Scaling Data from Multiple Sources: Analyzing FOMC Deliberations

Ted Enamorado[*] Gabriel López-Moctezuma[†] Marc Ratkovic[‡]

April 1, 2018

Abstract

We introduce a method for scaling two data sets from different sources. The proposed method estimates a latent component common to both datasets as well as an idiosyncratic component unique to each. In addition, it offers a flexible modeling strategy which permits the scaled locations to be a function of covariates, and efficient implementation allows for inference through resampling. A simulation study shows that our proposed method improves over existing alternatives that aim to use information from multiple sources to recover latent dimensions. In the main empirical application, we employ our method to recover a latent scaling of Federal Open Market Committee (FOMC) members around the financial crisis of 2008 combining both their speech and policy recommendations obtained from monetary policy meetings. We provide two additional analysis to empirically validate the proposed method: we use votes and speech from the 112th Senate to provide senators' locations in the shared space of votes and words. We also combine data on cross-country political and socioeconomic characteristics and scale countries according to both attributes.

[*]Ph.D. Candidate, Department of Politics, Princeton University, Princeton NJ 08544. Email: tede@princeton.edu

[†]Assistant Professor, Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena CA 91125. Email: glmoctezuma@caltech.edu

[‡]Assistant Professor, Department of Politics, Princeton University, Princeton NJ 08544. Phone: 608-658-9665, Email: ratkovic@princeton.edu, URL: http://www.princeton.edu/∼ratkovic

1 Introduction

Increasingly, political scientists confront not just large amounts of data but different *types* of data. As examples, political actors will often generate text data and vote data (e.g. Lauderdale and Clark, 2014); countries may have sets of qualitatively distinct attributes, such as political, social, and economic indicators (e.g. Coppedge et al., 2015); the same survey questions may be given to different groups of actors (e.g. Shor and McCarty, 2011); campaign contributions may flow from the same actors to both state and federal candidates (Bonica, 2014). In each case, the researcher must analyze data on different attributes for the same actors (say, Tweets and votes from legislators, Barbera (2016)), or the same attributes but on different actors (say, surveys given to both legislators and the mass public, Bafumi and Herron (2010)).

Combining data from different sources creates subtle theoretical and empirical issues. Jessee (2016) illustrated the underlying problem rather elegantly. Using survey data in the CCES with bridge questions for legislators, he showed that scaled locations can vary if different numbers of respondents from the two samples are pooled and used to estimate ideal points. The problem arises because different groups give different weights to each question, and generalizes to the problem of how to weight data coming from two different sources. Existing works have addressed, but not quite solved, the issue. Kim, Londregan and Ratkovic (2018) develop a choice theoretic model for combining words and votes, but a tuning parameter that balances the proportion of information coming from each source is not estimated within the model. Hobbs (2017) combines information from multiple text sources using a version of canonical correlation analysis (e.g. Hastie, Tibshirani and Friedman., 2013, Sec. 3.7), a method closely related to ours. The method, though, is tailored to short bursts of speech and does not offer means of inference. Earlier work pool multiple types of responses in a single scaling model (Quinn, 2004; Hoff, 2007; Jackman and Trier, 2008; Murray et al., 2013), but they all produce a pooled estimate that does not differentiate between qualitatively distinct types of inputs.

In this work, we develop a general framework for combining data from multiple sources. The method, *Multi-Dataset Multidimensional Scaling* (MD2S) simultaneously scales two datasets, decomposing the data into three separate scales: one spanning a latent space common to both datasets, and one more scale idiosyncratic to each of the two datasets. For example, combining votes and words on the same actors, MD2S estimates three latent scales. The first is a joint scale informed by both words and votes. The second is informed by words, but contains no information from votes. Likewise, the third is informed by votes, but not words.

We build off work in statistics and education focusing on recovering factors across multiple surveys or exams (Tucker, 1958; Browne, 1979; Anderson, 1989; Klami, Virtanen and Kaski, 2013; Bach and Jordan, 2005; Gupta et al., 2011; Tipping and Bishop, 1999). This model "inter-battery factor analysis," is precisely the model described above. Given two datasets, it returns a factor common to both, and two more factors, one unique to each dataset and uninformed by the other.

