(C^b) > 0,\ C^b \in CS^b\}$.

In words, the core solution is an agreement which prevents self-interested individuals to deviate from a specific coalition structure. This solution corresponds to a state in which there are a characteristic function maximizing the collective utility of a realized coalition (C with $|C| \geq \underline{n}$) and singleton individuals. Because every coalition and singleton individuals in CS weakly prefer CS to other possible coalition structures. With permission authority, members of C, in particular the formateur individual s^b, can screen out others if it hampers the collective payoff of the members in C.

3.3 Statistical Estimation Strategies for Latent Preference Recovery

In this section, I propose three approaches for statistical estimation of latent spatial preference from observed datasets under the proposed game-theoretic framework.

3.3.1 Exact Calculation of Optimal Positions

This subsection provides an exact optimization scheme which is directly derivable from the model proposed in the preceding section.

While a variety of factor analytic methods (e.g. singular value decomposition, principal component analysis, latent factor regression) do exist in political science to scale individual preference from covariance matrices, existing methods specify input matrix without concerning individual level strategic decisions. For example, Poole (2005) provides a factor decomposition method for agreement score distance matrix of roll call voting. Bonica (2014) uses similar specification to Poole (2005) for extracting ideology from political donation dataset but produce two types of symmetric product matrices each for

candidates and donors. Bond & Messing (2015) incorporate same approach to Bonica (2014) to infer ideal points of Facebook users and celebrity accounts.[1]

Although a large number of studies adopt the factor decomposition scheme for the recovery of latent positional preference, the exact connection between utility based reasoning and the cost minimization approach embedded in factor analytic models has been missing. To the best of my knowledge, there has been no work which established an individual level strategic foundation to produce factor decomposition matrices.

Here I show that a specific form of decomposition scheme is derivable from the stable binding agreement (Definition 3.2.1) of the preceding model.

The goal is to infer r-dimensional estimates $n \times r$ orthogonal ideal point matrix $\mathbf{X}$ from a set of observed coalitions structures $\{C^b \mid b \in B\}$. For the entire bills B, we can calculate the aggregate collective utility (characteristic function value) as

$$\sum_b v(C^b) = \sum_b u^{C^b} = -\sum_b \{ \sum_{k>l \in C^b} (x_k - x_l)^2 + |C^b| \} |C^b|^{-\gamma}. \tag{3.7}$$

With the belief that the individuals N would have realized coalitions in a way to maximize $\sum_b v(C^b)$, we now have a quadratic cost minimization problem for pairwise distance between the individual locations. Because a trivial solution to maximize $\sum_b v(C^b)$ is assigning all individuals an identical location, we need to impose an identification constraint $\mathbf{X}^T\mathbf{X} = \mathbf{I}$.

After neglecting the constant term, now the task is to maximize

$$-\sum_b \sum_{k>l \in C^b} (x_k - x_l)^2 |C^b|^{-\gamma} \tag{3.8}$$

with the identification constraint $\mathbf{X}^T\mathbf{X} = \mathbf{I}$. For notational convenience, let us define an

[1] Have not been debated intensively in political methodology community, however, specification of the input matrix for data scaling and clustering is itself a huge discipline in statistics and mathematics (Von Luxburg 2007). There exist a number of specifications for capturing covariance relationships among variables. The key procedure for these specifications is to eliminate a single rank from the full rank symmetric matrix representing

affinity matrix $\mathbf{A}$ with each element being $A_{ij} \equiv \sum_b \mathbb{1}(i,j \in C^b)|C^b|^{-\gamma}$ where $\mathbb{1}(i,j \in C^b)$ is an indicator function which becomes 1 if and only if individuals i and j both belong to coalition C^b. We also need to define a diagonal matrix of the affinity matrix whose diagonal elements correspond to row sum or column sum of the symmetric matrix $\mathbf{A}$: $\mathbf{V}$ where $V_{ii} \equiv \sum_j A_{ij}$.

Now we can rewrite $\sum_b v(C^b)$ in a matrix product format as $-\sum_k \mathbf{X}_{[,k]}{}^T(\mathbf{V}-\mathbf{A})\mathbf{X}_{[,k]}$ with the constant term removed. We can adopt Lagrangian multiplier for maximizing the function with the equality (identification) constraint. The equation simultaneously satisfying the two requirements is

$$\frac{\partial\{\mathbf{X}^T(\mathbf{V}-\mathbf{A})\mathbf{X}-\lambda(\mathbf{X}^T\mathbf{X}-\mathbf{I})\}}{\partial\mathbf{X}_{[,k]}} = 0. \tag{3.9}$$

The value of Lagrangian multiplier λ_k can be easily obtained by solving the equation and it equals $\lambda_k = \mathbf{X}_{[,k]}^T(\mathbf{V}-\mathbf{A})\mathbf{X}_{[,k]}$. This result lets us to write down λ_k for each dimension as its corresponding eigenvalue of the positive semi-definite matrix $\mathbf{V}-\mathbf{A}$. Since the initial goal was to minimize the quadratic cost function defined in equation 3.7, the order of eigenvalue λ_k becomes inversely proportional to its size. Hence for r dimensional estimates, the eigenvectors for $\mathbf{X}$ are the eigenvectors of $\mathbf{V}-\mathbf{A}$ with the r smallest non-zero eigenvalues.

A potential problem with the proposed estimation scheme is that the algorithm puts more importance to members who had more activities.[2] Since the aggregate utility function does not have a normalizing factor, we need to incorporate a principled approach for normalizing the extent each member's utility is reflected in the aggregate collective utility function $\sum_b v(C^b)$.

Here I propose a particular specification for normalization which corrects $\sum_b v(C^b)$

affinity (e.g. agreement score) among individuals. The eigenvector corresponding to the eliminated rank information is heavily related to the row sum or column sum of matrices. Specifying the matrix type for factor decomposition is particularly important in graph clustering. A variety of rank elimination schemes have been introduced including modularity, random walk and leading eigen-component removal (Fortunato 2010).

[2]This problem parallels with the non-response bias in item response theory models, which takes place when

as

$$\widetilde{\sum_b v(C^b)} \equiv \sum_{i,b} v(C^b)_i / \sum_{b'} v(C^{b'})_i. \tag{3.10}$$

By taking the same procedure of maximizing the unnormalized $\sum_b v(C^b)$ to $\widetilde{\sum_b v(C^b)}$, now the task of finding $n \times r$ orthogonal ideal point matrix $\mathbf{X}$ equals finding the r smallest non-zero eigenvalues of $\mathbf{L} = \mathbf{V}^{-0.5}(\mathbf{V} - \mathbf{A})\mathbf{V}^{-0.5}$.

Figure 3.5 compares the performance of the unnormalized and normalized approaches for ideal point recovery using a simulation example. In order to generate a synthetic affinity matrix of a multiparty legislature with realistic seat size distribution, I used heterogeneously clustered symmetric weighted graph generation algorithm proposed by Lancichinetti, Fortunato & Radicchi (2008). In order to make unequal seat size distribution for the parties, seat size was sampled from a power law distribution with exponent -1.5. There was partisan preference with 60% of affinity value assigned to intra-party weights ($\sum_j A_{ij}\mathbb{1}(p_i = p_j)/\sum_j A_{ij} \approx 0.6$). The total number of seats in the synthetic legislature was $n = 500$. As shown in the left panel, the number of parties was 11.

We can verify the benefit of normalization by observing the segregation of inter-party members on the 2-dimensional space. Because members of each party have strong partisan preference and inter-party members exchange fairly low level of affinity weights, a proper estimation should result clustering of the individuals by party label. However, the unnormalized estimates in the middle panel show inferior results in which the majority of parties are mixed in the space with each other. In the right panel, by contrast, parties are well segregated in the plane, reflecting partisan preference embedded in the data generating processes.

This subsection demonstrated that the coalitional game theoretic formulation leads to an exact derivation of specific factor analytic model for ideology inference. While the

there exist systemic patterns of abstention in the observed datasets (Rosas, Shomer & Haptonstahl 2015). The same approach can be applied to re-weight the likelihood function of standard item response theory models, but I do not detail here.

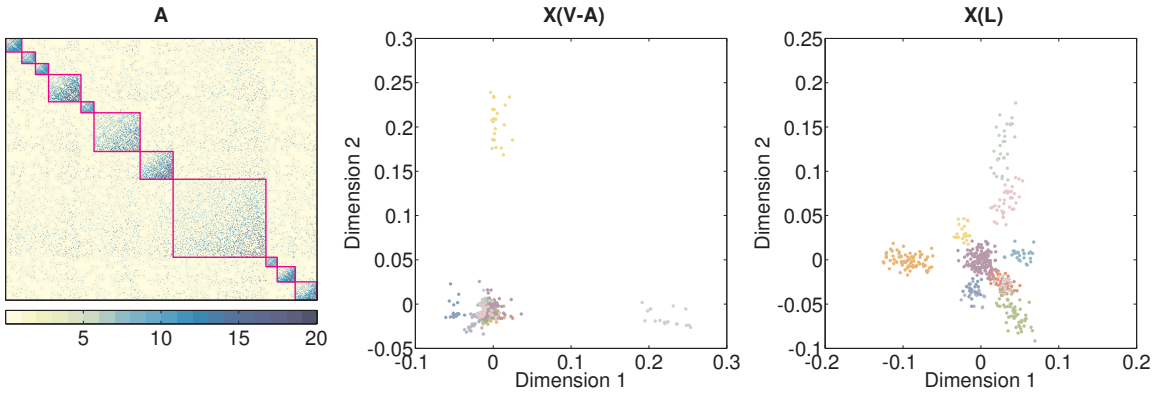

Figure 3.5: Synthetic coalitions and ideal point estimates. Synthetic example showing the impact of normalization on ideology recovery. A difficult, yet frequently observed, case of A with unequal party seat size distribution ($\langle \#(P_k) \rangle$) and heterogeneous affinity rates (colored in the left panel as pairwise weight): $n = 500$, $\#(P_k) \sim c\#^{-1.5}$, $A_{ij} \sim cw^{-1.5}$, $\sum_j A_{ij} \mathbb{1}(p_i = p_j) / \sum_j A_{ij} \approx 0.6$. Each color in the middle (unnormalized) and right (normalized) panels corresponds to party label. Party identity corresponds to dense blocks in the affinity matrix **A** at the left panel.

method results point estimates with maximum aggregate collective payoff, one can employ non-parametric bootstrap method to obtain uncertainty estimates of the model as introduced in Jacoby & Armstrong (2014).

3.3.2 Probabilistic Inference via Fully Bayesian Implementation

The exact calculation method introduced in the preceding section neglects a disturbance term as opposed to the two other statistical solutions I will introduce shortly. Empirical data generating process would not follow the perfect quadratic utility reasoning. There can be numerous sources of disturbance. In legislative coalition building, we can think about many situations in which the realized behavior of politicians do not represent their sincere preference, such as log-rolling and abrupt events that prevent them from choosing otherwise preferred options.

