ℓ_1-based Bayesian Ideal Point Model for Multidimensional Politics

Sooahn Shin*
Department of Government, Harvard University
and
Johan Lim†
Department of Statistics, Seoul National University
and
Jong Hee Park‡
Department of Political Science and International Relations, Seoul National University

October 5, 2024

Abstract

Ideal point estimation methods in the social sciences lack a principled approach for identifying multidimensional ideal points. We present a novel method for estimating multidimensional ideal points based on ℓ_1 distance. In the Bayesian framework, the use of ℓ_1 distance transforms the invariance problem of infinite rotational turns into the signed perpendicular problem, yielding posterior estimates that contract around a small area. Our simulation shows that the proposed method successfully recovers planted multidimensional ideal points in a variety of settings including non-partisan, two-party, and multi-party systems. The proposed method is applied to the analysis of roll call data from the United States House of Representatives during the late Gilded Age (1891-1899) when legislative coalitions were distinguished not only by partisan divisions but also by sectional divisions that ran across party lines.

Keywords: ideal point estimation, multidimensional ideal points, ℓ_1 norm, multivariate slice sampling, US Congress

*Corresponding author. Ph.D. candidate, Department of Government, Harvard University. Cambridge, MA, USA. Email: sooahnshin@fas.harvard.edu URL: https://sooahnshin.com

†Professor, Department of Statistics, Seoul National University. Seoul, South Korea. Email: johanlim@snu.ac.kr

‡Professor, Department of Political Science and International Relations, Seoul National University. Seoul, South Korea. Email: jongheepark@snu.ac.kr

1 Introduction

Ideal point estimation is a well-established methodological problem in the social sciences that maps the preferences of actors as geometric positions in a latent space. Since the seminal work of Poole and Rosenthal (1997), social scientists have widely applied ideal point estimation methods to measure the ideology of political actors including US legislators (Poole and Rosenthal, 1997; Clinton et al., 2004), Supreme Court justices (Bailey and Chang, 2001; Martin and Quinn, 2002), US state legislators (Shor and McCarty, 2011), US voters (Bafumi and Herron, 2010), the European Parliament members (Hix et al., 2006), and members of the UN General Assembly (Voeten, 2000).[1] These substantive applications are allowed by methodological advancements such as Bayesian estimation (Clinton et al., 2004) and dynamic modeling (Martin and Quinn, 2002) to name a few. Recent developments further include major progress in the model (e.g. Goplerud, 2019; Moser et al., 2021; Binding and Stoetzer, 2022) and estimation strategy (e.g. Carroll et al., 2013; Imai et al., 2016; Peress, 2020).

Despite significant developments in ideal point estimation methods in the social sciences, existing ideal point estimation methods lack a principled method to identify *multidimensional* ideal points. The source of the problem is *rotational invariance*, which indicates a case where the rotational transformation of ideal point estimates preserves the log-likelihood of the model. Attempts to estimate multidimensional ideal points, including the two most well-known methods Nominal Three-Step Estimation (NOMINATE; Poole and Rosenthal, 1997) and Bayesian IRT (BIRT; Clinton et al., 2004), had to utilize a priori constraints. Specifically, similar to factor analysis, NOMINATE sets the principal axis of maximum variance as the first dimension and finds the second dimension that best accounts for the remaining variance. In other words, each dimension should account for different variations, and thus the estimated dimensions cannot be 'correlated'. BIRT also induces a priori constraints such as anchoring multidimensional positions of a few actors based on researchers' substantive knowledge. For example, Clinton et al. (2004) note "In d-dimensional choice spaces, $d(d+1)$ linearly independent a priori restrictions on the ideal points $\mathbf{X}$ are required

[1] Recently, ideal point estimation methods have been extended to a rich set of non-voting data including text (Slapin and Proksch, 2008; Gerrish and Blei, 2011; Vafa et al., 2020), survey responses (Tausanovitch and Warshaw, 2013), campaign contributions (Bonica, 2014), Twitter data (Barberá, 2015), and Facebook data (Bond and Messing, 2015).

for identification" (Clinton et al., 2004, 357). A similar solution is applied in Peress and Spirling (2010), where the authors study movie critics using BIRT approach.

The use of a priori constraints, however, poses a major challenge in applied research. Bateman and Lapinski (2016) note that the "confusing and residual nature of what the second dimension of NOMINATE actually captures" poses a challenge in its interpretation (Bateman and Lapinski, 2016, 152). This is because, Benoit and Laver (2012) note, latent policy dimensions are often correlated in the sense that legislators' positions on one dimension can be predicted from their positions on some other dimension. Yet NOMINATE restricts such correlated dimensions of policy space as aforementioned. Anchoring of a fixed number of ideal points in BIRT also depends on the researchers' substantive knowledge of the policy space and is prone to controversies and misspecification.

In this paper, we propose a model-based solution to the problem of identifying multidimensional ideal points. To begin with, we show that the use of ℓ_2 distance in existing ideal point estimation models obstructs the identification of multidimensional ideal points since the ℓ_2 distance between two vectors is invariant under the rotation of the vectors. The use of ℓ_1 distance, in contrast, transforms the infinite rotational invariance into the signed perpendicular rotational invariance in which the configuration of the ideal points becomes a highly tractable problem. In this regard, we propose the Bayesian ℓ_1(or Manhattan distance)-based ideal point model (BMIM). The estimation of BMIM is done by a multivariate slice sampling method.[2] We show that BMIM produces posterior estimates that contract around a small area.

We are not the first to address the problem of rotational invariance in ideal point estimation. For example, Sohn (2017) introduces a model that samples parameters from the Matrix von Mises-Fisher (MvMF) distribution to address rotational invariance. The MvMF distribution has been used in factor analysis (Hoff, 2007, 2009), where it allows for dimension-specific estimates to align along maximum variance directions without redundancy. Our proposed method differs from this approach in that it allows for correlated dimensions of the ideal points, which may ease the interpretation of the latent policy space in social science applications.

We demonstrate that our proposed method effectively recovers various configurations of multidimensional ideal points including non-partisan, two-party, and multi-party systems

[2]The software that implements the proposed method is available as an `R` package.

in our simulation study. Then, we apply our method to the roll call data analysis of the US House of Representatives during the late Gilded Age (1891-1899), an era of the "battle of the standards". Our method successfully finds that the sectional division – Southern Democrats and the Populists from the West vs. Northeastern representatives – arose as a major cleavage, along with partisan division, in this period, which confirms the existing historical narratives on the debate over gold-standard.

The rest of the article is structured as follows. In Section 2, we illustrate three motivating examples that show the limitation of existing approaches. In Section 3, we introduce BMIM. We show that the use of ℓ_1 distance transforms the invariance problem of infinite rotational turns into the signed perpendicular problem and the posterior distribution of BMIM is concentrated around the "true" ideal points as N (the number of actors) and M (the number of votes) increase. In Section 4, we conduct simulation studies with synthetic data to check whether the proposed method recovers true multidimensional ideal points in a variety of settings. In Section 5, we analyze voting data from the US House of Representatives during the late Gilded Age. Finally, we conclude with remarks and discussion in Section 6.

2 Motivating Examples

This section examines three examples that demonstrate the challenges in estimating multidimensional ideal points using conventional ideal point estimation methods.

Example 1: Small Chamber with Two Uncorrelated Dimensions

Suppose we study ideal points in the context of a small legislature with only four legislators (A1, A2, B1, and B2) who cast votes on 400 roll calls. Despite being unrealistic, this example is helpful for understanding the practical challenge posed by the identification problem of multidimensional ideal point estimation. We assume there are two political cleavages in this policy space along the alphabet dimension (A vs. B) and the number dimension (1 vs. 2), respectively. The two cleavages are equally important and hence half of the roll call votes (RC #1 to # 200) are divided along the alphabet dimension and the other half (RC #201 to # 400) is divided along the number dimension. Table 1 displays the roll call votes in a matrix format in which a row indicates a legislator and a column indicates a roll call.

Table 1: Roll Call Voting Data of Small Chamber with Two Uncorrelated Dimensions: 0 denotes nay vote and 1 denotes Yea vote.

	RC #1	...	**#100**	**#101**	...	**#200**	**#201**	...	**#300**	**#301**	...	**#400**
A1	0	...	0	1	...	1	0	...	0	1	...	1
A2	0	...	0	1	...	1	1	...	1	0	...	0
B1	1	...	1	0	...	0	0	...	0	1	...	1
B2	1	...	1	0	...	0	1	...	1	0	...	0

Figure 1 (a) shows an ideal result of ideal point estimation, where the alignments of legislators' ideal points are well identified along the number dimension (A1/B1 and A2/B2) and the alphabet dimension (A1/A2 and B1/B2). Note that a perpendicular rotation of the results would not change substantive findings. In this case, since legislators' alphabet cannot be predicted from their number, we can say that the two dimensions are uncorrelated.

Now we apply conventional ideal point estimation methods, and compare the results with the ideal case of (a). Figure 1 (b) and (c) show the result of NOMINATE and BIRT respectively, using the same roll call voting data from Table 1.[3] In the results of (b) WNOMINATE, the first coordinates of A1 (yellow circle) and B2 (brown triangle) are similar. This is problematic since A1 should be absolutely discordant with B2 since they are always voting differently. In the results of (c) BIRT, any rotation of those four ideal points would yield the same log-likelihood, meaning that we cannot determine the dimension unless we impose some ad hoc constraints. Thus, both methods fail to capture the underlying dimensions, the alphabet, and the number dimension.

Example 2: Two Voting Clusters with Correlated Dimensions

Next, we consider a more realistic case with a larger number of legislators and two correlated policy dimensions. Suppose legislators' voting behavior arises from a two-dimensional policy space that consists of economic and social dimensions. In Figure 2 (a), each point corresponds to an ideal point of a legislator who votes on a set of roll calls based on their

[3] We implemented `wnominate()` in the R package `wnominate` (Poole et al., 2011) for the results in (b) and `ideal()` in the R package `pscl` (Jackman, 2020) for those in (c).