Our advances are threefold. First, unlike these earlier methods, we allow for inference on the number of latent dimensions. Distinguishing a dimension that is signal from one that is noise is a perennial problem, often unaddressed, in the scaling literature. We implement a permutation test that allows the researcher to distinguish a given dimension from noise. Second, we allow the researcher to model scaled locations as a function of covariates. This allows the researcher to conduct inference on whether or how scaled locations vary with some covariates of interest. Our third advance is in terms of estimation. Building on insights first advanced in Aldrich and McKelvey (1977), we implement an efficient estimation routine that performs well when the number of attributes grows large, as with text data where the researcher has a term document matrix with counts on thousands of n-grams for each speaker.

We illustrate the method's use and efficacy through a simulation exercise and three empirical applications. We show in the simulation study that MD2S outperforms existing methods in recovering latent dimensions (e.g. Klami, Virtanen and Kaski, 2013), especially as the number of attributes grows, as with text data. We then apply it to the rollcall votes and floor speech in the US Senate, where our shared dimensions aligns with the standard first ideological dimension running from liberals to conservatives (e.g. Poole, 2005). We also use MD2S to combine political and socioeconomic indicators, where we show that the widely used POLITY-IV measure of Gurr, Marshall and Jaggers (2010) is a cross-cutting measure spanning both dimensions.

In our primary applied example, we combine speech and vote data from the Federal Open Market Committee, the body of the US Federal Reserve tasked with setting monetary policy, during the period 2006-2010. MD2S returns a latent scaling combining words and votes that ranks members from inflation hawks to doves. Through a cross-validation exercise, we show that this scaling is a better predictor of observed policy choices than those estimates recovered exclusively from vote data. In addition, we recover the words that are most correlated with each extreme of the latent scale. Finally, we characterize changes in our latent scale before and after the 2008 financial crisis.

2 The Proposed Method

MD2S scales data coming from two different sources. To illustrate the basic problem and insight, consider the case where we observe two different streams of data, words and votes, that are observed on the same actors. Given a rollcall matrix and term-document matrix, we can combine the data in a single matrix.

$$\left[Votes \mid Words \right] \tag{1}$$

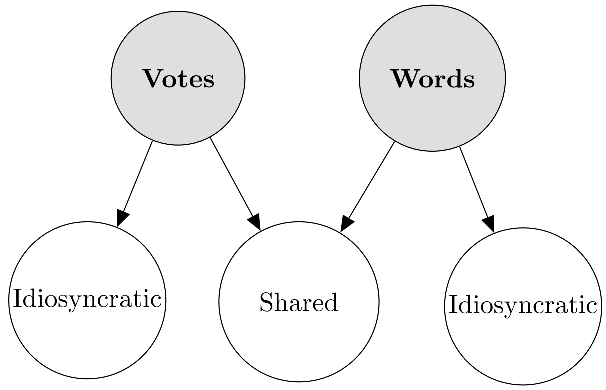
Consider three sets of principal components scalings: a scaling from the full matrix, one from the votes matrix alone, and one from the words matrix alone. In our primary empirical application (the FOMC), for example, there is a low correlation between the votes-only and words-only locations, suggesting the two are spanning different latent spaces. In addition, as is common in text data, we have many more words than votes. When we put the two datasets together, the joint scaling is closer to the words-only scaling than the votes-only scaling. The words contain more information, but we are not interested in all of the word data. We are most interested in the word data that contribute to explaining the joint variation in both types of data.

We could conduct multiple analyses after reweighting the matrix, scaling

$$\left[\mathit{\Large Votes} \mid Words \right] \text{ or } \left[Votes \mid \mathit{\Large Words}, \right] \tag{2}$$

as in Kim, Londregan and Ratkovic (2018), but this sidesteps the problem of relative weighting rather than solving it.

MD2S solves this problem through returning three scalings from the two datasets. The first is a joint scale, estimated to explain the largest amount of variance common to both datasets. The next two are idiosyncratic scalings, common only to each data source and uncorrelated with the common scaling:



These three scalings are designed such that they will not shift, except to grow more precise, with the addition of more information. This model is the Inter-Battery Factor Analysis model of Tucker (1958). MD2S builds on this model in several regards: modeling the scaled locations with covariates, estimating the number of latent dimensions, and providing an efficient estimation algorithm for large number of attributes.