To account for inherent uncertainty in the data generating process, I introduce Bayesian extension of the analytic method by incorporating model-based matrix factor analysis approach with fixed-rank condition (Hoff 2007, Lopes & West 2004). This probabilistic

inference scheme contains a remedy for the uncertainty concern by providing probabilistic density functions instead of point estimates.

Because it is sufficiently dense, I decompose the normalized Laplacian matrix $\mathbf{L} = \mathbf{XDX}^T + \mathbf{E}$ using continuous matrix distributions. Following Hoff (2007), the probability of matrix formation for $\mathbf{L}$ is given as

$$p(\mathbf{L}|\mathbf{X},\mathbf{D},\phi) = \left(\frac{\phi}{2\pi}\right)^{n^2/2} \exp\left(-\frac{1}{2}||\mathbf{E}_{-j}||^2 + \phi d_j \mathbf{X}_{[,j]}^T \mathbf{E}_{-j} \mathbf{X}_{[,j]} - \frac{1}{2}\phi d_j^2\right) \quad (3.11)$$

with priors given by $\mathbf{X} \sim \mathcal{U}(\nu_{r,n})$, $\mathbf{D} \sim \mathcal{N}(\mu, 1/\psi)$ and $\mathbf{E} \sim \mathcal{N}(0, 1/\phi)$ where $\mathcal{U}$ is a multi-dimensional uniform distribution and $\mathcal{N}$ is a multidimensional Gaussian distribution. As proposed by **?**, I incorporate matrix von Mises-Fisher density to estimate the orthonormal matrix $\mathbf{X}$.

Gibbs sampling chain for the recovery of orthonormal matrix $\mathbf{X}$ using von Mises-Fisher distribution, which results normalized (*i.e.* unit norm) vectors, is as follows. Except when sampling orthonormal matrices, I use standard distributions in Bayesian statistics for sampling other parameters. Notations and parameters are analogous to the model proposed in Chapter 1. See Chapter 1 for the details.

$$\begin{aligned}
(\mathbf{X}_{[,j]} \mid \mathbf{L}, \mathbf{X}_{[,-j]}, \mathbf{D}, \phi) &\equiv \mathcal{N}^x_{\{-j\}} \mathbf{x}_j, \mathbf{x}_j \sim \text{vMF}(\phi d_j \mathcal{N}^{x^T}_{-j} \mathbf{E}_{-j} \mathbf{X}_{[,j]}) \\
(d_j \mid \mathbf{L}, \mathbf{X}, \mathbf{D}_{[-j,-j]}, \phi, \mu, \psi) &\sim \mathcal{N}[(\mathbf{X}_{[,j]}^T \mathbf{E}_{-j} \mathbf{X}_{[,j]} \phi + \mu\psi)/(\phi + \psi), 1/(\phi + \psi)] \\
(\phi \mid \mathbf{L}, \mathbf{X}, \mathbf{D}) &\sim \Gamma[(\nu_0 + n^2)/2, (\nu_0 \sigma_0^2 + ||\mathbf{L} - \mathbf{XDX}^T||^2)/2] \\
(\mu \mid \mathbf{D}, \psi) &\sim \mathcal{N}(\psi \sum d_j + \mu_0/v_0^2)/(\psi r + 1/v_0^2), 1/(\psi r + 1/v_0^2) \\
(\mu \mid \mathbf{D}, \psi) &\sim \mathcal{N}(\psi \sum d_j + \mu_0/v_0^2)/(\psi r + 1/v_0^2), 1/(\psi r + 1/v_0^2) \\
(\psi \mid \mathbf{D}, \mu) &\sim \Gamma[(\eta_0 + r)/2, (\eta_0 \tau_0^2 + \sum (d_j - \mu)^2)/2]
\end{aligned}$$

What is very different from existing approaches for ideal point estimation in political science

is that the proposed sampling scheme aims to directly sample **X**, satisfying the perfect orthonormal condition (Chikuse 2012). This property automatically identifies the model. This is a critical difference of the present approach to conventional techniques for 2 parameter item response theory models which obtain approximately orthogonal multi-dimensional estimates through sequential sampling of each single dimensional estimate without any correlational restriction to other dimensional estimates (Jackman 2001, Poole 2005). Conventional approach can only obtain approximately orthogonal estimates (Hyvärinen, Karhunen & Oja 2004).

To the contrary, the present approach samples the multidimensional estimates **X** from a set of orthonormal matrices to satisfy the dimension-specific normalization condition, and, through the use of null operators, the algorithm ensures orthogonality between dimension-wise estimates. As a result, it always samples **X** satisfying the orthonormal condition.

3.3.3 Variational Approximation for Dimensionality Selection

In this section, I introduce a dimensionality selection method for rank approximation using a variational approximation scheme (Bishop 2006).

An unresolved issue in statistical methods for ideal point recovery is developing criteria for model selection. In particular, dimensionality selection for determining the number of aggregate dimensions is associated with important questions in political polarization (Poole & Rosenthal 1984), collective choice theory (Shepsle 1979) and comparative legislative behavior (Lijphart 2012). While several heuristics for rank determination exist, a principled approach has not been provided so far.[3]

Here I introduce an approach for rank determination of choice datasets. The approach I am introducing uses the strong regularization effect found in variational approximation methods. It has been known that this regularization effect can be used to dissociate Guassian

[3]Recently, Kim, Londregan & Ratkovic (2015) proposed a related method aimed at truncation of dimensions for ideal point estimation but did not explicitly incorporate the variational approximation scheme.

noise component from observed matrices.

Empirical Variational Bayesian Matrix Factorization (Nakajima, Sugiyama & Babacan 2011, Nakajima, Tomioka, Sugiyama & Babacan 2012) dissociates Gaussian noise component from shrinkage factors and infers the planted number of dimensions. The matrix factorization model starts by defining a posterior probability given by

$$p(\mathbf{X},\mathbf{D}|\mathbf{L}) = \frac{p(\mathbf{L}|\mathbf{X},\mathbf{D})p(\mathbf{X})p(\mathbf{D})}{p(\mathbf{L})} \tag{3.12}$$

where $p(\mathbf{L}) = \langle p(\mathbf{L}|\mathbf{X},\mathbf{D})\rangle_{p(\mathbf{X})p(\mathbf{D})}$ is the expected value of $p(\mathbf{L}|\mathbf{X},\mathbf{D})$ over $p(\mathbf{X})p(\mathbf{D})$.

The key difference of variational approximation to standard fully Bayesian procedure is the free energy (Kullback-Leibler divergence) minimization approach it adopts for the approximation of the full posterior distribution and the independence assumption among parameters. In the current problem, the goal is to minimize the following formula (*i.e.* free energy) driven from equation 3.12.

$$F(r) = \left\langle \log \frac{p(\mathbf{L}|\mathbf{X},\mathbf{D})p(\mathbf{X})p(\mathbf{D})}{p(\mathbf{L})} \right\rangle_{r(\mathbf{X},\mathbf{D})} = \left\langle \log \frac{p(\mathbf{L}|\mathbf{X},\mathbf{D})}{p(\mathbf{L})} \right\rangle_{r(\mathbf{X},\mathbf{D})} - \log p(\mathbf{L}) \tag{3.13}$$

where $r(\mathbf{X},\mathbf{D})$ is the trial distribution given throughout the optimization routine. Due to the independence assumption, I factorize $r(\mathbf{X},\mathbf{D})$ as $r(\mathbf{X})r(\mathbf{D})$.

Matrix factorization model is one of the few variational approximation problems in which exact analytic calculation of the lower bound is possible (Nakajima, Sugiyama & Babacan 2011, Nakajima et al. 2012). The shrinkage solution has identical $\mathbf{X}(\mathbf{L})$ with the exact solution but with a truncation threshold $1-\underline{\gamma}^{EVB}$. I let $\{\gamma_k\}$ the eigenvalues of $\mathbf{V}^{-0.5}\mathbf{A}\mathbf{V}^{-0.5} = \mathbf{I}-\mathbf{L} = \mathbf{I}-\mathbf{V}^{-0.5}(\mathbf{V}-\mathbf{A})\mathbf{V}^{-0.5}$ whose spectral property is identical to $\mathbf{L}$ but with $\gamma_k = 1-\lambda_k$. Now minimum factor loading discovery problem converts to maximum

[4] One can esitimate the variance with an additional step minimizing the free energy which satisfies $\hat{\sigma}^2 = \text{argmin}_{\sigma^2}(\min_{r,C_\mathbf{X},C_\mathbf{D}} F(r(\mathbf{X},\mathbf{D});C_\mathbf{X},C_\mathbf{D},\sigma^2|\mathbf{L}))$ where $C_\mathbf{X}$ and $C_\mathbf{D}$ respectively denote prior diagonal covariance matrices of $\mathbf{X}$ and $\mathbf{D}$.

factor loading discovery problem.

With $\hat{\sigma}^2$ being estimated variance[4] and κ^* being a constant satisfying the relationship $\frac{\log(\sqrt{\alpha}\kappa+1)}{\sqrt{\alpha}\kappa} + \frac{\log(\kappa/\sqrt{\alpha}+1)}{\kappa/\sqrt{\alpha}} = 1$, we can write the truncation criteria as follows.

$$\check{\gamma}_k^{EVB} = \begin{cases} \check{\gamma}_k^{EVB} & \text{if } \gamma_k^{EVB} \geq \underline{\gamma}^{EVB} \\ 0 & \text{if } \gamma_k^{EVB} < \underline{\gamma}^{EVB} \end{cases} \tag{3.14}$$

$$\underline{\gamma}^{EVB} = \hat{\sigma}\sqrt{2n + n(\kappa^* + 1/\kappa^*)} \tag{3.15}$$

$$\check{\gamma}_k^{EVB} = \frac{\gamma_k}{2}\left(1 - \frac{2n\hat{\sigma}^2}{\gamma_k^2} + \sqrt{\left(1 - \frac{2n\hat{\sigma}^2}{\gamma_k^2}\right)^2 - \frac{4n^2\hat{\sigma}^4}{\gamma_k^4}}\right) \tag{3.16}$$

The key idea for this rank determination approach is to use the inferred variance parameter along each dimension to screen out dimensions with low variance levels through model-based regularization. This approach is extremely convenient to employ compared to its simulation-based fully Bayesian counterpart, introduced above, which takes a long time for convergence for sampling the number of dimensions using Markov Chain Monte Carlo.

3.4 Strategic Coalition Formation in Bill Proposal Stage: Application to Historical Korean National Assembly Dataset

Korean National Assembly (KNA) is an attracting case to study the evolutionary path of a parliament due to its broad spectrum of democratic development over the history. Starting as a state with no democratic tradition at its liberation from Imperial Japan in 1945, as of 2012, South Korea's democracy index is ranked 20th in the globe, which is the highest

ranking in Asia and higher than that of the United States (Economist Intelligence Unit Democracy Index 2012). South Korea has experienced grand alterations such as a civil war, military dictatorship, and resurgence of democratization, thereby providing a unique opportunity for understanding ideological divergence and issue saliency dynamics associated with democratic transitions and historical events. Using newly combined dataset covering the entire legislative records of KNA and parliamentary elections, I provide quantitative analysis of both aggregate and issue-specific ideological patterns of the assemblies ranging from 1948 to 2013.