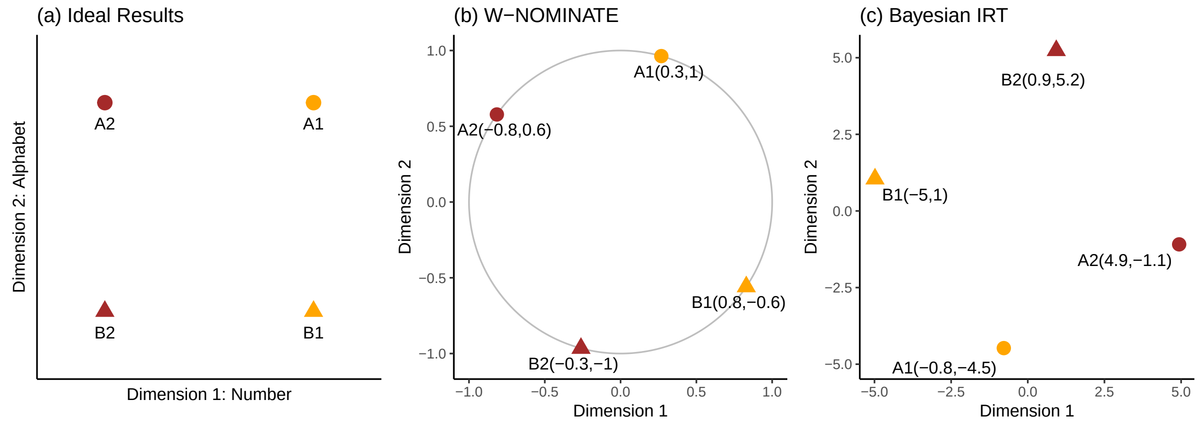

Figure 1: Estimated Ideal Points for Small Chamber with Two Uncorrelated Dimensions: Each panel shows the results of different multidimensional ideal points estimation. Each point indicates the ideal points of legislators: A1 (orange circle), A2 (brown circle), B1 (orange triangle), and B2 (brown triangle). (a) The example of ideal results where each dimension shows the cleavage by number and alphabet. (b) The results of WNOMINATE. The estimated weight of each dimension is 1 and 0.6 respectively. Note that WNOMINATE methods impose a restriction that each ideal point should lie within the unit circle (gray circle) for identification. (c) The results of BIRT without anchoring legislators.

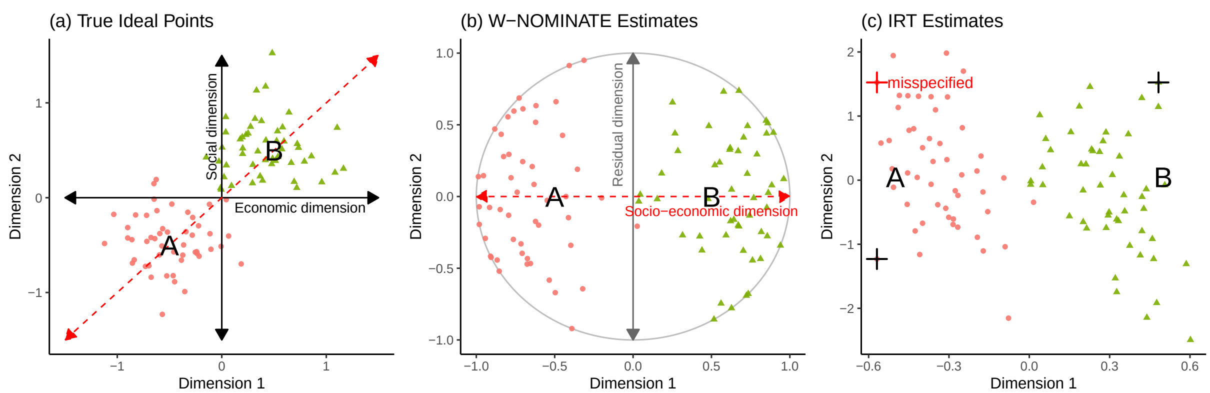

Figure 2: Two Voting Clusters with Correlated Dimensions: Red circles indicate the ideal points of synthetic legislators from a cluster A who are conservative on both economic (x-axis) and social dimensions (y-axis), whereas green triangles indicate those from cluster B liberal on both dimensions. (a) Ground truth where economic and social dimensions are the main cleavages. (b) WNOMINATE where economic and social dimensions are lumped. (c) BIRT with partial misspecification departing from the ground truth.

preferences. Here, an important thing to note is that the ideological positions on each coordinate are correlated. That is, legislators' alignment on economic and social dimensions constructs two correlated yet distinct cleavages that explain their voting behavior in this latent policy space.

In this example, we randomly generated synthetic roll-call votes based on the "true" ideal points in (a), and fit WNOMINATE and BIRT of which results are shown in (b) and (c) respectively. In (b), where we plot the estimated WNOMINATE scores of synthetic legislators, we can observe that the underlying multiple dimensions of policy space (economic and social dimensions) have been lumped into a single dimension (socio-economic dimension), hence losing important information about the essence of main cleavages. Furthermore, it poses a "loss in translation" of the second dimension, which accounts for the remaining variance (Bensel, 2016).

In (c), where we fixed 3 legislators after fitting BIRT for the identification, we assumed a partially misspecified case — among three legislators that have been used as anchors, one (red cross) has been misspecified whereas the other two (black crosses) are correctly specified. The result shows that BIRT also fails to recover the underlying policy space and the ideal points due to the misspecification of a single legislator. This shows that the results of BIRT are sensitive to the specification of anchors, thus posing a challenge in justifying the researcher's decision of anchors.

Example 3: The US House of Representatives During the Late Gilded Age

In real-world politics, the partisan division of legislators alone often cannot fully capture the underlying cleavages in latent policy space, motivating the use of multidimensional ideal point estimation. Researchers may miss a set of essential voting behaviors that jointly shape legislative politics if they only focus on a principal axis that lumps correlated yet sufficiently distinct issues in a single dimension. The politics in the US House of Representatives during the late Gilded Age, which expands from 1891 to 1899, is such an example.

During the late Gilded Age, the congressional and electoral politics in the US was centered around the battle of the standards": the adoption of a silver standard *versus* adherence to the gold standard (Bensel, 2000; Frieden, 2016). The Panic of 1893 and the rise of the Populists, two interrelated political economic turbulence, set the fire over the gold standard.

On one side, there were proponents for the free coinage of silver (the "Silverites") at a 16-1 rate against gold, implying a devaluation of the dollar that would buffer the sharp decline of farm prices and enhance the competitiveness of the products in world markets;

mostly supported by export-oriented farmers/miners from the South and the West. On the other side, there were proponents of the gold standard (the "Goldbugs") who believed that the gold standard could secure the stability of the currency. The goldbugs were concentrated in the international financial community and big cities in the Northeast. The partisan divide was not mirrored in the monetary standard debate. The division in the monetary standard was about a sectional division based on their financial interests. In the 1896 presidential election, William Jennings Bryan, the joint Democratic-Populists candidate who was against the gold standard, lost the support of the gold Democrats, whereas William Mckinley, the Republican candidate who was against the silver standard, lost the support of silver Republicans (Frieden, 2016, 118).

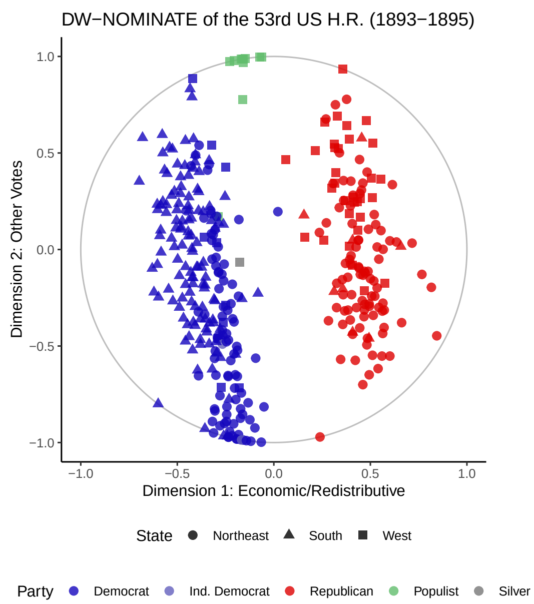

Figure 3: DW-NOMINATE Scores of the 53rd US House of Representatives (1893-1896). Each point denotes a two-dimensional ideal point of representative, measured by DW-NOMINATE method. The color indicates the party label of the representative: blue points indicate Democrats, red points Republican, green points Populists, and gray points Silver party. The shape indicates the state which each representative represents: circles indicate Northeastern states, triangles Southern states, and rectangles Western states.

Figure 3 shows DW-NOMINATE scores of the 53rd US House of Representatives (1893-1896) (Lewis et al., 2022). Shapes indicate geographical classifications of districts and colors indicate partisanship.[4] Here, only the partisan division is captured along the first

[4]Following Bensel (2000)'s classification, Southern states include all those that seceded into the Confederacy plus Kentucky, Missouri, Oklahoma, and West Virginia; Western states include all states west of Illinois and Wisconsin not included in the South; the Northeastern states include all the remaining states.

dimension, denoted as "Economic/Redistributive" dimension. Yet, the sectional division on the gold standard, the most important political debate during this period, is missing from this result.

Although the second dimension distinguishes the two major parties from the Populists (top), two observations hamper the interpretation of the second dimension as a sectional division over the monetary issue. First of all, Southern Democrats (blue triangles), who were in an anti-gold alliance with the Populists, are widely dispersed along the second dimension. Moreover, the only representative from the Silver party (gray square), Francis Griffith Newlands, is located as moderate on the second dimension, which makes the interpretation of the second dimension more confusing.

To sum up, three examples demonstrate the challenge of capturing interpretive dimensions of a latent policy space using existing multidimensional ideal point estimation methods, despite its indispensable necessity in social science studies. In the following, we show that the root cause of the problem is the use of ℓ_2 distance-based models.

3 Ideal Points Estimation in the ℓ_1-norm Space

3.1 Setup

Suppose we want to analyze the roll call votes of N legislators, $i = 1, 2, \ldots, N$, for M roll calls, $j = 1, 2, \ldots, M$. A standard multidimensional ideal point estimation can be motivated by the following spatial voting model based on random utility. Assume that each legislator i votes Yea on roll call vote j if the utility of voting Yea (U_{ijy}) is greater than that of voting Nay (U_{ijn}). Here, the Yea utility consists of a deterministic (u_{ijy}) part and a stochastic part (ϵ_{ijy}), where the deterministic part is defined by the distance between the ideal point of the legislator ($\mathbf{x}_i$) and that of the roll call ($\mathbf{o}_{yj}$) in the $s-$dimensional latent policy space, and similarly for the Nay utility ($\mathbf{x}_i, \mathbf{o}_{yj}, \mathbf{o}_{nj} \in \mathbb{R}^s$). That is, $u_{ijy} := f(d(\mathbf{x}_i, \mathbf{o}_{yj}))$ where $f(\cdot)$ is the utility function and $d(\cdot)$ is the distance measure. We assume a Gaussian random utility as follows:

$$
\begin{aligned}
U_{ijy} &= u_{ijy} + \epsilon_{ijy}, \\
U_{ijn} &= u_{ijn} + \epsilon_{ijn}
\end{aligned}
$$

Table 2: Utility Specification in WNOMINATE, BIRT, and BMIM (the Proposed Model)

Method	**Deterministic part** (u_{ijy})	**Utility function** (f)	**Distance** (d)
BMIM	$\sum_{k=1}^{s} -\lvert x_{ik} - o_{yjk}\rvert$	linear utility	ℓ_1 distance
WNOMINATE	$\beta \exp\left\{-\frac{1}{2}\sum_{k=1}^{s} w_k^2 \big(x_{ik} - o_{yjk}\big)^2\right\}$	Gaussian utility	ℓ_2 distance
BIRT	$\sum_{k=1}^{s} -\big(x_{ik} - o_{yjk}\big)^2$	quadratic utility	ℓ_2 distance

where $\epsilon_{ijy} - \epsilon_{ijn} \sim \mathcal{N}(0, 1)$. Let $y_{ij} = 1$ if legislator i vote Yea on roll call j, and $y_{ij} = 0$ otherwise. It follows that

$$
\begin{aligned}
\Pr(y_{ij} = 1) &= \Pr(U_{ijy} > U_{ijn}) \\
&= \Pr(u_{ijy} - u_{ijn} > \epsilon_{ijn} - \epsilon_{ijy}) = \Phi(u_{ijy} - u_{ijn})
\end{aligned}
$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.