Use Cases and Scope. MD2S generates a more reliable scale than methods that are sensitive to the relative size of disparate inputs, and it is not limited to combining words and votes. We give two examples. First, Jessee (2016) highlighted the problem of how to weight data from different sources when bridging across different sets of respondents (e.g. Tausanovitch and Warshaw, 2013; Shor and McCarty, 2011; Bafumi and Herron, 2010). If the two groups have different item discrimination parameters, simply pooling the two sets generates an ambiguity in their ideal point estimates. The estimates will vary based off the amount of information, and even number of respondents, in the two sets. As another instance, recent work has collected an impressive set of scales and measures for cross-national political, civic, and institutional comparison (Coppedge et al., 2015). Generating an index by aggregating from finer to coarser measures requires a method that is not sensitive to the number of items at each level.

2.1 The Setup

For each observation, we observe two vectors of outcomes, $Y^*_{(1)i}$ and $Y^*_{(2)i}$ with $i \in \{1, 2, \ldots, N\}$. To ease notation, we will use the index m to denote either 1 or 2, so $Y^*_{(m)i}$ is the generic notation for either $Y^*_{(1)i}$ or $Y^*_{(2)i}$. We assume $Y^*_{(m)i}$ is of length K_m, where K_1 may not equal K_2. For example, $Y^*_{(1)i}$ may be a vector of K_1 word counts uttered by legislator i, and $Y^*_{(2)i}$ may be a set of K_2 observed roll call votes for the same legislator.

We denote the outcome data matrices as:

$$Y^*_{(1)} = \begin{bmatrix} Y^{*\top}_{(1)1} \\ Y^{*\top}_{(1)2} \\ \cdots \\ Y^{*\top}_{(1)N} \end{bmatrix}; \;\; Y^*_{(2)} = \begin{bmatrix} Y^{*\top}_{(2)1} \\ Y^{*\top}_{(2)2} \\ \cdots \\ Y^{*\top}_{(2)N} \end{bmatrix}. \tag{3}$$

with individual outcomes in rows. We assume that each matrix $Y^*_{(m)}$ is on a common scale. This may be due to a natural scale, such as binary vote data, columns may be normed to have sample standard deviation one, or some other method may be used to place all columns of $Y^*_{(m)}$ on a common scale (e.g., Quinn, 2004; Hoff, 2007; Murray et al., 2013). The important point for our method is that all columns of $Y^*_{(m)}$ be on a common interval scale. While each matrix must be on a common scale, the two separate matrices may be on different scales. For example, $Y_{(1)}$ may contain roll call votes and $Y_{(2)}$ word counts.

As the intercept is rarely of interest, we pre-process the matrices by double-centering them, so that the row-mean, column-mean, and grand mean is zero. We denote the double-centered matrices as $Y_{(m)}$, with generic row $Y_{(m)i}$ and generic column $Y_{(m)k}$.[1]

[1]The maximum likelihood estimates for the intercept terms are simply the sample analogs (Tipping and Bishop, 1999), i.e. the double-centering matrix with each element the row-mean plus the column-mean, less the grand mean.

Also observed for each individual is X_i, a vector of K_X covariates. Denote as

$$X = \begin{bmatrix} X_1^\top \\ X_2^\top \\ \dots \\ X_N^\top \end{bmatrix} \tag{4}$$

the matrix of covariates. These covariates will be used to inform the scaled locations. For example, we will explore how committee members' scale location varies by political principal.

2.2 The Model

We model $Y_{(1)}$ and $Y_{(2)}$ in terms of their latent components as[2]

$$Y_{(1)} = Z_S L_{(1)} W_{(1)}^\top + Z_{(1)} D_{(1)} B_{(1)}^\top + \Omega_{(1)} \tag{5}$$

$$Y_{(2)} = Z_S L_{(2)} W_{(2)}^\top + Z_{(2)} D_{(2)} B_{(2)}^\top + \Omega_{(2)}. \tag{6}$$

We will refer to the $N \times Q_S$ matrix Z_S as the shared subspace and the $N \times Q_{(m)}$ matrix $Z_{(m)}$ as the idiosyncratic subspace. Z_S contains latent scaled locations in the shared subspace in columns for each of the Q_S dimensions. Similarly, each column of $Z_{(m)}$ contains the latent scaled locations in the idiosyncratic subspace for $Q_{(m)}$ latent dimensions. $L_{(m)}$ is a $Q_S \times Q_S$ nonnegative, diagonal matrix of loadings for the shared subspace for $Y_{(m)}$, and $W_{(m)}$ is an $K_{(m)} \times Q_S$ matrix of components for the shared subspace for dataset $Y_{(m)}$. $D_{(m)}$ is a $Q_{(m)} \times Q_{(m)}$ diagonal matrix of loadings for the idiosyncratic subspace, and $B_{(m)}$ is the $K_{(m)} \times Q_{(m)}$ of components for the idiosyncratic subspace. The matrix $\Omega_{(m)}$ is a matrix of mean-zero, independent, equivariant noise.