3.4.1 Data Acquisition

The dataset is obtained from National Assembly Legislative Information System (NALIS)[5] and National Election Commission (NEC)[6] web pages.[7] Each legislator's name and party affiliation from NALIS are matched with information from NEC which include election results and candidate biographical information. Party affiliation of each individual is coded on the basis of one's affiliation in the beginning of each chamber. Bill-specific activity records are parsed from each pdf document containing information on every single bill proposed. Unique ID was assigned to each individual legislator although one served in multiple chambers with different party affiliations.

Bill cosponsorship records of legislators and bill titles in Korean language are the primary datasets used in the present study. For estimating latent ideological preference of individual legislators, I use cosponsorship data as their choice records.

Majority of ideal point estimation techniques are developed to analyze roll call votes in US Congresses (Poole 2005). A few previous research which aimed to uncover latent ideological preference of actors in recent chambers of KNA also utilized one of the most

[5] http://likms.assembly.go.kr/bill

[6] http://www.nec.go.kr

[7] I thank the members of team Popong, in particular, Eunjeong Park and Cheol Kang for their help in acquisition and prepossessing of the dataset.

popular methods, NOMINATE, to analyze the 16th and the 17th chambers (Hix & Jun 2009). However, there are clear reasons why one needs to use cosponsorship data for the analysis of KNA over a longer time span.

First and an obvious reason is that *recorded voting started from the 16th session (year 2000) in KNA*. Before this period, the majority of floor votes were not recorded. Due to its non-existence one cannot utilize roll call voting data for ideological scaling of members in KNA prior to year 2000.

Second, KNA exhibits extremely strong party line vote ever since when the roll call vote was institutionalized. For instance, the members of 17th and 18th chambers casted more than 90% votes as party line votes (Moon 2011). In contrast, the exclusiveness of cosponsorship coalitions yields higher granularity in individual choices.

Third, abstention in roll call voting is strongly correlated with a legislator's seniority. Due to the fact that senior legislators avoid vote safe bills (Moon 2011), utilizing roll call vote as a major source of ideological estimation can incur systemic bias.

Fourth, bill cosponsorship is highly visible to Korean citizens. In the world wide web, One can browse more than 20,000 news articles and tweets mentioning "cosponsorship" in Korean. This fact supports the validity of applying the assumption of the proposed coalition signaling model to understand the generating process of KNA cosponsorship dataset. Given the fact that these articles and tweets usually contain the list of cosponsors for each bill, ideological leaning of individual legislators is likely to be perceived by the citizens on the basis of whom they collaborate with.

To summarize, these factors together lead researchers to use cosponsorship records as a primary legislative choice dataset for ideological estimation of individual legislators. In contrast to roll call voting data, every single bill proposed since the first chamber of KNA is listed in NALIS including the information on who sponsored and cosponsored the bills.

In total, I analyze 82,452 bill proposal records of 2,334 individual legislators from

the 2nd to the 19th session. Contrary to US Congresses, bills can be also proposed by collective bodies in the Government (i.e. ministers) and the assembly (i.e. committee chairs). I exclude bills introduced by collective bodies, since these bills do not have any cosponsor and it is not applicable to use this information for ideological scaling of legislators.

KNA has a unique setting in bill proposal stage. From the 2nd session, only with a small number of exceptions for particular types of bills, every bill proposed by individual legislators must have 9 collaborators or 19 collaborators at minimum.

As opposed to cosponsorship institution of US Congresses, in which legislators can freely sign for a bill if they want to, bill cosponsorship is closed in KNA. A legislator can be a cosponsor of a bill only when she is included in the bill draft when it is initially submitted. This provides permission authority to sponsors for screening out cosponsors whose preference deviates significantly from sponsor ideology.

3.4.2 Individual Preference Locations in Aggregate Policy Space

Figure 3.6 illustrates legislator ideal points inferred for each session of KNA. By looking at the overall patterns of ideological landscape, one can trace the evolution of party system over the past half century in terms of within-party ideological coherence and inter-party ideological differentiation. Party dominance in making bill proposals increased starting around 1960. The segregation of parties in the ideological space is apparent in the latter chambers.

The last panel of Figure 3.6 provides quantitative measurement of the party polarization trend. Except the 2nd, 3rd and 5th sessions, a high level of party polarization is observed in all chambers.

The party polarization trend has important implications with respect to the change in party strength through democratic transitions. We can find several interesting observations in relation to the political turmoils in Korean history. The spring of Seoul, denoting the

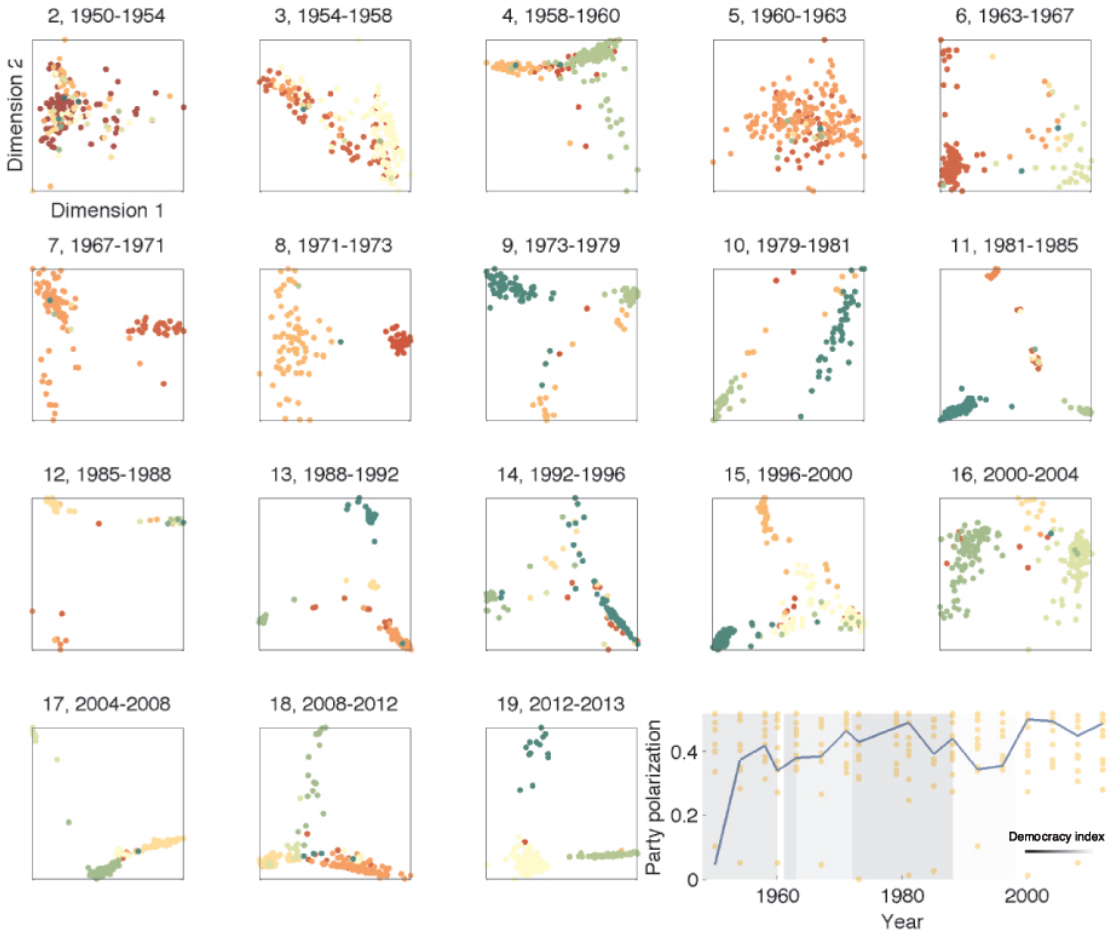

Figure 3.6: Ideological distributions of legislators from the 2nd to the 19th KNA. For each session, a unique color is given to each party. Same color across sessions does not represent same party label. Members without party affiliation are excluded from the figure. In the last panel, I plot the level of party polarization in bill coalition formation. The procedure for computing the specific nonparametric segregation index for polarization is explained in the appendix. The blue line indicates mean value of party polarization for each session obtained over 10 simulations. A higher value indicates that the pattern of segregation in cosponsorship coalition formation is well explained by the members' party affiliations. In other words, high value indicates high level of party polarization. Shade in a given period indicates Polity4 democracy index (Marshall & Cole 2014) of the corresponding period. Dark shades indicate low democracy score whereas bright shades indicate high democracy score.

short period of democratization after Park Jung-hee's assassination in 1979, yielded constant increase in the party polarization index until the end of authoritarian governance by military leaders (i.e. Jun Doo-hwan and Noh Tae-woo). The 14th chamber, which started in 1992

followed by the election of the first non-militant president Kim Young-sam in 1993 ever since 1962, exhibits the lowest level of party polarization, indicating that the democratization (i.e. direct presidential election) incurred turmoil in the existing party disciplines. Further investigation, such as counting the number of party switchers or the proportion of incumbents, would be necessary to better understand the trend.

3.4.3 Comparison of the Recovered Ideology to W-NOMINATE Scores

To compare the W-NOMINATE scores and the proposed estimates, I highlight the case of the 17th session, which had the most broad ideological spectrum in the history of KNA with the largest number of assembly members from the labor party. In the beginning of the chamber, 10 members were affiliated with Democratic Labor Party (DLP) largely due to the change in the electoral system (i.e. the introduction of party-list proportional representation elections).

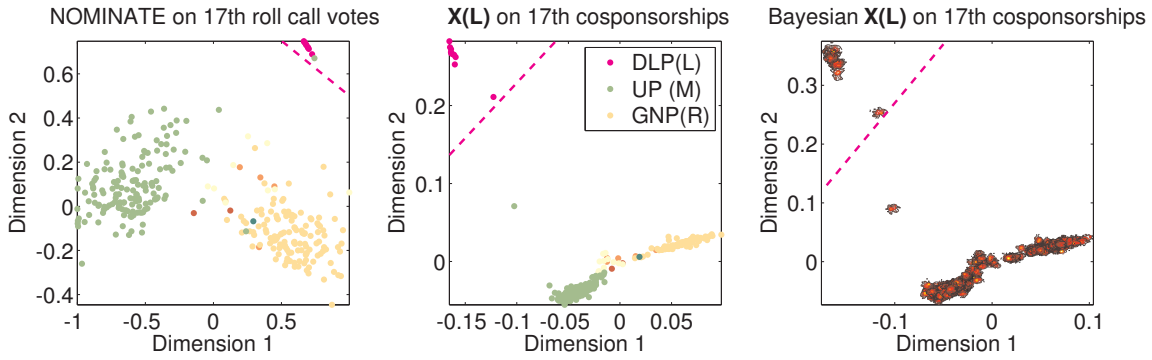

Figure 3.7: Comparison with W-NOMINATE. In W-NOMINATE, government-versus-opposition dominates the principal axis: DLP (labor) is located on the right of UP (median, majority, presidential) and GNP (conservative). Such opposition is marginalized into the 2nd dimension of the bill proposal space, and the parties are well aligned in the 1st dimension (DLP-UP-GNP), as observed in NOMINATE on other sessions without UP majority.