Poole and Rosenthal (1997)'s NOMINATE and Clinton et al. (2004)'s BIRT are the two most well-known statistical models for multidimensional ideal point estimation.[5] WNOMINATE and BIRT both can be explained as arising from the Euclidean spatial voting model using different random utility function f: WNOMINATE uses the Gaussian utility which has a fatter tail compared to BIRT's quadratic utility. This implies that as the Yea position and Nay position move sufficiently far from the legislator's ideal point, the difference of utilities between voting Yea and Nay decreases in WNOMINATE whereas it increases in BIRT (Carroll et al., 2009, 2013).

Our proposed model departs from existing methods in its way to model the latent distance d in voting space. Table 2 shows how each method (WNOMINATE, BIRT, and BMIM) is different in its way to model the deterministic part. The main difference between BMIM and the other methods is in how we model the latent distance between the legislator's ideal points and roll call positions. See Appendix E for more discussion on the choice of utility function. In the next section, we propose our method and then show that the use of ℓ_2 norm leads to the rotational invariance problem whereas ℓ_1 norm does not.

[5]In this paper, we focus on WNOMINATE (weighted NOMINATE) among several variants of NOMINATE methods. The result in this paper can be easily generalized to DW-NOMINATE, a dynamic version of the method. See Section 5.3.2 of Armstrong et al. (2020) for more details about different types of NOMINATE methods.

3.2 ℓ_1 Distance and Rotation Invariance

We first specify the linear utility function (f) of a spatial voting model using ℓ_1 distance measure (d):

$$U_{ijy} = \underbrace{\sum_{k=1}^{s} -\left|x_{ik} - o_{yjk}\right|}_{u_{ijy}} + \epsilon_{ijy}, \quad \epsilon_{ijy} \sim \mathcal{N}(0,1) \tag{3.1}$$

$$U_{ijn} = \underbrace{\sum_{k=1}^{s} -\left|x_{ik} - o_{njk}\right|}_{u_{ijn}} + \epsilon_{ijn}, \quad \epsilon_{ijn} \sim \mathcal{N}(0,1). \tag{3.2}$$

A voting model is followed from the above specification:

$$\begin{aligned}
\Pr(y_{ij} = 1) &= \Pr(U_{ijy} > U_{ijn}) && (3.3)\\
&= \Pr(\sum_{k=1}^{s} -\left\{\left|x_{ik} - o_{yjk}\right| - \left|x_{ik} - o_{njk}\right|\right\} > \epsilon_{ijn} - \epsilon_{ijy}) && (3.4)\\
&= \Phi(\sum_{k=1}^{s} -\left\{\left|x_{ik} - o_{yjk}\right| - \left|x_{ik} - o_{njk}\right|\right\}). && (3.5)
\end{aligned}$$

Recall that the non-stochastic utility of voting Yea is a function of the distance between the ideal point of the legislator and that of the roll call: $u_{ijy} = f(d(\mathbf{x}_i, \mathbf{o}_{nj}))$. Thus the log-likelihood of the spatial voting model can be expressed as a function of u_{ijy} and u_{ijn}. If there exist multiple sets of ideal points $\{\mathbf{x}_i, \mathbf{o}_{nj}, \mathbf{o}_{yj}\}$ of which log-likelihood is exactly the same, we cannot identify the ideological configuration under this model. Existing literature lists three invariance conditions of log-likelihood (Sohn, 2017):

1. Rotation invariance exists if $f(d(\mathbf{x}_i, \mathbf{o}_{nj})) = f(d(A\mathbf{x}_i, A\mathbf{o}_{nj}))$ for some 2×2 matrix A.[6]

2. Addition invariance exists if $f(d(\mathbf{x}_i, \mathbf{o}_{nj})) = f(d(\mathbf{x}_i + \mathbf{c}, \mathbf{o}_{nj} + \mathbf{c}))$ for some vector $\mathbf{c} \in \mathbb{R}^2$.

3. Scaling invariance exists if $f(d(\mathbf{x}_i, \mathbf{o}_{nj})) = f(d(c\mathbf{x}_i, c\mathbf{o}_{nj}))$ for a scalar $c \in \mathbb{R}$.

[6]From here, we assume the number of latent dimension to be 2 ($s = 2$) when we discuss applications. When we discuss a model, we denote the number of latent dimension as s for generality.

It is well known that addition and scaling invariances can be easily resolved by normalizing the ideal point estimates or by an informative prior (Bafumi et al., 2005). In contrast, rotational invariance has no easy fix. For example, researchers of BIRT impose the so-called Kennedy-Helms restriction that puts Kennedy at -1 and Helms at $+1$. WNOMINATE sets the principal axis of maximum variance as the first dimension and finds the second dimension that best accounts for the remaining variance, similar to factor analysis.

Figure 4 illustrates how we tackle the rotation invariance problem using ℓ_1 distance. As illustrated in Figure 4 (a), if we use ℓ_2 distance for $d(\cdot)$, any arbitrary rotation of legislator ($\mathbf{x}_1$; red circle) and nay position ($\mathbf{o}_{n1}$; blue square) would preserve the distance between those two points, i.e. $d(\mathbf{x}_1, \mathbf{o}_{n1}) = d(A\mathbf{x}_1, A\mathbf{o}_{n1})$, for any rotation matrix A or an orthogonal matrix with determinant 1 equivalently. Thus, an infinite number of sets of two such rotated points (a red circle and blue square connected with a line) would yield the same value of u_{11n}, implying that $|\{A\}| = \infty$. In contrast, if we use ℓ_1 distance for $d(\cdot)$, only the signed perpendicular rotations of two points as in (b) would yield the same distance, and hence $|\{A\}| = 8$. In short, the total number of likelihood-preserving rotations and reflections is reduced from ∞ to 8 as we model $d(\cdot)$, instead of ℓ_2 as in BIRT and WNOMINATE, with ℓ_1 distance.

Figure 5 further illustrates how we tackle the rotation invariance problem by ℓ_1 distance and Bayesian framework, using Example 1 from Section 2 in which four legislators (A1, A2, B1, and B2) are divided by the alphabet and the number divisions. The three panels in Figure 5 show that the identification of multidimensional ideal points is crucially dependent upon the choice of ℓ_p. In this specific case with four legislators, what determines the configuration of ideal points are tangent points of a contour of the probability density function of the bivariate normal distribution (our prior) with a sphere of ℓ_p norm. Intuitively speaking, this is because the direction of maximizing the bivariate normal *prior* centered at the origin is opposite of the direction of maximizing *likelihood* function. We further illustrate each of these two components below.

The blue circled contour lines in each panel of Figure 5 indicate that the probability of ideal points under a bivariate normal prior has its maximum (the darkest point) at the mean. Thus, in the absence of strong information from data, the prior pulls ideal points toward the prior mean. Imagine a circle on which a point would have a probability of ϵ, an arbitrarily small and strictly positive real number, under the standard bivariate normal

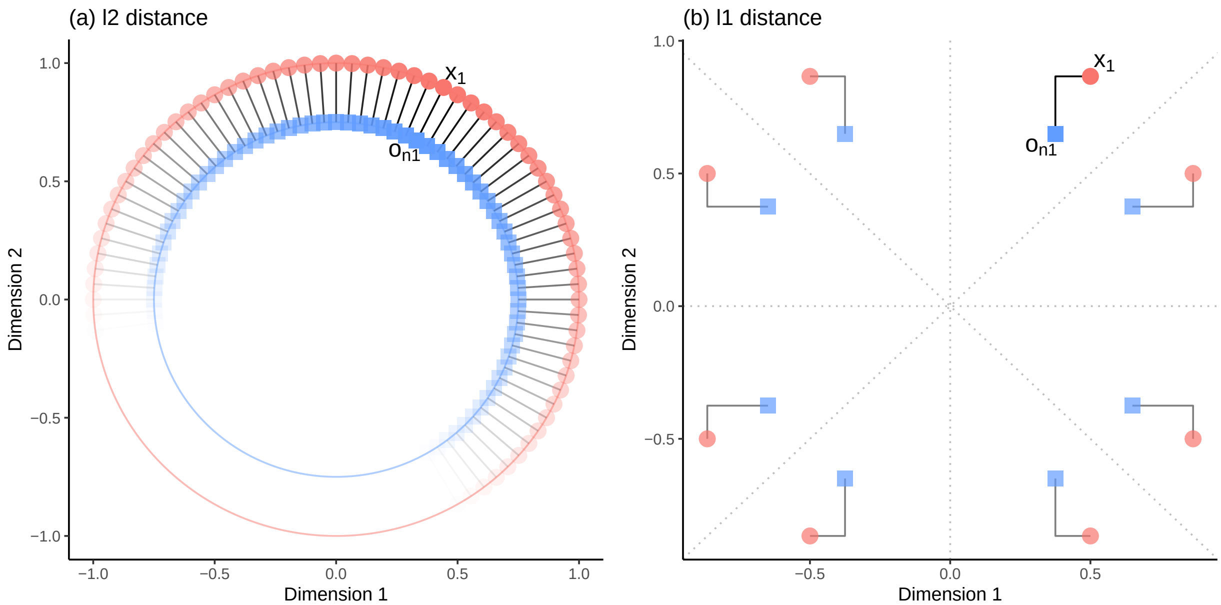

Figure 4: Illustration of Rotation Invariance Problem. Red circles indicate possible positions of the legislator ($\mathbf{x}_1$), and blue squares indicate those of nay position ($\mathbf{o}_{n1}$). That is, each pair of a red circle and blue square connected with a black line (which tracks ℓ_2 distance in (a) and ℓ_1 distance in (b)) yields the same likelihood of the model. (a) The total number of likelihood-preserving rotations/reflections is ∞ for ℓ_2 based model (e.g. WNOMINATE and BIRT). (b) It is reduced to 8 for the ℓ_1 based model (the proposed model).

distribution: $\{\mathbf{x} \in \mathbb{R}^2 : \phi_2(\mathbf{x}) = \epsilon\}$. Under the bivariate normal prior, the ideal points of two opposing legislators, A1 and B2 (A2 and B1), fall within this circle.