See Poole and Rosenthal (1997) for a discussion of double-centering.

[2] We chose notation consistent with Murphy (2012) and Klami, Virtanen and Kaski (2013). We denote all observed outcomes as $Y_{(m)}$ instead of Y and X.

We have modeled each observed data matrix $Y_{(m)}$ in terms of a shared subspace Z_S and individual subspaces $Z_{(m)}$. The researcher may believe, though, that the estimated scaled locations may vary systematically with some set of known covariates. Inspired by the work of Roberts et al. (2014), we model the scaled locations as

$$Z_S = X_S\beta_S + \Omega_{Z_S} \tag{7}$$

$$Z_{(1)} = X_{(1)}\beta_{(1)} + \Omega_{Z_{(1)}} \tag{8}$$

$$Z_{(2)} = X_{(2)}\beta_{(2)} + \Omega_{Z_{(2)}}. \tag{9}$$

The covariates $X_{(S)}$, $X_{(1)}$, and $X_{(2)}$ structure the systematic components of $Z_{(S)}$, $Z_{(1)}$, and $Z_{(2)}$, respectively. Throughout the analysis and simulations, we assume $X = X_{(S)} = X_{(1)} = X_{(2)}$, but our model reflects that this need not be the case.

We make four assumptions for identifying the model (for a discussion of identification, see Tipping and Bishop, 1999, Appendix A.1):

$$Z_S^\top Z_S = B_S^\top B_S = I_{Q_S} \tag{10}$$

$$Z_{(m)}^\top Z_{(m)} = W_{(m)}^\top W_{(m)} = I_{Q_{(m)}} \text{ for } m \in \{1, 2\} \tag{11}$$

$$Z_{(m)}^\top Z_S = \mathbf{0}_{Q_{(m)} \times Q_S} \tag{12}$$

$$L_{(1)}, L_{(2)}, D_{(1)}, D_{(2)} \text{ are diagonal with non-negative entries} \tag{13}$$

Assumptions (10) - (11) state that, within a given subspace, the latent scalings and components are uncorrelated and length one. Assumption (12) states that the common subspace spanned by Z_S is not correlated with the idiosyncratic scalings. This assumption allows us to differentiate the shared subspace from each idiosyncratic subspace. Assumption 13 identifies the particular rotation that we estimate. Specifically, we are assuming that $W_{(m)}$ and $B_{(m)}$ are principal components of the shared and idiosyncratic subspaces of $Y_{(m)}$, respectively.

We only identify the latent components $Z_S, Z_{(m)}, W_S$ and $B_{(m)}$ up to sign.[3] We follow convention and assume the elements of $L_{(m)}$ and $D_{(m)}$ are nonnegative and ordered in decreasing order.

2.3 A Probabilistic Framework

We next move from a geometric, least squares framework to a probabilistic framework. We fit the model in this project by the method of maximum likelihood, but the probabilistic formulation allows extension to a Bayesian framework, see in particular Hare et al. (2015); Klami, Virtanen and Kaski (2013); Gupta et al. (2011); Jackman and Trier (2008). Placing the method in this framework will leave it commensurate with popular random utility models (e.g. Ladha, 1991; Clinton, Jackman and Rivers, 2004), and allow us to use data augmentation schemes to accommodate binary, ordered, and count data (e.g. Albert and Chib, 1993; Goplerud, Forthcoming).