When applying W-NOMINATE to the 17th roll call voting data, we observe a strong government-opposition tendency of the two minority parties, Grand National Party (GNP) and DLP, toward the majority Presidential party Uri Party (UP) (Figure 3.7). Believed to be

the most conservative party in the 17th chamber, GNP holds an indistinguishable location to the leftist, DLP. This left-right ideological characteristic is partially observed on the second dimension. UP members hold the moderate position, while there exists a high level of overlap between UP and GNP members.

Such opposition along the principal axis of the W-NOMINATE space is marginalized into the 2nd dimension of the bill proposal space constructed using the proposed method. The three parties are well aligned in the principal dimension (DLP-UP-GNP) (i.e. labor-median-conservative), as observed in NOMINATE on other sessions without UP majority. Fully Bayesian implementation also gives similar estimates to the exact results.

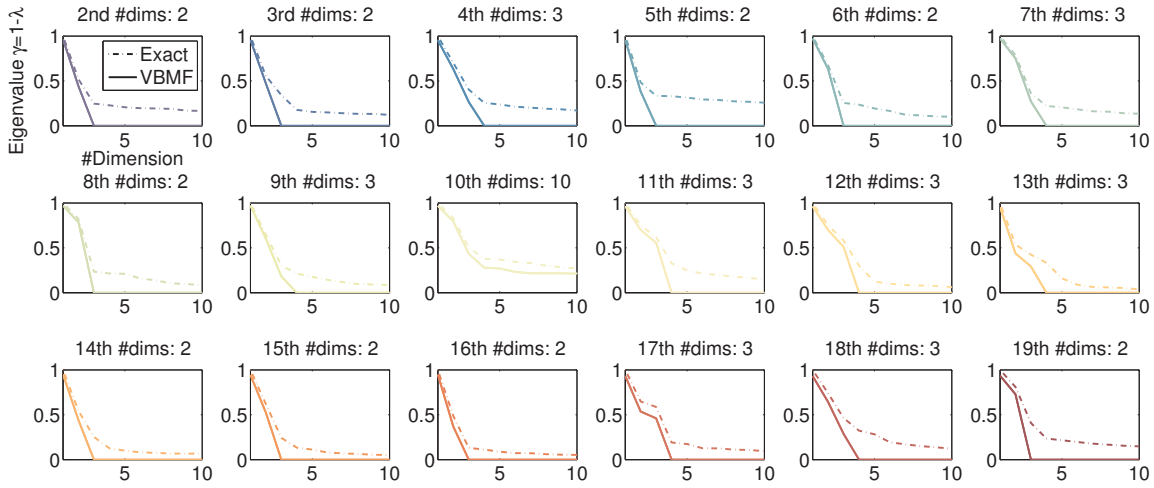

Figure 3.8: Dimensionality of KNA. Number of dimensions inferred through dimensionality truncation using variational approximation. Dotted lines indicate factor loading obtained using the exact framework whereas the solid lines indicate truncated loading obtained using Variational Bayesian Matrix Factorization (VBMF).

3.4.4 Dimensionality Selection via Variational Bayes

I apply the empirical variational approximation approach introduced in the previous section to determine the number of aggregate policy dimensions embedded in legislative bill proposal decisions.

Figure 3.8 compares the distributions of factor loading for each dimension obtained

with exact calculation and truncated factor loading obtained from variational approximation. As shown, except the extremely short-lived 10th session, all sessions exhibit stable number of truncated dimensions ranging from 2 to 3.

3.4.5 Identifying Substantive Issue Coverage Trend Through Bill-Issue Matching

A significant piece of information missing in the aggregate bill proposal data analysis is that we do not know substantive issues underlying the revealed ideological preference of legislators. One way to understand what constitutes their ideological characteristics is to study the issue contents of the bills proposed. Because it is very difficult to find labels that classify the bills into issue categories, we need to employ inference techniques to know issue contents of bills.

While more rigorous analysis can be conducted using the entire set of words used in bill documents and committee referral information, as a first start, I use bill titles to identify issue contents of the bill proposals. Because bill titles are short text, due to the sparsity in word-to-word associations, it is very difficult to infer topics using uni-gram level topic modeling techniques. In order to overcome this difficulty, I utilize phrase-level mining approach (Wang, Danilevsky, Desai, Zhang, Nguyen, Taula & Han 2013) which extracts multi-gram level semantic associations from word usage in bill titles. Details of the algorithm are introduced in the appendix.

Table 3.1 shows top phrases inferred from the phrase level topic modeling approach. Despite some misclassification errors present in the results, high ranking phrases in each topic cluster seem to represent common issue. For example, the first topic is mainly on regional development, and tax and investment issues on cities and rural areas. The 4th topic is on educational issue. The 2nd, 5th, 6th and 10th topics cover budget related laws, enterprise regulation laws, business and account related laws, and earnings and tax-

Table 3.1: Phrase ranking by topic. Top ranking phrases of the 12 topics revealed using the topic hierarchy model from the pooled bill title dataset. I, D and F denote increasing, decreasing and fluctuating trends respectively, that are classified using the time series clustering method proposed by Caliński & Harabasz (1974).

Topic 1	I	Provincial Tax-Law	Region Installation	City
Topic 2	D	Organization	Additional Yearly Budget	Supplementary Yearly Budget
Topic 3	I	Tax Exemption Limit	Public-Official	Housing
Topic 4	I	Elementary and Secondary Education	Infant Care	Higher Education
Topic 5	I	Monopoly Regulation Law	Protection-Law	Enterprise Act
Topic 6	F	Special Limits for Development Administration	Account-Law	Business and Land Public Acquisition
Topic 7	F	Special Committee Composition	Fact-finding Committee	Confirmation Hearing
Topic 8	F	Secretary-of-State Committee Hearing	Secretary-of-State Committee Economy Hearing	Minister Candidate Greetings
Topic 9	I	Public Health Insurance Legislation	Food Sanitation-Law	Urban Environment Dwelling
Topic 10	D	Repayment Warranty	Earnings	Tax-Evasion Prevention
Topic 11	F	Inspection Approval	Inspection Report	Congressman Resignation
Topic 12	I	Information Utilization Protection	Information Utilization Promotion	Communications-network Protection Promotion

evasion laws respectively. The 9th topic is health and food related laws. The 11th topic includes inspection related laws. The 12th topic involves the majority of phrases related to information technology and communications industry. The 7th and 8th topics are about opening hearings for candidate appointment and fact confirmation. The 3rd and 11th topics have relatively mixed issue-representing phrases.

The issue-relevant probability assigned to each bill allows one to trace the evolutionary path of issue coverage over the chambers. To classify the overall trends of the topic coverage, I apply distance-based time series clustering method, which classifies a set of one-dimensional time series data into discrete groups by looking at their patterns of progression over time (Peng & Müller 2008). The classification result is shown in Table 1. The clustering metric proposed by Caliński & Harabasz (1974) is maximized when the number of clusters equals 6. 2 out of the 6 clusters have clear increasing and decreasing patterns respectively, each having 6 and 2 topics. The remaining clusters consist of single topic member with fluctuating trends.

Issues with increasing trends are associated with regional development, enterprise,

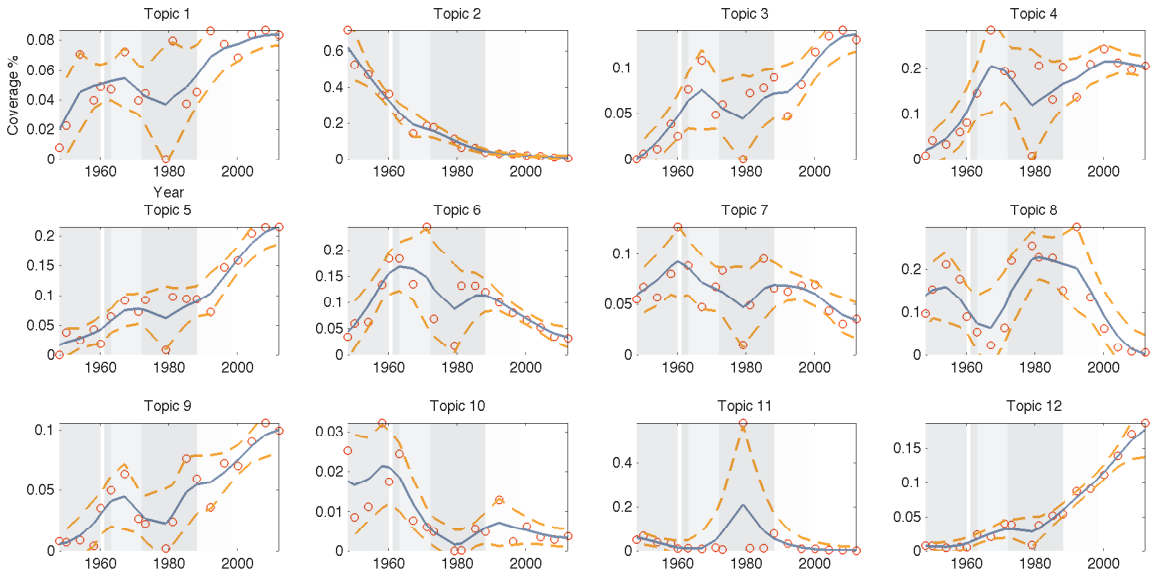

Figure 3.9: Topic coverage over time. Trends are plotted using robust locally weighted scatterplot smoothing (RLOESS). Intervals denote $\pm 2\sigma$ regions. Shade in a given period indicates Polity4 democracy index of the corresponding period. Dark shades indicate low democracy scores whereas bright shades indicate high democracy scores.

education, information and communication technology, and health and food (Figure 3.9). The increasing trends found for all these issues look plausible given the increasing demand for making laws relevant to the issues. For example, over the course of the rapid industrialization, the extreme concentration of capital, population, and infrastructures on metropolitan areas has become the major social issue in South Korea. The increasing trend of Topic 1 demonstrates that the government and the assembly have invested increasing efforts on balanced development of regions. Another case is the constant increase of information and technology-related issue (Topic 12). The coverage size of this issue starts a sharp increase around 1980, which coincides with the supply of televisions and telephones in South Korea.