Next, recall that A1 and B2 (A2 and B1) always vote differently as defined in Table 1. Accordingly, each model in Figure 5 attempts to locate the ideal point of A1 (B1) farthest from that of B2 (A2) in order to maximize the likelihood, where the definition of "farthest" depends on the choice of ℓ_p. To illustrate this, imagine a sphere of ℓ_p norm tangent to the aforementioned circle outward: $\{\mathbf{x} \in \mathbb{R}^2 : \left\|\mathbf{x} - \mathbf{0}\right\|_p = c\}$ where $\mathbf{0} = (0, 0)$ and c is a scalar. Any pair of tangent points pointing in opposite directions would have the greatest distance between them within the circle. To put it differently, the tangent points represent the geometric locations where two ideal points can be placed based on both the prior information and the observed voting pattern, with A1 (A2) opposing B2 (B1) being the most desirable configuration.

Three panels of Figure 5 illustrate how these tangent points vary by the choice of ℓ_p in the case of four legislators: (a) $p = 1$, (2) $p = 2$, and (c) $p = \infty$, respectively. In panel (a), the sphere of ℓ_1 norm (red diamond) is tangent to the prior (blue) contours at the midpoints of each side. In this way, the proposed method successfully recovers "configurations" of four legislators' ideal points.[7]

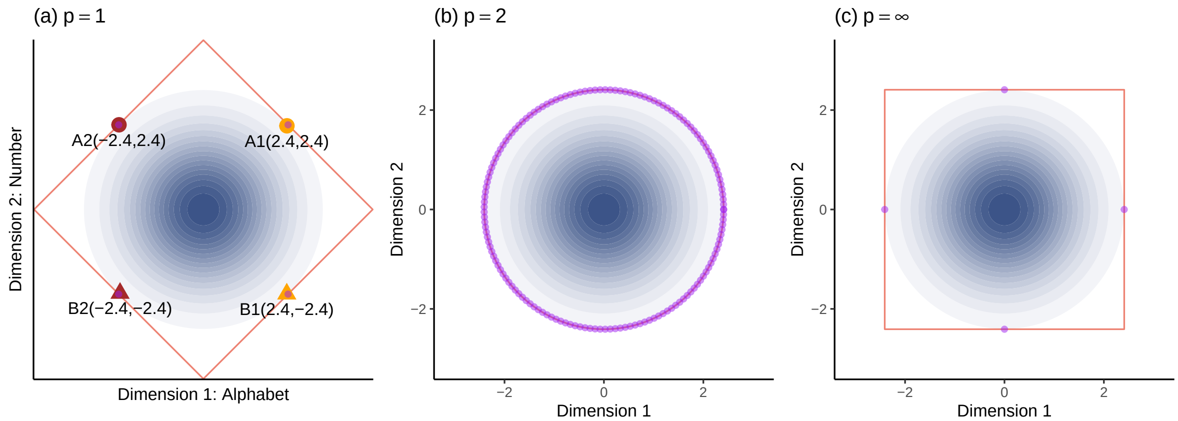

Figure 5: Illustration of Cleavages Captured by Different Choices of Distance Measure with Example 1 (Section 2). Each panel shows the tangent points of a sphere of ℓ_p norm and a contour of pdf of the bivariate normal distribution with different choices of p. In each panel, blue contours indicate the contours of the pdf of a bivariate normal distribution, of which the mean is $(0, 0)$ and the covariance matrix is an identity matrix. Red lines indicate a sphere of ℓ_p norm: $\{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x} - \mathbf{0}\|_p = 4.8\}$ with different choice of p. Purple circles indicate tangent points of the outer contour and the sphere. (a) A choice of $p = 1$ (i.e. ℓ_1 distance), as in the proposed model (BMIM). This panel also shows the estimation results of BMIM using the toy example (Example 1). Each point indicates the ideal points of legislators: A1 (orange circle), A2 (brown circle), B1 (orange triangle), and B2 (brown triangle). The sphere (red lines) is in a diamond shape and there exist four tangent points (purple circles) between the sphere and the blue contour. (b) A choice of $p = 2$ (i.e. ℓ_2 distance), as in NOMINATE and BIRT. The sphere (red lines) is in a circle shape and there exist infinite tangent points (purple circles) between the sphere and blue contour since these two overlap. (c) A choice of $p = \infty$ (i.e. ℓ_∞ distance). The sphere (red lines) is in a square shape and there exist four tangent points (purple circles) between the sphere and the blue contour.

If we use ℓ_2 distance (panel (b)), the sphere takes a shape of a circle that overlays over the blue contours. There is an *infinite* number of pairs of tangent points that are in opposite directions and have the same distance on the circle. There is no principled way to choose one configuration in the absence of a priori constraints.

The sphere of ℓ_∞ distance takes a shape of a square, as shown in (c). Tangent points to the blue contours are located at the *midpoints* of each side (purple dots). Although the four positions of legislators are identifiable (to the perpendicular rotation) in this case, the recovered dimensions (or cleavages) are significantly distorted from the ground truth concerning the underlying cleavages. To be specific, in panel (c), suppose that the ideal

[7] We use the term "configurations" here because four legislators' ideal points in panel (a) are invariant to the *signed permutation*.

points of A1 and B2 are located at the midpoints of the first dimension (top and bottom dots). Then, by construction, the remaining two opposing legislators' (A2 and B1) ideal points are located at the midpoints of the second dimension (leftmost and rightmost dots). This configuration tells us that (1) A1 and B2 (A2 and B1) have identical ideal points in the first (second) dimension and (2) A2 and B1 (A1 and B2) have the most extreme ideal points in the second (first) dimension. None of these inferential results is consistent with the underlying cleavages.

Thus, it is clear in this example that the use of an ℓ_1 distance-based likelihood function and a normal prior allows us to avoid the rotational invariance problem by reducing the likelihood-preserving rotational turns from infinity possible cases to perpendicular turns while preserving a desired configuration of the underlying cleavages. We focused on the case with four legislators to illustrate the intuition behind it, but the reduction of the likelihood-preserving rotational turns holds in general. In the following, we formally state the theoretical justifications for our BMIM.

3.3 Model

We assume $\mathbf{x}_i$, $\mathbf{o}_{yj}$, and $\mathbf{o}_{nj}$, $i = 1, 2, \ldots, N, j = 1, 2, \ldots, M$, are jointly sampled from the multidimensional standard normal distribution under the constraint $\sum_i \mathbf{x}_i = \mathbf{0}_s$, where $\mathbf{0}_s$ is the $s-$dimensional vector with all 0 elements. To speak more precisely,

$$\boldsymbol{\theta} \sim \frac{1}{\phi_s(\mathbf{0}_s)} \prod_{i=1}^{N} \phi_s\big(\mathbf{x}_i\big) \prod_{j=1}^{M} \phi_s\big(\mathbf{o}_{yj}\big) \prod_{j=1}^{M} \phi_s\big(\mathbf{o}_{nj}\big) \cdot \delta\Big(\sum_i \mathbf{x}_i\Big), \tag{3.6}$$

where $\boldsymbol{\theta} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N, \mathbf{o}_{y1}, \mathbf{o}_{y2}, \ldots, \mathbf{o}_{yM}, \mathbf{o}_{n1}, \mathbf{o}_{n2}, \ldots, \mathbf{o}_{nM})$ is an $(N + 2M)s \times 1$ vector, $\phi_s\big(\mathbf{z}\big) = (2\pi)^{-s/2} \exp\big(-\|\mathbf{z}\|_2^2/(2)\big)$ for $\mathbf{z} \in \mathcal{R}^s$ and δ is a $s-$dimensional dirac measure centered at 0. Hereinafter, we write $\boldsymbol{\theta}$ as $\big\{\mathbf{x}_i, \mathbf{o}_{yj}, \mathbf{o}_{nj}\big\}$ if necessary.

We then have $\Pr\big(\boldsymbol{\theta} \mid \mathbf{Y}\big)$, the posterior distribution of $\boldsymbol{\theta}$ given $\mathbf{Y} = \big\{y_{ij}, i = 1, 2, \ldots, N, j =$

$1, 2, \ldots, M\}$, is proportional to

$$\prod_{i=1}^{N}\prod_{j=1}^{M}\left\{\Phi(u_{ijy}-u_{ijn})^{y_{ij}}\big(1-\Phi(u_{ijy}-u_{ijn})\big)^{1-y_{ij}}\right\}$$
$$\times\left\{\prod_{i=1}^{N}\phi_s(\mathbf{x}_i)\prod_{j=1}^{M}\phi_s(\mathbf{o}_{yj})\prod_{j=1}^{M}\phi_s(\mathbf{o}_{nj})\right\},$$
$$= \prod_{i=1}^{N}\prod_{j=1}^{M}p_{ij}^{y_{ij}}(1-p_{ij})^{1-y_{ij}}\left\{\prod_{i=1}^{N}\phi_s(\mathbf{x}_i)\prod_{j=1}^{M}\phi_s(\mathbf{o}_{yj})\prod_{j=1}^{M}\phi_s(\mathbf{o}_{nj})\right\}, \tag{3.7}$$

where $p_{ij} = \Pr(Y_{ij} = 1) = \Phi(u_{ijy} - u_{ijn})$.

3.4 Model Identifiability and Posterior Contraction Property

In this subsection, we provide two theoretical justifications for our BMIM. First, we show that the log-likelihood function in Equation (3.9) below and also the posterior distribution (3.7) are identifiable up to the signed permutation of the $s-$dimensional ideal point space. Second, we prove the posterior distributions of ideal point estimation models are concentrated around the "true" ideal points as N and M increase.

Let us start with the definition of the signed permutation transformation of the ideological space with dimension s, under which the log-likelihood function is invariant. Let $\mathbf{z}$ be a point in the $s-$dimensional ideological space, for example, $\mathbf{z}$ could be either $\mathbf{x}_i$, $\mathbf{o}_{yj}$, or $\mathbf{o}_{nj}$. We consider the location shift and signed permutation transformation of $\mathbf{z}$ that is

$$\mathbf{z} \to \mathbf{z}' = \mathrm{P}(\mathbf{z}_i + \Delta), \tag{3.8}$$

where P is a signed permutation matrix and $\Delta \in \mathcal{R}^s$. Here, a matrix $\mathrm{P} = (p_{kl}, k, l = 1, 2, \ldots, s)$ is a signed permutation matrix if and only if p_{kl}'s are 0 or ± 1 and $\sum_{k=1}^{s}|p_{kl}| = \sum_{l=1}^{s}|p_{kl}| = 1$ for every $k, l = 1, 2, \ldots, s$. The total number of $s-$dimensional signed permutation matrices is finite as $2^s \cdot s!$. For $s = 2$, there are $2^2 \cdot 2! = 8$ different matrices, which are $[1, 0; 0, 1]$, $[-1, 0; 0.1]$, $[1, 0; 0, -1]$, $[-1, 0; 0, -1]$, $[0, 1; 1, 0]$, $[0, -1; 1, 0]$, $[0, 1; -1, 0]$, and $[0, -1; -1, 0]$.

The log-likelihood function of $\big(\{\mathbf{x}_i\}, \{\mathbf{o}_{yj}\}, \{\mathbf{o}_{nj}\}\big)$ and β is

$$\ell\Big(\big(\{\mathbf{x}_i\}, \{\mathbf{o}_{yj}\}, \{\mathbf{o}_{nj}\}\big), \beta\Big) = \sum_{i=1}^{N}\sum_{j=1}^{M}\left\{y_{ij}\log\left(\frac{\Phi(d_{ij})}{1-\Phi(d_{ij})}\right) - \log(1-\Phi(d_{ij}))\right\} + C \quad (3.9)$$

where C is a constant and $d_{ij} = \Phi^{-1}(p_{ij}) = u_{ijy} - u_{ijn} = \beta\|\mathbf{x}_i - \mathbf{o}_{yj}\|_1 - \beta\|\mathbf{x}_i - \mathbf{o}_{nj}\|_1$.

Following Peress and Spirling (2010), we call the log-likelihood function is *identifiable* in $\big(\{\mathbf{x}_i\}, \{\mathbf{o}_{yj}\}, \{\mathbf{o}_{nj}\}\big)$ if and only if, for every realizations of $\big\{y_{ij}\big\}$, there does not exist $\big(\{\mathbf{x}_i\}, \{\mathbf{o}_{yj}\}, \{\mathbf{o}_{nj}\}\big)$ and $\big(\{\mathbf{x}'_i\}, \{\mathbf{o}'_{yj}\}, \{\mathbf{o}'_{nj}\}\big)$ such that $\big(\{\mathbf{x}_i\}, \{\mathbf{o}_{yj}\}, \{\mathbf{o}_{nj}\}\big) \neq \big(\{\mathbf{x}'_i\}, \{\mathbf{o}'_{yj}\}, \{\mathbf{o}'_{nj}\}\big)$ but

$$\ell\Big(\big(\{\mathbf{x}_i\}, \{\mathbf{o}_{yj}\}, \{\mathbf{o}_{nj}\}\big), \beta\Big) = \ell\Big(\big(\{\mathbf{x}'_i\}, \{\mathbf{o}'_{yj}\}, \{\mathbf{o}'_{nj}\}\big), \beta'\Big) \quad (3.10)$$

for some β'.

We have the following results on the identifiability of the log-likelihood function in Equation (3.9).

Theorem 1. *The log-likelihood function as a function of* $\big(\{\mathbf{x}_i\}, \{\mathbf{o}_{yj}\}, \{\mathbf{o}_{nj}\}\big)$ *is identifiable up to the class of variables formed by the location-shift and signed permutation transformation of the ideological space.*

Theorem 1 relieves the identifiability problem of a general ideal point model with actor-specific ideal points and bill parameters by introducing the ℓ_1-norm for the ideological space. The invariance of (3.9) (only up) to the signed permutation transforms is induced from the isometry property of the ℓ_1-normed space (only up) to the signed permutation transformation. In geometry, suppose we consider an unit ball in the $s-$dimensional ℓ_1-normed space, it has 2^s extremal points and the 'only' transformation to preserve the distance from zero to these extremal points is the signed permutation transformation. On the other hand, the identifiability on the location shift transformation is obtained by imposing a constraint to the average of all ideal scores equals to the zero vector.

Note that Theorem 1 also provides the identifiability of Yea/Nay positions, using which researchers can test theories of lawmaking (see Peress, 2013 for a relevant discussion). We provide further discussion on the spatial voting theory and the intuition behind the identification of such roll call positions in Appendix F and Appendix G.

Next, we show the contraction of the posterior distribution of $\boldsymbol{\theta} = \big(\{\mathbf{x}_i\}, \{\mathbf{o}_{yj}\}, \{\mathbf{o}_{nj}\}\big)$ to their true values. For this, we prepare two normed spaces of two sets of parameters. First, we consider the parameter space of $\mathbf{Y}$, an (NM)-dimensional space of $\boldsymbol{p} = \big(p_{11}, \dots, p_{1M}, p_{21}, \dots, p_{2M}, \dots, p_{N1}, \dots, p_{NM}\big)^\top$, which is equipped with the vector ℓ_2-norm, $\|\boldsymbol{p}\|^2 = \sum_{ij} p_{ij}^2$. Second, we consider a weighted matrix $\ell_{1,2}$-norm for the $(N+2M) \times s$-dimensional space of the ideological scores $\boldsymbol{\theta}$ in (3.6) that is

$$\big\|\boldsymbol{\theta}\big\|_{\mathrm{m}}^2 := M\left(\sum_i \|\mathbf{x}_i\|_1^2\right) + N\left(\sum_i \|\mathbf{o}_{yj}\|_1^2\right) + N\left(\sum_i \|\mathbf{o}_{nj}\|_1^2\right). \tag{3.11}$$

In (3.11), $\sum_i \|\mathbf{x}_i\|_1^2$, $\sum_i \|\mathbf{o}_{yj}\|_1^2$, and $\sum_i \|\mathbf{o}_{nj}\|_1^2$ are the matrix $\ell_{1,2}-$norm of

$$\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_N \end{pmatrix}, \begin{pmatrix} \mathbf{o}_{y1} \\ \mathbf{o}_{y2} \\ \vdots \\ \mathbf{o}_{yM} \end{pmatrix}, \text{ and } \begin{pmatrix} \mathbf{o}_{n1} \\ \mathbf{o}_{n2} \\ \vdots \\ \mathbf{o}_{nM} \end{pmatrix}.$$

The newly defined norm $\big\|\boldsymbol{\theta}\big\|_{\mathrm{m}}$ is the convex combination of $\sum_i \|\mathbf{x}_i\|_1^2$, $\sum_i \|\mathbf{o}_{yj}\|_1^2$, and $\sum_i \|\mathbf{o}_{nj}\|_1^2$, and thus it is well defined. We show that the ideological space of $\boldsymbol{\theta}$ with the weighted matrix $\ell_{1,2}$-norm is compatible with the $(NN)-$dimensional parameter space of $\boldsymbol{p}$ with the vector ℓ_2-norm.

Now, we state the contraction property of the posterior distribution.

Theorem 2. *Suppose* $\mathbf{p}^* = \{p_{ij}^*, i = 1, 2, \dots, N, j = 1, 2, \dots, M\}$ *satisfies*

$$\max_{ij} \frac{1}{p_{ij}^*(1 - p_{ij}^*)} \le u'. \tag{3.12}$$

Then, for $\delta_{N,M} = (N+2M)^{1+\nu}$ *with any* $\nu > 0$*, we have*

$$\lim_{N,M\to\infty} \Pr\left(\boldsymbol{\theta} \in \overline{\mathcal{B}}(\boldsymbol{\theta}^*, \delta_{N.M}) \middle| \mathbf{Y}\right) = 0, \quad \textit{almost surely in } P_{\mathbf{Y}}, \tag{3.13}$$

where $\overline{\mathcal{B}}(\boldsymbol{\theta}^*, \delta_{N,M}) := \mathcal{B}(\boldsymbol{\theta}^*, \delta_{N.M})^c$,

$$\mathcal{B}(\boldsymbol{\theta}^*, \delta_{N,M}) := \Big\{\boldsymbol{\theta} \in \mathcal{R}^q \;\Big|\; \min_{\mathrm{P}\in\mathcal{P}} \big\|\mathrm{P}(\boldsymbol{\theta}) - \boldsymbol{\theta}^*\big\|_{\mathrm{m}} < \delta_{N,M}\Big\},$$

for $\mathcal{P}$ is the set of all $2^s \cdot s!$ signed permutation matrices and $\mathrm{P}(\boldsymbol{\theta}) := \left(\mathrm{P}\mathbf{x}^\top, \mathrm{P}\mathbf{o}_y^\top, \mathrm{P}\mathbf{o}_n^\top\right)^\top$.

Theorem 2 proves that, under the assumption of a proper choice of the perpendicular rotation P (let $\mathrm{P} = I_s$ without loss of generality), the posterior estimates of BMIM based on ℓ_1-norm distance almost surely contract on a small error around the "true" parameter $\boldsymbol{\theta}$, whose weighted matrix $\ell_{1,2}$ norm in the 'prior' distribution is concentrated around its mean with a high probability as

$$\left\|\boldsymbol{\theta}\right\|_{\mathrm{m}} \in NM\left\{\mu_{1,2} \pm O\left(\frac{1}{\sqrt{\max(N,M)}}\right)\right\} \tag{3.14}$$

with $\mu_{1,2} = \frac{1}{NM}\left\{\mathrm{E}\left\|\mathbf{x}_1\right\|_1^2 + \mathrm{E}\left\|\mathbf{o}_{y1}\right\|_1^2 + \mathrm{E}\left\|\mathbf{o}_{n1}\right\|_1^2\right\}^{1/2}$. In contrast, $\ell_{1,2}$ norm of $\boldsymbol{\theta}$ in the 'posterior' distribution is concentrated in the ball having the center at the true value $\boldsymbol{\theta}^*$ and the radius $(N+2M)^{1+\nu}$. The radius $(N+2M)^{1+\nu}$ is much smaller than the radius $NM/\sqrt{\max(N,M)}$ and thus we have the posterior contraction to the true value $\boldsymbol{\theta}^*$.

The proof of Theorem 2 could be done with two lemmas, Lemma 1 and Lemma 2 in Appendix A. The first lemma shows the compatibility between two norms, the weighted matrix $\ell_{1,2}$-norm of $\boldsymbol{\theta}$ and ℓ_2-norm of $\mathbf{p}$ and this makes the contraction of the posterior probability of $\boldsymbol{\theta}$ in the weighted matrix $\ell_{1,2}$-norm is equivalent to that of the posterior probability of $\mathbf{p}$ in the ℓ_2-norm. The normed space of $\mathbf{p}$ with the ℓ_2-norm is the parameter space of $\mathbf{Y}$ and we obtain the posterior contraction of $\mathbf{p}$ around the true value $\mathbf{p}^*$ by following regular steps. The detailed proofs of Theorem 1 and 2 are provided in Appendix A and Appendix B.