On the contrary, one of the two decreasing issues is related to budget. This observation can be understood as an artifact of the constant increase in the number of bills proposed. Since the number of legislations discussing budget-related issues may remain constant across the sessions, whereas bills related to other issues exhibit steady increase

in their total volume, the proportion of legislation discussing budget-related issue would decrease.

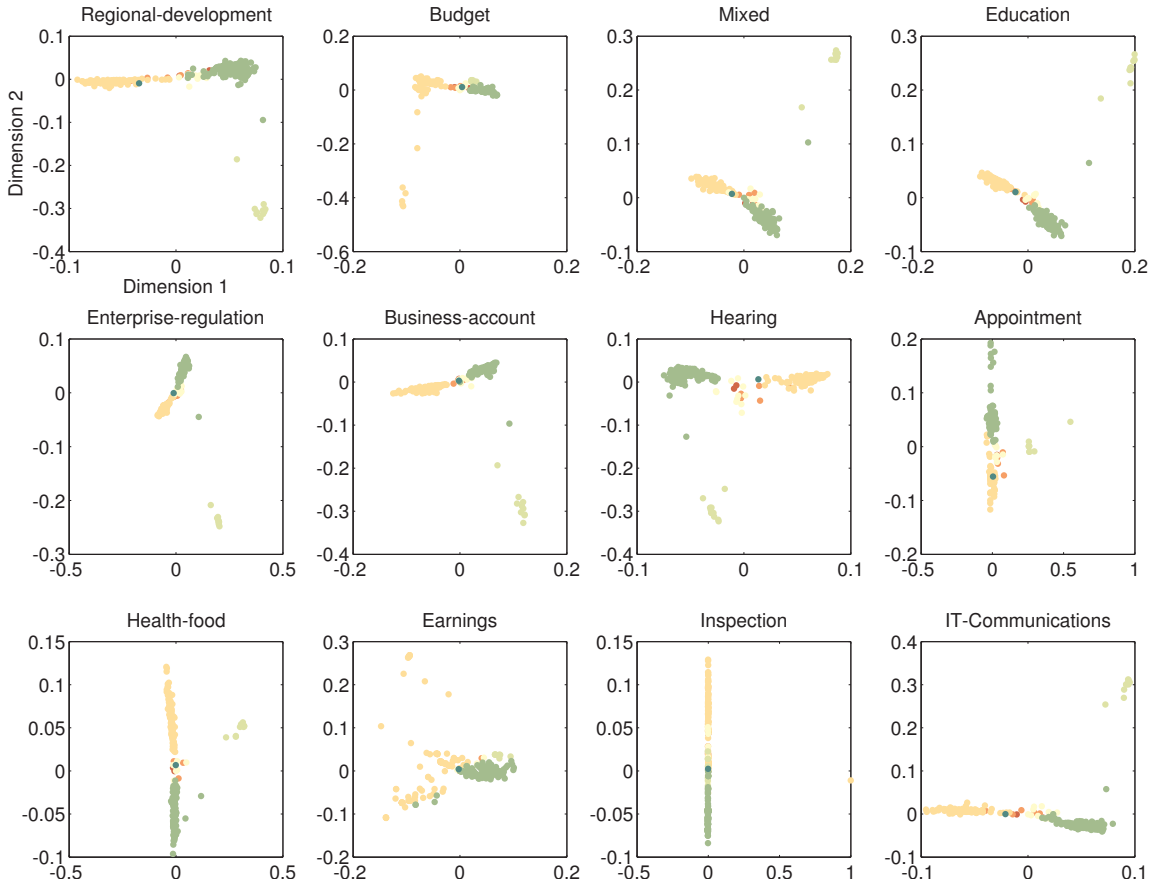

Figure 3.10: Issue-specific ideology inferred from the 17th session bill proposal dataset. Green: Uri Party (median,presidential,majority). Yellow: Grand National Party (conservative). Light Green: Democratic Labor Party (labor).

3.4.6 Issue-specific Ideology Estimation

Similar in spirit to Lauderdale & Clark (2014) who extracted issue-specific ideology of supreme court justices by sequentially extracting issues from opinion texts and issue-specific ideology from subsets of voting matrices, one can infer issue specific ideology of each session of KNA using the topics in Table 3.1. The proposed ideal point estimation scheme is flexible enough to accommodate the estimation of issue-specific ideology. As

an example, I conduct ideal point estimation of issue-specific ideology of the 17th session. Details of the process for issue-specific ideology extraction is given in the appendix.

Figure 3.10 shows issue-specific ideology inferred from the proposed approach. We can see a high level of issue-correlation between the issue-specific estimates. Party centered behavior of legislators are fairly well preserved across the issues. There are several notable patterns. As shown in the previous subsection, DLP was the most liberal party in the 17th session in their aggregate ideological scores. While such pattern is well preserved in issues including regional-development, education and IT-communications bills, we can see a strong idiosyncratic pattern of legislative activity in health-food related ideology estimates. That is, the presidential party UP and the conservative minority party GNP form a strong coalition in the first dimension while DLP is isolated in the right side of the spectrum. In the 17th session, one of the most debatable political issue over the country was establishing a Free Trade Agreement (FTA) with US. Among many products that would be affected by the agreement, food was the most critical product category in the debate. DLP formed a strong coalition with agricultural interest groups and their major agenda during the session was anti-FTA and agricultural protection. This unique issue-specific ideological preference, while hidden in the aggregate estimates, is well reflected in the issue-specific ideology scores.

3.4.7 Empirical Case Study: Strategic Positioning of a Satellite Party Under Authoritarian Regime

While there are potentially many relevant research questions to answer using the present dataset, I probe an exemplary question in relation to the strategic behavior of authoritarian governments.

One of the puzzles in authoritarian governance is that it is very common for them to hold elections and maintain party systems despite the fact that these decisions can be

risky (Boix & Svolik 2013, Svolik 2009). Strategic power sharing behavior of authoritarian governments has been known to increase their legitimacy by pretending to allow ordinary political activities.

A related empirical regularity found in many historical examples is the existence of satellite parties (i.e. bloc parties). Satellite parties are government-built parties that are different from the ruling parties. Although the communist party is known to be the only active party in China, China is formally a multi-party state with a number of minority parties. Approved by the communist party, there are about 10 official minority parties.

Satellite parties have existed in other authoritarian regimes including Soviet Union, East Germany and North Korea. Despite such frequent existence, the legislative behavior of satellite party members is not well known.

Here I conduct a simple analysis of a satellite party in KNA to understand how different their ideological characteristics were to the ruling party members. This is an interesting question to ask because satellite parties have opposing incentives for being cohesive to the government party and distant to the government party. On the one hand, satellite party members would have similar ideological interest to the ruling party because they are supported by the ruling party. On the other hand, their revealed preference needs to make the audience believe that they are independent to the ruling party. As a result, such strategic position-taking will reduce the level of ideological overlap of the satellite party with the ruling party and make the satellite party to compete with other parties for voters with anti-government sentiments. With the ideological estimates inferred through the analysis, we can answer which incentive dominated the other.

There have been more than five satellite parties over the history of KNA. Among

[8]Perhaps, the most well-known satellite party would be Yushin Party (YP) which was formed after the revision of constitution by dictator Park Jung-hee. In 1972, the government revised the constitution to allow the president to dominate parliament, judiciary and government administration. One notable action was the president's authority to appoint $1/3$ of the seats in the legislature. Each cohort was appointed three times in 1973, 1976 and 1979 respectively. However it was a semi-party organization which took a peculiar status in the institution. As a result, it was not recorded as a party in NALIS.

those, I focus on one of the most well-known parties.[8] The satellite party, Democratic Korea Party (DKP), was made in 1981 by the order of dictator Jun Doo-hwan. It has been known that National Intelligence Service of Korea financially supported the party and the ruling party, Democratic Justice Party (DJP), appointed the leader of DKP. In the 11th session, DKP earned 81 seats becoming the largest minority party and 35 seats in the 12th session.

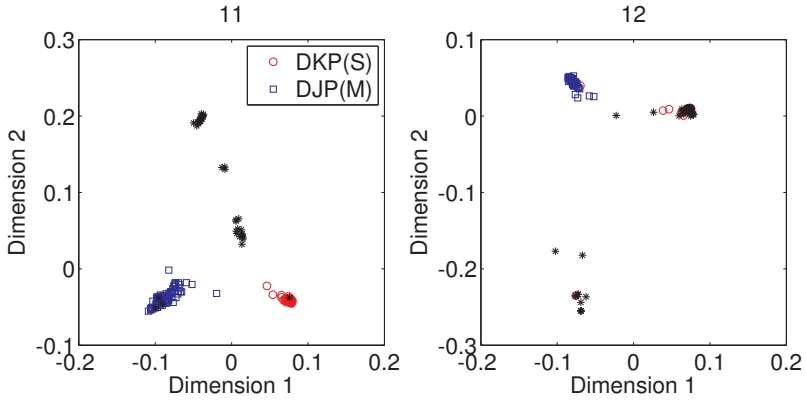

Figure 3.11: Satellite party ideal points. Ideological characteristics of the majority (government) party (DJP) and the satellite party (DKP). Other minority parties are colored in black.

Here I show the relationship between DKP-DJP pair in terms of their affinity in legislative activities. As shown in Figure 3.11, the two parties conducted almost unrelated legislative actions to each other in the 11th session. In the principal axis, DKP and DJP are the parties with the highest level of ideological difference without any overlap in member ideology. We can observe similar patterns in the 12 session. The ruling party and the satellite party still maintain a substantial level of ideological distance to each other.

This result supports the hypothesis that the satellite party took an almost pure perfunctory role in the legislature to promote the legitimacy of the authoritarian government, thereby having a position-taking incentive for ideological independence to DJP. This result is appealing particularly due to the fact that 11th and 12th sessions held multi-member district elections. With the danger of having sincerely opposing parties in the legislature, the ruling party and the government had an incentive to present a satellite party whose ideological

characteristic was different to the ruling party. In this way, the satellite party could compete with other minority parties in the similar ideological niche.

3.5 Concluding Remarks: Toward a Framework for Systemic Analysis of Coalition Datasets

In this paper, I develop an informational rationale for coalition formation among position-taking motivated individuals and derive statistical approaches for ideal point estimation from observed coalition structures.

The method is applied to newly collected bill proposal coalition dataset of Korean National Assembly. Through the analysis of the empirical dataset, I have shown that the proposed approach can offer an effective tool for extracting latent preference of political actors, especially when voting decision data are unavailable due to institutional conditions. The proposed methodological framework is flexible enough to extract ideological scores from a weighted subset of the original dataset. I incorporate text topic modeling approach to bill contents to match bills with substantive issues and derive issue-specific ideology scores.

The empirical data analysis results alone may answer important questions about legislative behavior in various political environments. Researchers may find relevant cases of natural experiments incurred by changes in election types, presidential/parliamentary systems or regime types. Other than institutional changes, there are notable patterns related to the emergence of regional parties and rapid industrialization. The updated dataset will include additional information including party lineage and bill committee referral.