3.5 Estimation

We can write the joint posterior distribution of the proposed model as follows:

$$p(\cdot|\text{data}) \propto \prod_{i=1}^{N}\prod_{j=1}^{M}\left\{\Phi(u_{ijy}-u_{ijn})^{y_{ij}} \times (1-\Phi(u_{ijy}-u_{ijn}))^{1-y_{ij}}\right\} \tag{3.15}$$

$$\times \prod_{i=1}^{N} p(\mathbf{x}_i) \times \prod_{j=1}^{M}\left\{p(\mathbf{o}_{yj}) \times p(\mathbf{o}_{nj})\right\}. \tag{3.16}$$

Theorem 1 tells the likelihood is invariant to the location-shift and perpendicular rotation transformations. We often resolve the invariance of the location-shift transformation and its difficulty by imposing a set of s linear constraints to $\boldsymbol{\theta} = (\mathbf{x}^\top, \mathbf{o}_y^\top, \mathbf{o}_n^\top)^\top$. Here, we choose $\sum_i \mathbf{x}_i = \mathbf{0}_s$ following among many sets of constraints. Note that the likelihood's invariance to the signed permutation transformation does not affect the relative configuration or substantive interpretation of the ideal points. This can be addressed with a simple post-processing step described in Appendix J.

Since the full conditional distributions of ideal point parameters do not belong to any standard distributions, we use the multivariate slice sampler (MSS) proposed by Neal (2003) for the estimation of ideal point parameters. Slice sampling is a generic method that can be easily implemented in case where the sampling distributions do not have a standard form. It has been used in the implementation of a variant of the NOMINATE method Carroll et al. (2013), where Gibbs sampling is not applicable as in BIRT. MSS is a straightforward generalization of a single-variable slice sampler that updates multiple variables at the same iteration. We choose MSS using hyper-rectangle to update multidimensional parameters simultaneously so that the sampler produces an ergodic chain despite the tight dependencies of parameters across dimensions. Specifically, in our estimation using MSS, a multidimensional ideal point of a single legislator $(x_{i1}, \ldots, x_{is})$ is sampled together for each i, and similarly for Yea $(o_{jy1}, \ldots, o_{jys})$ and Nay position $(o_{jn1}, \ldots, o_{jns})$.

A generic description of MSS we use is as follows. For simplicity of notation, let $\lambda = (\lambda_1, \cdots, \lambda_s)$ denote a parameter to be updated. At the t-th iteration, the method replaces the current state, $\lambda^{(t)}$, with a new state, $\lambda^{(t+1)}$, following the three-steps (Neal, 2003, 721):

1. Draw l from $\mathcal{U}(0, f(\lambda^{(t)}))$, thereby defining a horizontal "slice": $S = \{\lambda : l < f(\lambda)\}$.

2. Find a hyper-rectangle $H = (L_1, R_1) \times \cdots \times (L_s, R_s)$ around $\lambda^{(t)}$ that contains all, or much, of the slice.

3. Draw the new point, $\lambda^{(t+1)} \in A = \{\lambda : \lambda \in S \cap H \text{ and } \Pr(\text{Select } H \mid \lambda) = \Pr(\text{Select } H \mid \lambda^{(t)})\}$.

For example, to sample a 2-dimensional parameter from the full conditional distribution, a horizontal slice of the density is drawn uniformly in the first step. Then, we find a rectangle that contains all of the slice we made. Lastly, we draw the new parameter by picking uniformly from the rectangle until a point inside the slice is found. The MSS algorithm for BMIM estimation is described in Algorithm 1. We choose the tuning parameter c to be 4 throughout the simulation study and real data analysis (see Appendix I for more details).

Algorithm 1: t-th Iteration of MSS for BMIM Estimation

input : $f(\lambda) =$ the full conditional distribution of λ

$\lambda^{(t)} =$ the current point

$c =$ estimate of the typical size of a slice

output: $\lambda^{(t+1)}$ = the new point

initialization: $l \leftarrow f(\lambda^{(t)}) - \text{rexp}(1)$

For $i = 1, \cdots, s$:

$L_i \leftarrow \lambda_i^{(t)} - c \times \text{runif}(0, 1)$

$R_i \leftarrow L_i + c$

repeat

  for $i = 1, \cdots, s$ **do**

    $\lambda_i^{(t+1)} \leftarrow L_i + \text{runif}(0, 1) \times (R_i - L_i)$

  if $l < f(\lambda^{(t+1)})$ **then**

    exit loop

  for $i = 1, \cdots, s$ **do**

    if $\lambda_i^{(t+1)} < \lambda_i^{(t)}$ **then**

      $L_i \leftarrow \lambda_i^{(t+1)}$

    else

      $R_i \leftarrow \lambda_i^{(t+1)}$

In case of 2-dimensional latent space, the sampling algorithm of the proposed method using the MSS can be summarized as follows:

1. Sample yea position of j-th bill ($\mathbf{o}_{yj}$) from $f(\mathbf{o}_{yj} \mid \{\mathbf{o}_{yj'}\}_{j' \neq j}, \{\mathbf{o}_{nj}\}, \{\mathbf{x}_i\})$ for $j \in \{1, \ldots, M\}$ using MSS.

2. Sample nay position of j-th bill ($\mathbf{o}_{nj}$) from $f(\mathbf{o}_{nj} \mid \{\mathbf{o}_{nj'}\}_{j' \neq j}, \{\mathbf{o}_{yj}\}, \{\mathbf{x}_i\})$ for $j \in \{1, \ldots, M\}$ using MSS.

3. Sample ideal point of i-th legislator ($\mathbf{x}_i$) from $f(\mathbf{x}_i \mid \{\mathbf{x}_{i'}\}_{i' \neq i}, \{\mathbf{o}_{yj}\}, \{\mathbf{o}_{nj}\})$ for $i \in \{1, \ldots, N\}$ using MSS.

4 Simulation Study

We consider three simulation settings, each of which represents three distinct cases of multidimensional politics. The first case is a non-partisan system where legislators' ideal points are randomly distributed across the latent space, independent of legislators' party affiliations. The second case is a two-party system where legislators' multidimensional ideal points are different in both dimensions along the party line. The last case is a multi-party system in which ideal points are concentrated in four different-sized clusters. Simulated examples of three cases are visualized in the first column of Figure 6.

For simulation, we assume a two-dimensional latent space with normally distributed ideal points and uniformly distributed Yea/Nay positions.

1. Draw ideal points $\mathbf{x}_{g[i]}$ from $\mathcal{N}(\boldsymbol{\mu}_{g[i]}, \Sigma_{g[i]})$, where $g[i] \in \{1, 2, \cdots, \#g\}$ denotes a cluster to which the legislator i belongs and $\Sigma_{g[i]} = \begin{pmatrix} \sigma^2_{g[i]} & 0 \\ 0 & \sigma^2_{g[i]} \end{pmatrix}$.
2. Draw each-dimensional coordinates of Yea and Nay positions $\mathbf{o}_{yjk}, \mathbf{o}_{njk}$ from $\mathcal{U}(-1, 1)$.
3. Compute $\Pr(y_{ij} = 1)$ for each (i, j) with randomly sampled $\epsilon_{ijy} - \epsilon_{ijn}$ from $\mathcal{N}(0, 0.5)$.
4. Sample γ_{ij} from $\mathcal{U}(0, 1)$. If $\gamma_{ij} < \Pr(y_{ij} = 1)$ set $y_{ij} = 1$, otherwise set $y_{ij} = 0$.

In words, we first generate synthetic ideal points and Yea/Nay positions. Next, we compute each legislator's probability of voting Yea for each roll call vote. Then, we generate synthetic roll call votes data using the synthetic ideal points, Yea/Nay positions, and voting probabilities. We set $N = 100$ with $M = 1000$. All samples are obtained from 4 chains of $50,000$ MCMC iterations with $20,000$ burn-in trials. Every 10th draw is stored for analysis. See Appendix H for a summary of computational cost and convergence diagnostics.

The results of the simulation are visualized in Figure 6. Note that the results shown in the figure are after adjusting the signed permutation of the estimated ideal points for ease of comparison. Three different cases, (a) the non-partisan system, (b) the two-party system, and (c) the multi-party system, are presented row-wise. The first column shows the ground truth and the second column shows the posterior means of multidimensional ideal point estimates from BMIM. The third and fourth columns compare the posterior means of BMIM with true ideal points for each coordinate.

The results clearly demonstrate that BMIM successfully recovers multidimensional ideal points across three different settings. The correlations between posterior means and the true ideal points in the third and fourth columns indicate that the rank order of posterior estimates for each coordinate closely matches that of the true ideal points. Specifically, the second row, the case of a two-party system, shows that the proposed method successfully recovers dimension-specific estimates of correlated coordinates, which is not feasible when

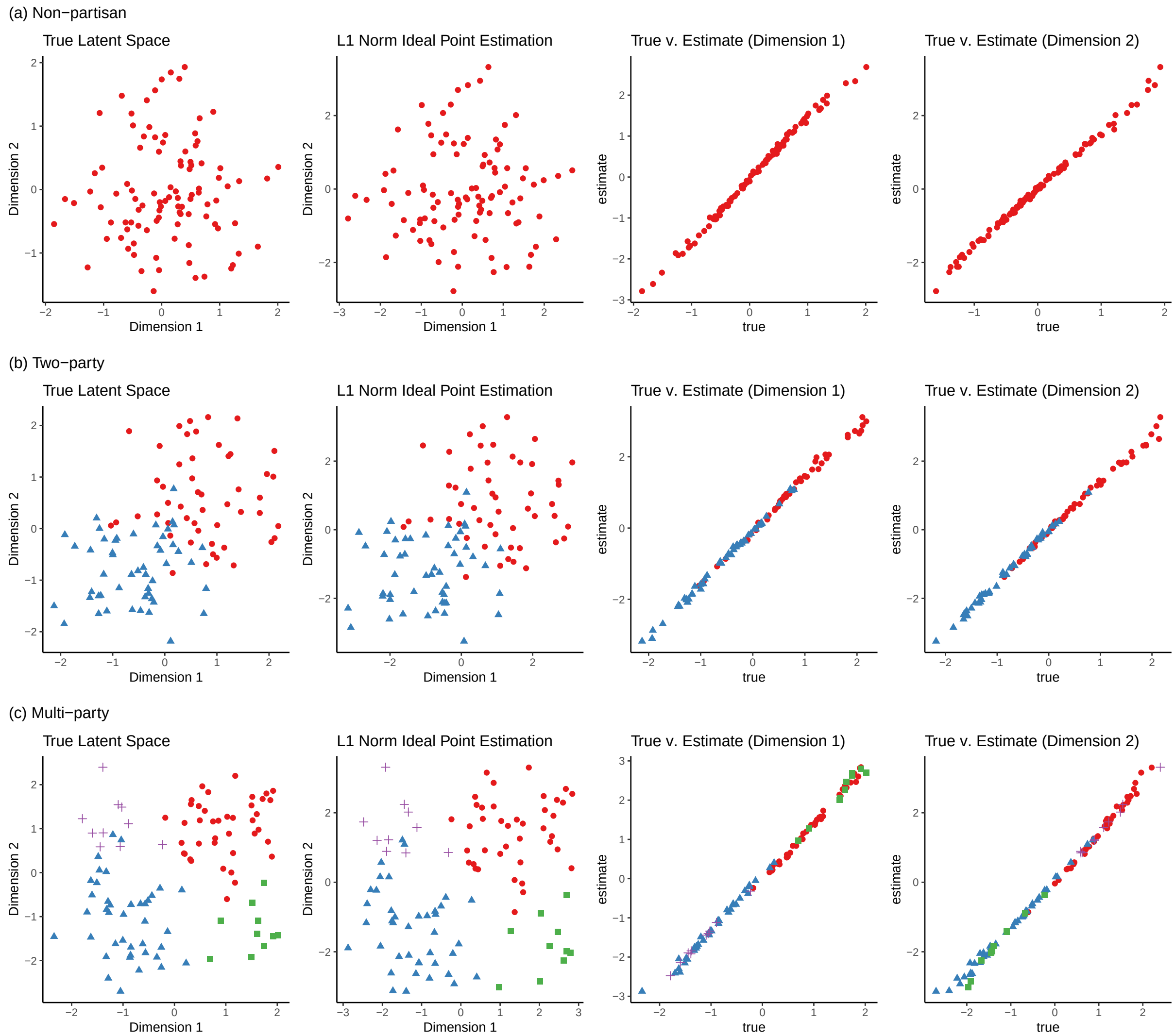

Figure 6: Simulation Studies with Synthetic Data: Clusters are identified by different colors and shapes. Each row indicates different cluster structures and the first column shows the ground truth. The second to fourth columns compare BMIM estimates with the ground truth. Specifically, the second column shows estimated ideal points, and the third and fourth columns dimension-wise comparison results. The first row (a) shows the case of the non-partisan system where synthetic legislators are dispersed across the space. The second row (b) shows the case of a two-party system where two parties oppose each other across the dimensions. The third row (c) shows the case of the multi-party system with two opposing main parties and two minor parties.

ℓ_2 distance is used in the likelihood. The third row also shows that the proposed method successfully recovers a complex cleavage structure with multiple voting clusters. Two small clusters in the off-diagonal direction (purple crosses and blue squares) are not lumped together with two larger clusters (red circles and green triangles) even though we do not impose any a priori identifying constraint. Comparison with BIRT and WNOMINATE is available in Appendix K. Additionally, we conduct further simulation studies of misspecified cases where the roll call data is generated based on the Gaussian/quadratic utility model and the non-uniformly distributed Yea/Nay positions. The results are available in Appendix L and Appendix M.

5 Applications

In this section, we analyze roll call data of the 53rd US House of Representatives during the late Gilded Age. As discussed in Section 2, the congressional and electoral politics during this period were largely shaped by the battle over the gold standard. The division over gold aligned with a sectional division rather than a partisan division, forming a multidimensional policy space among representatives. We use the roll call votes of the 53rd US House of Representatives collected by `voteview` (Lewis et al., 2022). The roll call data consists of 372 representatives with 336 roll calls. All samples are obtained from $100,000$ MCMC iterations with $50,000$ burn-in and thinning every 10th draw.

Figure 7 compares BMIM estimates (a) with those of DW-NOMINATE (b). We denote legislators' states by shape and voting clusters by color. The results of our method reveal two notable patterns that are missing in that of DW-NOMINATE. First, one of the major cleavages during this time, the battle of the standards, is captured in the second dimension by our method. In panel (a), Southern Democrats (orange triangles) and the Populists from the West (green rectangles) formed an anti-gold alliance with the Silver representative (gray rectangle) which is located at the top of the figure. This is in contrast with the results of DW-NOMINATE shown in panel (b) where Southern Democrats and the Populists are not clustered on either side of the second dimension. Additional analysis focusing on two core Silver bills: Sherman Act Repeal (H.R.1) and Free Silver Override (H.R.4956) (Frieden, 2016, p.129), is presented in Appendix O.

Second, the first dimension of the panel (a) follows the partisan division, indicating that the partisan division is still a significant cleavage among the representatives. Panel (b) shows that the politics of the 53rd Congress was mainly about partisan division, with no clear pattern in the second dimension.

To further examine the trajectory of the battle of standards, we extended the time frame to four Congresses from 1891 to 1899 and conducted BMIM analysis. The results presented in Figure 8 clearly indicate that the sectional division on the monetary standard

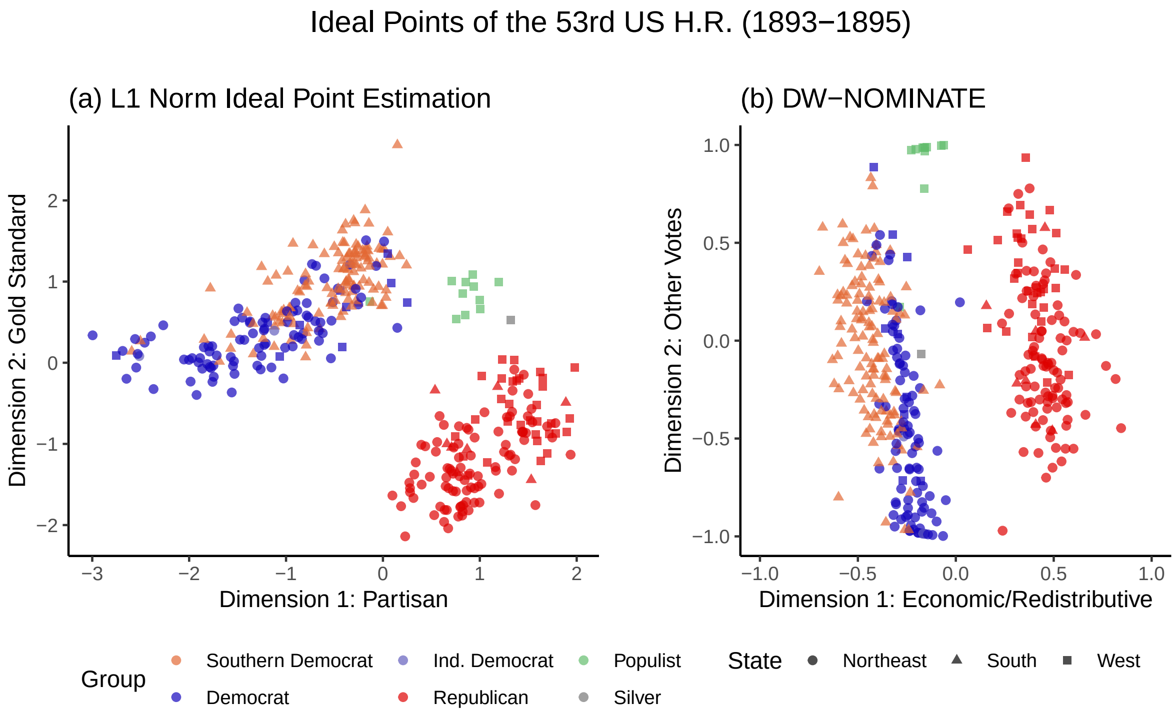

Figure 7: Multidimensional Ideal Point Estimation of the 53rd US House of Representatives (1893-1896). Each point denotes a two-dimensional ideal point of representative, measured by (a) BMIM, and (b) DW-NOMINATE method respectively. The color indicates the party label of the representative: blue points indicate Democrats, red points Republicans, green points Populists, and gray points the Silver party. The shape indicates the state which each representative represents: circles indicate Northeastern states, triangles Southern states, and rectangles Western states. For ease of visualization, we used orange triangles to denote Southern Democrats.

was eventually integrated with the partisan divide. This change in the battle of standards occurred before and after the presidential election of 1896, in which Republican William McKinley defeated Democratic-Populist William Jennings Bryan. The second sectional cleavage of the House, distinct from the first partisan cleavage, was shaped by the battle of standards during the 52nd and 53rd Congresses, as shown in panels (a) and (b) of Figure 8. However, the gold standard issue was absorbed by the first partisan dimension in panel (c), which separated Goldbug Republicans on the right from the majority of southern Democrats, Populists, and some Western Republicans on the left. In Appendix O and Appendix P, we further illustrate a historical background of the transition and provide additional analysis fitting the entire time frame.

Estimates of BIRT are largely consistent with estimates of WNOMINATE, both of which are available in Appendix O. Ideal points are split mainly by the partisan division (the first dimension), but the second dimension does not capture the sectional division on monetary issue. Southern Democrats and populists are dispersed over the far-left region without forming a distinct group in the estimates of WNOMINATE and BIRT. For this reason, we conclude that the proposed method has a comparative advantage in representing the multidimensional politics during this period over two existing methods of multidimensional ideal point estimation.

6 Discussion

In this article, we propose a novel method for the multidimensional ideal point estimation using the ℓ_1 distance and Bayesian inference. We show that the use of ℓ_1 distance transforms the invariance problem of infinite rotational turns, which is the main challenge to the identification and interpretation of multidimensional ideal points, into a milder problem of the signed perpendicular problem. We showed that the total number of likelihood-preserving rotations and reflections is reduced from ∞ to eight in the case of the two-dimensional voting space as we change the distance measure from ℓ_2 to ℓ_1, where these remaining signed perpendicular problem does not alter the relative configuration and substantive interpretation of the ideal points.

Our simulation studies show that the proposed method successfully recovers various types of multidimensional ideal points while existing ideal point estimation methods fail to recover the planted structures of multidimensional ideal points. The proposed method turned out to be highly effective in identifying complex congressional cleavages during the late Gilded Age (1891-1899), an era of the battle of the standards. Our method finds that the sectional division – Southern Democrats and the Populists from the West vs. Northeastern representatives – arose as a major cleavage, along with partisan division, during the 52nd and 53rd Congresses, which is consistent with existing historical narratives on this period (Bensel, 2000; Frieden, 2016).