I have to admit that the proposed approach provides only a basic setup for understanding coalition formation with ideological interest. A more sophisticated model can serve as a framework for developing the statistical models. For example, the present framework lacks

ideological learning component as opposed to theoretical models of coalition formation which consider belief updates about other player types by observing repeated coalition formation (Chalkiadakis 2007). While this setting could increase the model complexity substantially, a future development could consider incorporating it for more realistic model building.

Finally, I would like highlight that the basic framework and the issue-specific inference framework can be used to analyze a parliamentary system or a non-legislative dataset without a recorded voting institution. Not to mention parliaments in the past, even many contemporary parliaments, especially those of authoritarian regimes, do not have voting institutions or the voting results do not convey meaningful information as the voting institution takes only a perfunctory role. In this case, researchers need to identify datasets which contain meaningful information on ideological heterogeneity among the individuals. Also, the framework can be extended to model latent preference of individuals from non-legislative datasets. For instance, the framework can be applied to analyze the ideological leaning of interest groups and citizens whose usual political activities do not involve voting. I hope the present framework can be a starting point for helping the researchers to formulate their datasets for further analytic inquiry.

3.6 Appendix

3.6.1 Raw Data of KNA Bill Proposal Dataset

Figure 3.12 shows a typical bill proposal document archived in NALIS. By parsing the original pdf file, one can extract relevant information of each bill including bill title, proposal date, list of cosponsors, bill contents, proposal status and etc.

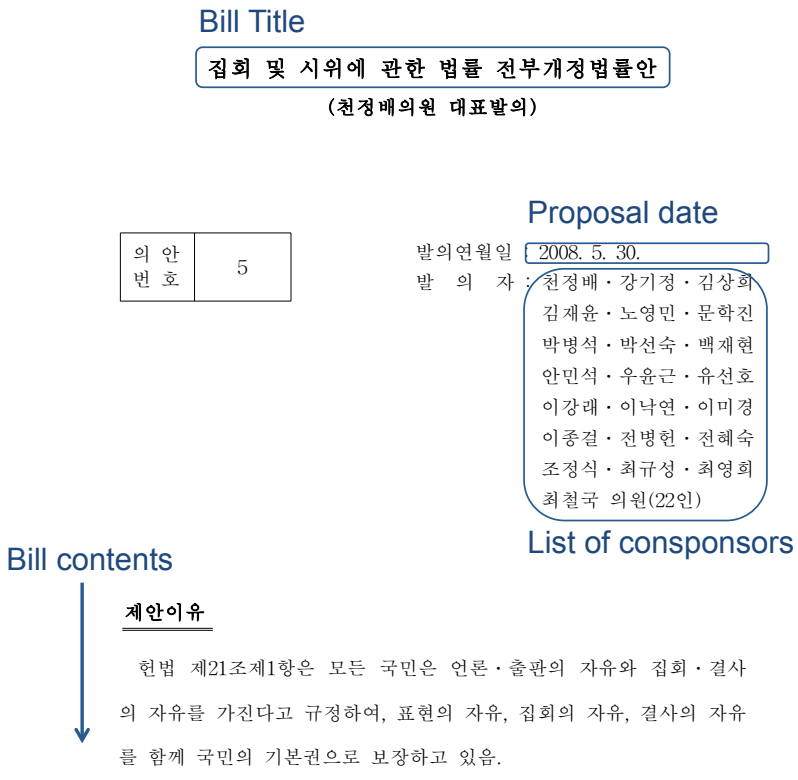

Figure 3.12: An exemplary bill document.

3.7 Calculation of Party Polarization Index

One can compute the level of party polarization for each session by comparing actual party label vector (P) of the legislators and coalition label vector (C) inferred from the bill proposal coalition dataset. Inferred coalition label is a group partition vector that is believed

to result the highest level of matrix segregation. I utilize non-parametric clustering approach for simultaneous estimation of coalition label and the segregation level corresponding to the inferred coalition label.

There are a plenty of matrix segregation methods available. I adopt a stochastic optimization algorithm for graph clustering (i.e. matrix segregation) using Metropolis-Hasting algorithm (Guimera & Amaral 2005). The algorithm tries to find a partition vector which maximizes an objective function

$$Q = \frac{1}{2m}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2m}\right)\delta(C_i, C_j),$$

where $\mathbf{A}$ is an affinity matrix, $\mathbf{k}$ is a vector corresponding to the array of column sum of $\mathbf{A}$, $2m$ is the total sum of elements in $\mathbf{A}$, and $\delta(C_i, C_j)$ is a Kronecker delta which becomes 1 if and only if nodes i and j belong to the same coalition.

The algorithm maximizes Q following a simulated annealing path with small backward transition probability for inferior solutions in order to escape from local maxima. To be specific it jumps to a new state (i.e. flipping the coalition membership of an individual) with probability 1 if the value of Q of the new membership (partition) vector is greater or equal to the value of Q for the previous membership vector. If the new partition vector results smaller value of Q than the previous one, then it moves to the new state with probability $\exp -\frac{Q^{new}-Q^{old}}{T}$ with T representing the temperature of the optimization system which decreases as the optimization routine repeats. As a result, the algorithm searches heterogeneous partition space and gradually converges to search a narrow region corresponding to the global maximum.

Because the quality functional landscape of matrix partitions is known to be extremely rugged (Fortunato 2010), I generate multiple coalition partition vectors to quantify the level of segregation in the cosponsorship records of each session.

Using the inferred partition vector (C), I compute the level of party polarization by

measuring the similarity between the partition vector (C) and the actual party affiliation vector (P). In this way, one can quantify to what extent the level of polarization in the dataset is explained by party label. In particular, the metric I employ, variation of information, is an index which always scales between 0 and 1. It hits its minimum value 0 when the two partitions are identical and hits 1 when the two partitions are different at maximum. Since it is more convenient to have higher value when party polarization level is high, I use $1-VI(P,C)$ to represent the level of party polarization. Variation of information is defined as

$$VI(P,C) = \frac{H(P)+H(C)-2MI(P,C)}{\log|P|}$$

where variation of information $VI(P,C)$ indicates variation of information between partition vectors P and C (Meila 2003). $H(P)$ is the entropy of party label vector and $MI(P,C)$ is the mutual information between the pair of partitions P and C. Variation of information decreases as the similarity between two partitions grows. When two partitions are equal, the value becomes 0. The maximum possible value of VI is adjusted to 1 by dividing it with $\log|P|$.

3.7.1 Phrase Mining Approach for Automated Bill-Issue Matching

I adopt a text modeling framework for bill-issue (i.e. document-topic) matching specialized for short content-representative documents (i.e. titles).

An extant number of algorithms have been proposed to infer semantic characteristics of unstructured text data (Grimmer & Stewart 2013, Salton & Buckley 1988). Yet, the majority of models exhibit poor performance for short documents. This problem is due to single term (i.e. uni-gram) level modeling framework incorporated in the most of the models. Contrary to this approach, the topical hierarchy framework enables phrase (i.e. a phrase consists of multiple terms) level modeling. This treatment enables topic extraction of short content-representative documents. The first successful application of this algorithm

was sub-field classification of the titles of computer science articles (Wang et al. 2013).

In order to extract issue-relevant information, each bill title is parsed into multiple terms using HAM Korean natural language processor implemented in C++ (Kang 1996). The obtained M by W matrix, $\mathbf{B}$ contains information on which words are used in each bill title: $B_{ij} = 1$ if jth term appears in ith bill title.

On the basis of the bag of words conception which states that terms belonging to similar topics have a high frequency of co-occurrence, the algorithm starts from building a term co-occurrence matrix $\mathbf{G}$ which is defined as a product between $\mathbf{B}^T$ and $\mathbf{B}$. Thus $\mathbf{G}$ is a W by W matrix where G_{ij} denotes the number of times term i and j co-occur in same bill titles. Notable distinction of this approach to other single term level modeling scheme is that the topic hierarchy model uses the projection matrix $\mathbf{G}$, which represents word-to-word relationship, instead of the word-to-document matrix $\mathbf{B}$. As a result, the topic hierarchy approach incorporates bi-term level association as a basic unit of information in the dataset.

The core idea of topic hierarchy extraction approach is to decompose the original matrix $\mathbf{G}$ into a set of sub-matrices, $\{\mathbf{G}^1, \ldots, \mathbf{G}^K\}$. For example, $\mathbf{G}^1$ embodies the list of term associations that are expected to represent the 1st topic. The algorithmic procedure is to discover a dot product representation of $\mathbf{G}^z$, so that the random variable $G_i j^z \sim$ Poisson$(\rho^z \theta_i^z \theta_j^z)$ and $\sum_{i=1}^{W} \theta_i^z = 1$. By imposing non-negativity constraint on ρ^z and θ^z vectors, the estimation procedure for obtaining $\widehat{\mathbf{G}^z}$ follows a standard Non-negative Matrix Factorization (NMF) algorithm of which a general estimation scheme is proposed by Lee & Seung (1999).

The specific Expectation-Maximization procedure proceeds as follow. Due to the expectation property of the Poisson distribution, the expected value of G_{ij}^z, $E(G_{ij}^z) = \rho_z \theta_i^z \theta_j^z$, and we get a simple relationship, $E(\sum_z G_{ij}^z) = \rho_z \sum_i \theta_i^z \sum_j \theta_j^z = \rho_z$. The original term co-occurence matrix $\mathbf{G}$, which is the only directly observable data in the estimation procedure, can be represented as $G_{ij} = \sum_z G_{ij}^z \sim$ Poisson$(\sum_z \rho_z \theta_i^z \theta_j^z)$.

For a given set of parameters θ and ρ, the likelihood of observing $\mathbf{G}$ is

$$p(\mathbf{G}|\theta,\rho) = \prod_{i,j} p(G_i j|\theta_i,\theta_j,\rho) = \prod_{i,j} \frac{(\sum_z \rho_z \theta_i^z \theta_j^z)_{ij}^G \exp(-\sum_z \rho_z \theta_i^z \theta_j^z)}{G_{ij}!}.$$

In order to maximize $p(\mathbf{G}|\theta,\rho)$, E-step and M-step of the EM algorithm retrieve maximum likelihood estimates of Poisson distribution by employing

$$\text{E-step: } \widehat{G_{ijz}} = \frac{\rho_z \theta_i^z \theta_j^z}{\sum_t \rho_t \theta_i^t \theta_j^t}$$

$$\text{M-step: } \rho_z = \sum_{ij} \widehat{G_{ijz}},\ \theta_i^z = \frac{\sum_j \widehat{G_{ijz}}}{\rho_z}.$$

Through iterative updating of $\widehat{G_{ijz}}$, ρ_z and θ_i^z we obtain a K by 1 vector for each phrase, which represents the corresponding phrase's probability of association to each topic. For a single term phrase i, θ_i^z is the probability i belonging to topic z and ρ_z is a normalizing constant representing the density of bi-term associations in topic z. For the case of a phrase in arbitrary length, $\mathbf{P} = \{w_{x_1},,w_{x_n}\}$, the probability that it belongs to topic z equals $\frac{\rho_z \prod_{i=1}^n \theta_{x_i}^z}{\sum_t \rho_t \prod_{i=1}^n \theta_{x_i}^t}$. And the expected number of topical frequency for $\mathbf{P}$ becomes the product of phrase frequency and this value:

$$f_z(\mathbf{P}) = f_{par(z)}(\mathbf{P}) \frac{\rho_z \prod_{i=1}^n \theta_{x_i}^z}{\sum_t \rho_t \prod_{i=1}^n \theta_{x_i}^t}$$

where $f_{par(z)}(\mathbf{P})$ is the number of times $\mathbf{P}$ appears in the entire set of documents. For a given topic, to extract its representative phrases, I use this topical frequency measure so that a phrase exhibiting a large topical frequency value in topic z is more likely to be a representative phrase of topic z. For details of the ranking procedure see Wang et al. (2013).

To select the optimal number of topics, I choose one with the lowest Bayesian Information Criterion (BIC) value after running the algorithm with different number of

topic clusters (Gelman et al. 2014). BIC is defined as $-2LL + M\log(W)$ where LL is the maximum value of the likelihood function $p(\mathbf{G}|\theta,\rho)$ and M and W respectively denote the number of bill titles and the number of words.

By employing the topic hierarchy framework introduced above, I compute topic-association probability of terms and phrases which appear in bill titles. For 82,452 bill titles and 6,871 terms analyzed, the minimum BIC is obtained when the number of clusters is 12 (Figure 3.13). Table 1 shows the highest ranking phrases belonging to each topic. Terms without issue-relevant information (e.g. amendment, special-act) are removed. For presentational purpose, Korean words are translated into English after finishing all algorithmic procedures.

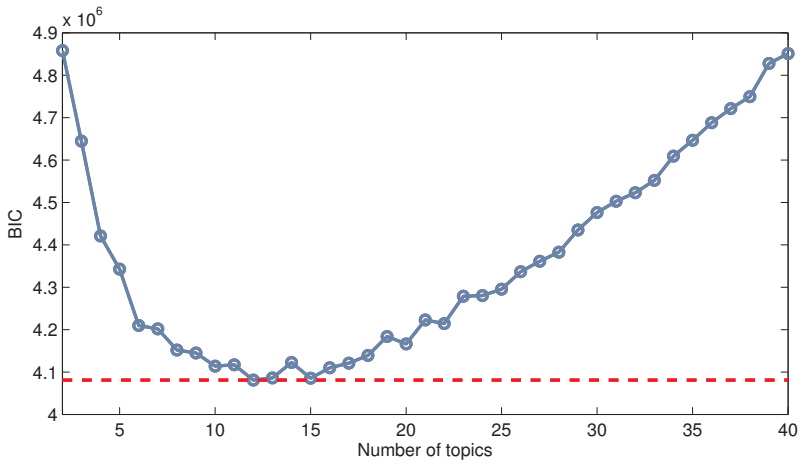

Figure 3.13: The number of topic clusters versus BIC. Minimum value is obtained when the number of the clusters is 12.

3.7.2 Issue-specific Ideology Inference

For the recovery of issue-specific ideology, I adopt the canonical idea of generalized Singular Vale Decomposition (SVD) (Jolliffe 2002). Generalized SVD extends SVD by decomposing a subset of the original matrix through reweighting the original matrix. In the present case, I reweight the affinity matrix $\mathbf{A}$ using the information obtained in the

topic modeling stage to retrieve issue-specific ideology estimates. For a given issue s, we have an m by 1 vector p^s whose element denotes the probability that to what extent the corresponding bill is about issue s. By applying the proposed ideal point method on the reweighted (issue-specific) affinity matrix $\mathbf{A}^s$ where $A^s_{ij} \equiv \sum_b p^s_b \mathbb{1}(i, j \in C^b)|C^b|^{-\gamma}$, I obtain issue-specific ideology $\mathbf{X}^s$ of the legislators.

Chapter 3 is currently in preparation for submission for publication of the material. Sohn, Yunkyu. The dissertation author was the sole researcher and author of this material.

Bibliography

Albert, James H. 1992. "Bayesian estimation of normal ogive item response curves using Gibbs sampling." *Journal of Educational and Behavioral Statistics* 17(3):251–269.

Alemán, Eduardo, Ernesto Calvo, Mark P. Jones & Noah Kaplan. 2009. "Comparing Cosponsorship and Roll-Call Ideal Points." *Legislative Studies Quarterly* 34(1):87–116.

Armstrong, David A, Ryan Bakker, Royce Carroll, Christopher Hare, Keith T Poole & Howard Rosenthal. 2014. *Analyzing spatial models of choice and judgment with R*. CRC Press.

Arnold, R Douglas. 1992. *The logic of congressional action*. Yale University Press.

Axelrod, Robert. 1967. "Conflict of interest: an axiomatic approach." *Journal of Conflict Resolution* pp. 87–99.

Axelrod, Robert. 1981. "The emergence of cooperation among egoists." *American political science review* 75(02):306–318.

Bafumi, Joseph, Andrew Gelman, David K Park & Noah Kaplan. 2005. "Practical issues in implementing and understanding Bayesian ideal point estimation." *Political Analysis* pp. 171–187.

Bai, Zhidong & Jack W Silverstein. 2010. *Spectral analysis of large dimensional random matrices*. Vol. 20 Springer.

Bailey, Michael A. 2007. "Comparable preference estimates across time and institutions for the court, congress, and presidency." *American Journal of Political Science* 51(3):433–448.

Barberá, Pablo. 2015. "Birds of the Same Feather Tweet Together. Bayesian Ideal Point Estimation Using Twitter Data." *Political Analysis* 23(1):76–91.

Bartels, Larry M. 1991. "Constituency Opinion and Congressional Policy Making: The Reagan Defense Buildup." *American Political Science Review* 85(02):457–474.

Bateman, David A & John Lapinski. 2016. "Ideal Points and American Political Development: Beyond DW-NOMINATE." *Studies in American Political Development* pp. 1–25.

Bishop, Christopher M. 2006. *Pattern recognition and machine learning*. springer.

Bock, R Darrell & Marcus Lieberman. 1970. "Fitting a response model forn dichotomously scored items." *Psychometrika* 35(2):179–197.

Boix, Carles & Milan W Svolik. 2013. "The foundations of limited authoritarian government: Institutions, commitment, and power-sharing in dictatorships." *The Journal of Politics* 75(02):300–316.

Bond, Robert & Solomon Messing. 2015. "Quantifying Social Media's Political Space: Estimating Ideology from Publicly Revealed Preferences on Facebook." *American Political Science Review* 109(01):62–78.

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

Botev, Z. I. 2016. "The normal law under linear restrictions: simulation and estimation via minimax tilting." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* pp. n/a–n/a.

Brady, Henry E. 1989. "Factor and ideal point analysis for interpersonally incomparable data." *Psychometrika* 54(2):181–202.

Brooks, Stephen P & Andrew Gelman. 1998. "General methods for monitoring convergence of iterative simulations." *Journal of computational and graphical statistics* 7(4):434–455.

Caliński, Tadeusz & Jerzy Harabasz. 1974. "A dendrite method for cluster analysis." *Communications in Statistics-theory and Methods* 3(1):1–27.

Chalkiadakis, Georgios. 2007. "A Bayesian Approach to Multiagent Reinforcement Learning and Coalition Formation under Uncertainty.".

Chib, Siddhartha. 1998. "Estimation and comparison of multiple change-point models." *J Econom* 86(2):221–241.

Chikuse, Yasuko. 2012. *Statistics on special manifolds*. Vol. 174 Springer Science & Business Media.

Chung, Fan RK. 1997. *Spectral graph theory*. Vol. 92 American Mathematical Soc.

Clark, Tom S & Benjamin Lauderdale. 2010. "Locating Supreme Court opinions in doctrine space." *American Journal of Political Science* 54(4):871–890.

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

Clinton, Joshua, Simon Jackman & Douglas Rivers. 2004. "The statistical analysis of roll call data." *American Political Science Review* 98(02):355–370.

Converse, Philip E. 1962. *The nature of belief systems in mass publics*. Survey Research Center, University of Michigan Ann Arbor.

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

Crespin, Michael H & David W Rohde. 2010. "Dimensions, issues, and bills: Appropriations voting on the House floor." *The Journal of Politics* 72(04):976–989.

De Ayala, Rafael Jaime. 2013. *The theory and practice of item response theory*. Guilford Publications.

Demange, Gabrielle. 2005. *Group formation: The interaction of increasing returns and preferences diversity*. Cambridge: Cambridge University Press.

Desposato, Scott W., Matthew C. Kearney & Brain F. Crisp. 2011. "Using Cosponsorship to Estimate Ideal Points." *Legislative Studies Quarterly* 36(4):531–565.

Dewan, Torun & Arthur Spirling. 2011. "Strategic opposition and government cohesion in Westminster democracies." *American Political Science Review* 105(02):337–358.

Enelow, James & Melvin J Hinich. 1981. "A new approach to voter uncertainty in the Downsian spatial model." *American journal of political science* pp. 483–493.

Fariss, Christopher J. 2014. "Respect for human rights has improved over time: Modeling the changing standard of accountability." *American Political Science Review* 108(2):297–318.

Fortunato, Santo. 2010. "Community detection in graphs." *Phys Rep* 486(3):75–174.

Fowler, James H. 2006. "Connecting the Congress: A study of cosponsorship networks." *Political Analysis* 14(4):456–487.

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

Gamm, Gerald & Thad Kousser. 2013. "No Strength in Numbers: The Failure of Big-City Bills in American State Legislatures, 1880–2000." *American Political Science Review* 107(04):663–678.

Gelman, Andrew, John B Carlin, Hal S Stern & Donald B Rubin. 2014. *Bayesian data analysis*. Chapman & Hall/CRC.

Gentzkow, Matthew & Jesse M Shapiro. 2010. "What drives media slant? Evidence from US daily newspapers." *Econometrica* 78(1):35–71.

Geweke, John. 1991. *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*. Vol. 196 Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN, USA.

Gower, John C & Garmt B Dijksterhuis. 2004. *Procrustes problems*. Oxford University Press on Demand.

Grimmer, Justin & Brandon M Stewart. 2013. "Text as data: The promise and pitfalls of automatic content analysis methods for political texts." *Political Analysis* p. mps028.

Grimmer, Justin, Solomon Messing & Sean J Westwood. 2012. "How words and money cultivate a personal vote: The effect of legislator credit claiming on constituent credit allocation." *American Political Science Review* 106(04):703–719.

Guimera, Roger & Luis A Nunes Amaral. 2005. "Functional cartography of complex metabolic networks." *Nature* 433(7028):895–900.