A principled statistical method to identify multidimensional politics is important both for methodological and substantive reasons. First, methodologically speaking, attempts to uncover multidimensional ideal points using a priori constraints hinder scientific communication. Estimates of multidimensional ideal points will be conditional upon a priori constraint, the uncertainty of which is not accounted for in the estimation. Thus, it is cru-

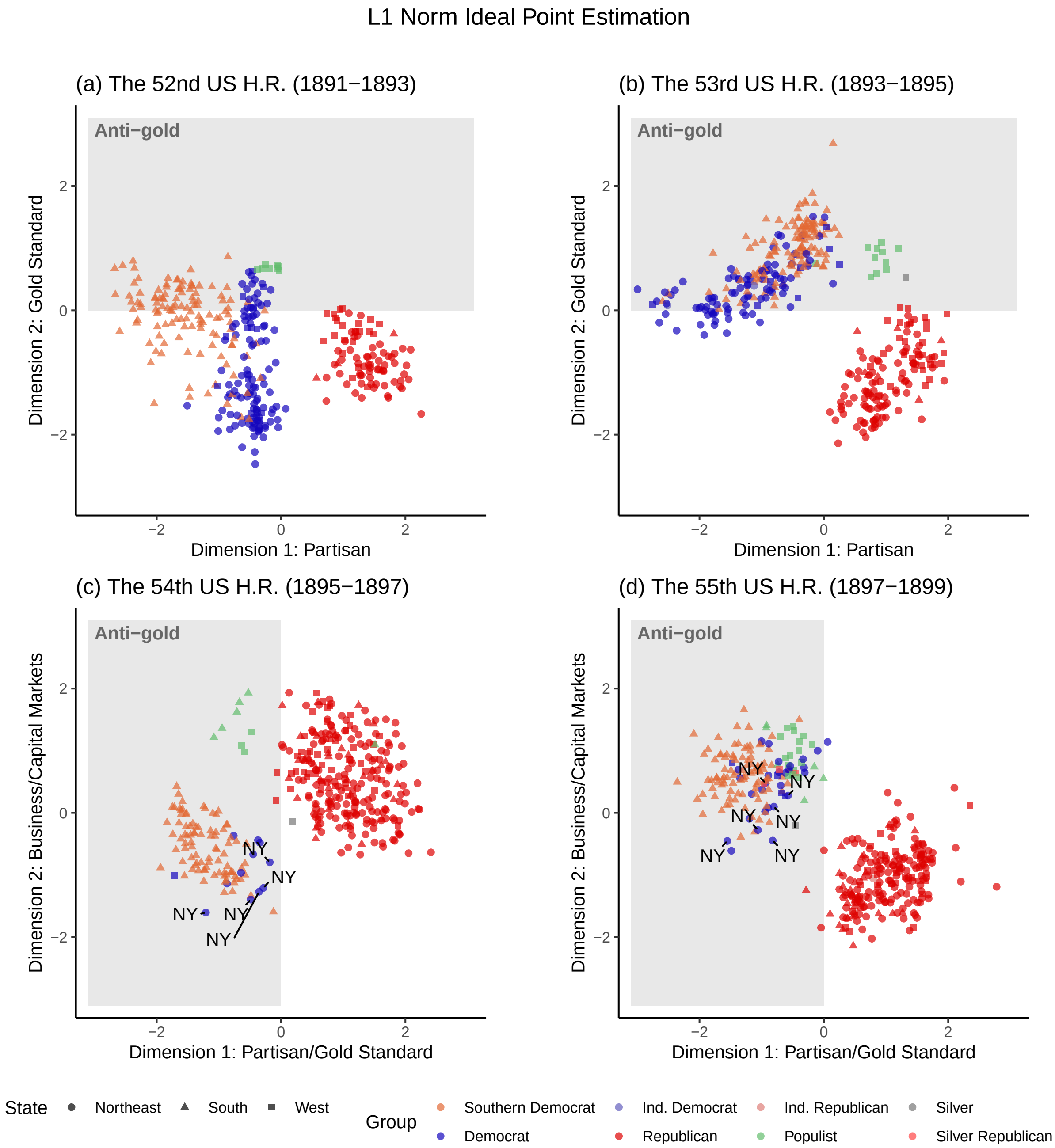

Figure 8: Multidimensional Ideal Point Estimation of US House of Representatives using the Proposed Method. (a) The 52nd US H.R. (1891-1893). (b) The 53rd US H.R. (1893-1895). (c) The 54th US H.R. (1895-1897). (d) The 55th US H.R. (1897-1899). The color indicates the party label of the representative: blue points indicate Democrats, red points Republican, green points Populists, and gray points Silver party. The shape indicates the state which each representative represents: circles indicate Northeastern states, triangles Southern states, and rectangles Western states. For ease of visualization, we used orange triangles to denote Southern Democrats.

cial to have a valid statistical method that minimizes the impact of researchers' subjective judgment on estimates. A false description of politics is made possible by the absence of a systematic approach to multidimensional ideal points, which is the second and equally important issue. Due to the rotational invariance problem and researcher-specified constraints, conventional ideal point estimation methods may overstate the significance of the first dimension, which is typically the partisan divide, identify political cleavages incorrectly, or underestimate the likelihood of a political compromise. Our method shed light on an alternative solution for the rotational invariance problem, thus providing a principled way of estimating multidimensional ideal points.

Supplementary Materials

The supplementary materials include (1) a pdf file containing the appendix referenced in the paper and (2) replication codes, along with detailed instructions for reproducing the results.

Acknowledgments

We are grateful to the Editor, Associate Editor, and two anonymous reviewers for their constructive comments and suggestions.

Disclosure Statement

The authors report there are no competing interests to declare.

References

Armstrong, D. A., Bakker, R., Carroll, R., Hare, C., Poole, K. T., and Rosenthal, H. (2020). *Analyzing Spatial Models of Choice and Judgment*. Chapman and Hall/CRC, 2nd edition.

Bafumi, J., Gelman, A., Park, D. K., and Kaplan, N. (2005). Practical issues in implementing and understanding bayesian ideal point estimation. *Political Analysis*, 13(2):171–187.

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

Bailey, M. and Chang, K. H. (2001). Comparing presidents, senators, and justices: interinstitutional preference estimation. *Journal of Law, Economics, and Organization*, 17(2):477–506.

Barberá, P. (2015). Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Political Analysis*, 23(1):76–91.

Bateman, D. A. and Lapinski, J. (2016). Ideal points and american political development: Beyond dw-nominate. *Studies in American Political Development*, 30(2):147–171.

Benoit, K. and Laver, M. (2012). The dimensionality of political space: Epistemological and methodological considerations. *European Union Politics*, 13(2):194–218.

Bensel, R. (2016). Lost in translation: an epistemological exploration of the relation between historical analysis and the nominate algorithm. *Studies in American Political Development*, 30(2):185–201.

Bensel, R. F. (2000). *The Political Economy of American Industrialization, 1877–1900*. Cambridge University Press.

Binding, G. and Stoetzer, L. F. (2022). Non-separable preferences in the statistical analysis of roll call votes. *Political Analysis*, Forthcoming:1–14.

Bond, R. and Messing, S. (2015). Quantifying social media's political space: Estimating ideology from publicly revealed preferences on facebook. *American Political Science Review*, 109(1):62–78.

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

Carroll, R., Lewis, J. B., Lo, J., Poole, K. T., and Rosenthal, H. (2009). Comparing nominate and ideal: Points of difference and monte carlo tests. *Legislative Studies Quarterly*, 34(4):555–591.

Carroll, R., Lewis, J. B., Lo, J., Poole, K. T., and Rosenthal, H. (2013). The structure of utility in spatial models of voting. *American Journal of Political Science*, 57(4):1008–1028.

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

Frieden, J. A. (2016). *Currency Politics: The Political Economy of Exchange Rate policy*. Princeton University Press.

Gerrish, S. M. and Blei, D. M. (2011). Predicting legislative roll calls from text. In *Proceedings of the 28th International Conference on Machine Learning, ICML 2011*.

Goplerud, M. (2019). A multinomial framework for ideal point estimation. *Political Analysis*, 27(1):69–89.

Hix, S., Noury, A., and Roland, G. (2006). Dimensions of politics in the european parliament. *American Journal of Political Science*, 50(2):494–520.

Hoff, P. D. (2007). Model averaging and dimension selection for the singular value decomposition. *Journal of the American Statistical Association*, 102(478):674–685.

Hoff, P. 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.

Imai, K., Lo, J., and Olmsted, J. (2016). Fast estimation of ideal points with massive data. *American Political Science Review*, 110(4):631–656.

Jackman, S. (2020). *pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory*. United States Studies Centre, University of Sydney, Sydney, New South Wales, Australia. R package version 1.5.5.

Lewis, J. B., Poole, K., Rosenthal, H., Boche, A., Rudkin, A., and Sonnet, L. (2022). Voteview: Congressional roll-call votes database. Technical report.

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

Moser, S., Rodríguez, A., and Lofland, C. L. (2021). Multiple ideal points: Revealed preferences in different domains. *Political Analysis*, 29(2):139–166.

Neal, R. M. (2003). Slice sampling. *The Annals of Statistics*, 31(3):705–767.

Peress, M. (2013). Estimating proposal and status quo locations using voting and cosponsorship data. *The Journal of Politics*, 75(3):613–631.

Peress, M. (2020). Large-scale ideal point estimation. *Political Analysis*, 30(3):346–363.

Peress, M. and Spirling, A. (2010). Scaling the critics: Uncovering the latent dimensions of movie criticism with an item response approach. *Journal of the American Statistical Association*, 105(489):71–83.

Poole, K., Lewis, J., Lo, J., and Carroll, R. (2011). Scaling roll call votes with wnominate in R. *Journal of Statistical Software*, 42(14):1–21.

Poole, K. T. and Rosenthal, H. (1997). *Congress: A Political-Economic History of Roll-Call Voting*. Oxford University Press, Oxford.

Shor, B. and McCarty, N. (2011). The ideological mapping of american legislatures. 105(3):530–551.

Slapin, J. B. and Proksch, S.-O. (2008). A scaling model for estimating time-series party positions from texts. *American Journal of Political Science*, 52(3):705–722.

Sohn, Y. (2017). *Measuring Ideology, Dimensionality and Polarization in Politics.* Ph.D. Disseration, Department of Political Science, University of California, San Diego.

Tausanovitch, C. and Warshaw, C. (2013). Measuring constituent policy preferences in congress, state legislatures, and cities. *The Journal of Politics*, 75(2):330–342.

Vafa, K., Naidu, S., and Blei, D. M. (2020). Text-based ideal points. *Proceedings of the 2020 Conference of the Association for Computational Linguistics, ACL 2020.*

Voeten, E. (2000). Clashes in the assembly. *International Organization*, 54(2):185–215.