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

Hix, Simon & Hae-Won Jun. 2009. "Party Behaviour in the Parliamentary Arena The Case of the Korean National Assembly." *Party Politics* 15(6):667–694.

Hoff, Peter D. 2007. "Model averaging and dimension selection for the singular value decomposition." *Journal of the American Statistical Association* 102(478).

Hoff, Peter D. 2009. "Simulation of the Matrix Bingham-von Mises-Fisher Distribution, With Applications to Multivariate and Relational Data." *Journal of Computational and Graphical Statistics* 18(2):438 – 456.

Hyvärinen, Aapo, Juha Karhunen & Erkki Oja. 2004. *Independent component analysis*. Vol. 46 John Wiley & Sons.

Imai, Kosuke, James Lo & Jonathan Olmsted. 2016. "Fast estimation of ideal points with massive data." *American Political Science Review* 110(4):631–656.

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

Jackman, Simon. 2009. *Bayesian analysis for the social sciences*. Vol. 846 John Wiley & Sons.

Jackman, Simon & Lynn Vavreck. 2010. "Primary politics: Race, gender, and age in the 2008 democratic primary." *Journal of Elections, Public Opinion and Parties* 20(2):153–186.

Jacoby, William G & David A Armstrong. 2014. "Bootstrap confidence regions for multidimensional scaling solutions." *American Journal of Political Science* 58(1):264–278.

Jolliffe, Ian. 2002. *Principal component analysis*. Wiley Online Library.

Kang, S. S. 1996. Ham: Korean morphological analyzer and automatic indexing system. Technical report.

Kim, In Song. 2017. "Political cleavages within industry: firm-level lobbying for trade liberalization." *American Political Science Review* 111(1):1–20.

Kim, In Song, John Londregan & Marc Ratkovic. 2015. "Voting, Speechmaking, and the Dimensions of Con ict in the US Senate." *Unpublished paper* .

Kim, Y. 2011. *The Politics of Coalition in South Korea*. Taylor and Francis.

Koopmans, Tjalling C. 1949. "Identification problems in economic model construction." *Econometrica, Journal of the Econometric Society* pp. 125–144.

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

Lancichinetti, Andrea, Santo Fortunato & Filippo Radicchi. 2008. "Benchmark graphs for testing community detection algorithms." *Physical Review E* 78(4):046110.

Lauderdale, Benjamin E & Tom S Clark. 2014. "Scaling politically meaningful dimensions using texts and votes." *American Journal of Political Science* 58(3):754–771.

Laver, Michael & Kenneth A Shepsle. 1996. *Making and breaking governments: Cabinets and legislatures in parliamentary democracies*. Cambridge University Press.

Laver, Michael & Norman Schofield. 1998. *Multiparty government: The politics of coalition in Europe*. University of Michigan Press.

Lee, Daniel D & H Sebastian Seung. 1999. "Learning the parts of objects by non-negative matrix factorization." *Nature* 401(6755):788–791.

Lijphart, Arend. 2012. *Patterns of democracy: Government forms and performance in thirty-six countries*. Yale University Press.

Lo, James, Sven-Oliver Proksch & Thomas Gschwend. 2014. "A common left-right scale for voters and parties in Europe." *Political Analysis* 22(2):205–223.

Lopes, Hedibert Freitas & Mike West. 2004. "Bayesian model assessment in factor analysis." *Statistica Sinica* pp. 41–67.

Lowe, Will, Kenneth Benoit, Slava Mikhaylov & Michael Laver. 2011. "Scaling policy preferences from coded political texts." *Legislative studies quarterly* 36(1):123–155.

Lowell, Abbott Lawrence. 1902. *The influence of party upon legislation in England and America*. American Historical Association.

Magnus, JR & H Neudecker. 1988. *Matrix differential calculus with applications in statistics and econometrics*. New York: Wiley.

Marschak, Jacob & Roy Radner. 1972. *Economic theory of teams*. Yale University Press New Haven.

Marshall, Monty G & Benjamin R Cole. 2014. "Global Report 2014: Conflict, governance, and state fragility.".

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

Mayhew, David R. 1974. *Congress: The electoral connection*. Yale University Press.

McCarty, Nolan. 2001. "The hunt for party discipline in congress." *American Political Science Review* 95(3):673–687.

McCarty, Nolan, Keith T Poole & Howard Rosenthal. 2016. *Polarized America: The dance of ideology and unequal riches*. mit Press.

Meila, Marina. 2003. Comparing Clusterings by the Variation of Information. In *Learning Theory and Kernel Machines: 16th Annual Conference on Computational Learning Theory and 7th Kernel Workshop, COLT/Kernel 2003, Washington, DC, USA, August 24-27, 2003, Proceedings*. Vol. 16 Springer p. 173.

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

Minka, Thomas P. 2000. Automatic choice of dimensionality for PCA. In *Nips*. Vol. 13 pp. 598–604.

Moon, Woojin. 2011. "Party votes and party loyalty in multiparty party systems: An analysis of the 17th and the 18th national assemblies in South Korea." *Journal of Legislative Studies* 17(2):7–37.

Murphy, Kevin P. 2012. *Machine learning: a probabilistic perspective*. MIT press.

Murphy, Susan A & Aad W Van der Vaart. 2000. "On profile likelihood." *Journal of the American Statistical Association* 95(450):449–465.

Nadakuditi, Raj Rao & Mark EJ Newman. 2012. "Graph spectra and the detectability of community structure in networks." *Physical review letters* 108(18):188701.

Nadakuditi, Raj Rao & Mark EJ Newman. 2013. "Spectra of random graphs with arbitrary expected degrees." *Physical Review E* 87(1):012803.

Nakajima, Shinichi, Masashi Sugiyama & S Derin Babacan. 2011. Global solution of fully-observed variational Bayesian matrix factorization is column-wise independent. In *Advances in Neural Information Processing Systems*. pp. 208–216.

Nakajima, Shinichi, Ryota Tomioka, Masashi Sugiyama & S Derin Babacan. 2012. Perfect dimensionality recovery by variational Bayesian PCA. In *Advances in Neural Information Processing Systems*. pp. 971–979.

Ng, Andrew Y, Michael I Jordan & Yair Weiss. 2002. "On spectral clustering: Analysis and an algorithm." *Advances in neural information processing systems* 2:849–856.

Nowak, Martin A. 2006. "Five rules for the evolution of cooperation." *science* 314(5805):1560–1563.

Ostrom, Elinor. 1990. *Governing the commons: The evolution of institutions for collective action*. Cambridge university press.

Peixoto, Tiago P. 2013. "Eigenvalue spectra of modular networks." *Physical review letters* 111(9):098701.

Peng, Jie & Hans-Georg Müller. 2008. "Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions." *The Annals of Applied Statistics* pp. 1056–1077.

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

Poole, Keith T. 2016. "Voteview.".
URL: *http://voteview.com/dwnl.htm*

Poole, Keith T & Howard Rosenthal. 1984. "The polarization of American politics." *The Journal of Politics* 46(04):1061–1079.

Poole, Keith T & Howard Rosenthal. 1991. "Patterns of congressional voting." *American journal of political science* pp. 228–278.

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

Poole, Keith T, Jeffrey B Lewis, James Lo & Royce Carroll. 2008. "Scaling roll call votes with w-nominate in r." *Available at SSRN 1276082* .

Putnam, Robert D, Robert Leonardi & Raffaella Y Nanetti. 1994. *Making democracy work: Civic traditions in modern Italy*. Princeton university press.

Rao, N Raj & Alan Edelman. 2008. "The polynomial method for random matrices." *Foundations of Computational Mathematics* 8(6):649–702.

Reckase, Mark. 2009. *Multidimensional item response theory*. Vol. 150 Springer.

Riker, William H. 1962. *The theory of political coalitions*. Vol. 578 Yale University Press New Haven.

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

Rohe, Karl, Sourav Chatterjee & Bin Yu. 2011. "Spectral clustering and the high-dimensional stochastic blockmodel." *Ann Stat* 39(4):1878–1915.

Rosas, Guillermo, Yael Shomer & Stephen R Haptonstahl. 2015. "No News Is News: Nonignorable Nonresponse in Roll-Call Data Analysis." *American Journal of Political Science* 59(2):511–528.

Rossi, Peter Eric, Greg Martin Allenby & Robert Edward McCulloch. 2005. *Bayesian statistics and marketing*. Wiley New York.

Rothenberg, Thomas J. 1971. "Identification in parametric models." *Econometrica: Journal of the Econometric Society* pp. 577–591.

Salton, Gerard & Christopher Buckley. 1988. "Term-weighting approaches in automatic text retrieval." *Information processing & management* 24(5):513–523.

San Martın, Ernesto & Jorge González. 2010. "Bayesian identifiability: Contributions to an inconclusive debate." *Chilean Journal of Statistics* 1(2):69–91.

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

Shepsle, Kenneth A. 1979. "Institutional arrangements and equilibrium in multidimensional voting models." *American Journal of Political Science* pp. 27–59.

Shor, Boris & Nolan McCarty. 2011. "The ideological mapping of American legislatures." *American Political Science Review* 105(03):530–551.

Sin, Gisela. 2014. *Separation of Powers and Legislative Organization*. Cambridge University Press.

Slapin, Jonathan B & Sven-Oliver Proksch. 2008. "A scaling model for estimating time-series party positions from texts." *American Journal of Political Science* 52(3):705–722.

Strøm, Kaare & Benjamin Nyblade. 2007. "Coalition theory and government formation." *The Oxford Handbook of Comparative Politics. Oxford, Oxford UniversityPress* .

Sundquist, James L. 1983. *Dynamics of the party system*. Brookings Institute.

Svolik, Milan W. 2009. "Power sharing and leadership dynamics in authoritarian regimes." *American Journal of Political Science* 53(2):477–494.

Tipping, Michael E & Christopher M Bishop. 1999. "Probabilistic principal component analysis." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61(3):611–622.

Train, Kenneth E. 2009. *Discrete choice methods with simulation*. Cambridge university press.

Treier, Shawn & Simon Jackman. 2008. "Democracy as a latent variable." *American Journal of Political Science* 52(1):201–217.

Von Luxburg, Ulrike. 2007. "A tutorial on spectral clustering." *Statistics and computing* 17(4):395–416.

Wang, Chi, Marina Danilevsky, Nihit Desai, Yinan Zhang, Phuong Nguyen, Thrivikrama Taula & Jiawei Han. 2013. A phrase mining framework for recursive construction of a topical hierarchy. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM pp. 437–445.

Wood, Andrew TA. 1994. "Simulation of the von Mises Fisher distribution." *Communications in statistics-simulation and computation* 23(1):157–164.

Zeileis, Achim, Christian Kleiber & Simon Jackman. 2008. "Regression models for count data in R." *Journal of statistical software* 27(8):1–25.

Zhu, Mu & Ali Ghodsi. 2006. "Automatic dimensionality selection from the scree plot via the use of profile likelihood." *Computational Statistics & Data Analysis* 51(2):918–930.