The probabilistic MD2S model. The probabilistic MD2S model can be written as

$$Y_{(m)k}|W_{(m)k}, B_{(m)k} \sim \mathcal{N}(Z_S L_{(m)} W_{(m)j} + Z_{(m)} D_{(m)} B_{(m)j}, \sigma^2_{(m)} I_N) \quad (14)$$

$$Z_{Siq}|X_{Si}, \beta_{Sq} \sim \mathcal{N}(\ X_{Si}\beta_{Sq}, \sigma^2_{Sq}) \quad (15)$$

$$Z_{(m)iq}|X_{(m)i}, \beta_{(m)q} \sim \mathcal{N}(\ X_{(m)i}\beta_{(m)q}, \sigma^2_{(m)q}) \quad (16)$$

This model is an extension of the Probabilistic Principal Components model of Tipping and Bishop (1999); see also Bach and Jordan (2005). We differ from these models in several regards. First, we are most interested in the actors' spatial locations $(Z_S, Z_{(m)})$, so we treat the weights $W_{(m)}$ and $B_{(m)}$ as random and the spatial locations as fixed (See also Aldrich and McKelvey, 1977, p. 117). This differs from factor analytic methods that are primarily interested in the weight matrices, and therefore treat the observation location as fixed. Second, we allow for modeling the spatial locations

[3] This means that the data cannot differentiate between a model with estimates $\{Z_S, Z_{(m)}, W_S, B_{(m)}\}$ and $\{-Z_S, -Z_{(m)}, -W_S, -B_{(m)}\}$.

with covariates. We maintain the assumption that the errors are of equal variance, so $\sigma^2_{(m)ik} = \sigma^2_{(m)}$ and therefore does not vary systematically across individuals or questions. This differs from Hare et al. (2015, Eqs. 2, 4), who allow for $\sigma^2_{(m)ik} = \sigma^2_{(m)i}\sigma^2_{(m)k}$.

Implementation. In the single dataset setting, Tipping and Bishop (1999) show that the maximum likelihood estimates for each factor are principal components of the data. We extend the result to the MD2S model. Doing so allows for an efficient estimation strategy, whereby we can estimate $Z_S, Z_{(1)}$, and $Z_{(2)}$ directly, then recover the remaining estimates afterwards.

Our algorithm allows the computational advantage of working with whichever is smaller, $N \times N$ or $K_{(m)} \times K_{(m)}$ matrices.[4] For example, in the example below we observe 29 voting members, 33 votes, but 6,931 words. Our algorithm is fit through manipulating matrices of size 29×29 instead of 33×33 or $6,931 \times 6,931$. The computational gains are sizable. Our algorithm estimates the MD2S model using an iterative procedure that updates the estimate of each subspace one at a time, enforcing the constraints in Equations 10–13 along the way. We include in the Supplemental Materials a full description of the algorithm and a proof that it recovers the desired estimates.

Our primary technical contribution is in establishing that our algorithm recovers the correct estimates. The algorithm starts with an estimate of the idiosyncratic dimension for each dataset, partials these out in order to estimate the shared space, and then updates the idiosyncratic space estimate given the new shared space estimate. We prove the validity of this strategy in the following proposition,

PROPOSITION 1 *The maximum likelihood estimates for the shared and idiosyncratic subspaces can be written as singular vectors of functions of the data. Specifically:*

1. *The maximum likelihood estimates for $Z_{(m)}$ are proportional to principal components of $Y_{(m)}^\top(M(Z_S))$*

[4] See Aldrich and McKelvey (1977, p. 117) and Tipping and Bishop (1999, Appendix B) for similar insights.

for $m = 1, 2$.

2. *Denote* $Z_{S|m}$ *as the first* L_S *principal components of* $Y_{(m)}^\top M(Z_{(m)})$*. Then, (a)* $Z_S \propto Z_{S|1}\vec{w}_1 + Z_{S|2}\vec{w}_2$ *with* $\vec{w}_{1j} + \vec{w}_{2j} = 1; w_{1j}, w_{2j} > 0$ *and (b) selected to maximize* $\operatorname{tr}\left(Z_S^\top Y_{(1)} Y_{(1)}^\top Y_{(2)} Y_{(2)}^\top Z_S\right)$.

where $M(A)$ *is the annihilator matrix for matrix* $A : I - A(A^\top A)^- A^\top$ *with* $-$ *denoting the generalized inverse and denote as* $H(A) = I - M(A)$*, with* I *the commensurate identity matrix.*

Proof. *See Supplemental Materials.*

The proposition lends itself directly to our estimation strategy, for which we provide details in the supplemental materials.

2.4 Uncertainty

We estimate uncertainty for two parts of the MD2S model: the scaled locations and the number of dimensions. For the scaled locations, we rely on the bootstrapping methodology introduced by Jacoby and Armstrong II (2014). Let $\{\widetilde{Y}^*_{(1),b}, \widetilde{Y}^*_{(2),b}\}$ denote two b^{th} bootstrapped sample, with $b \in \{1, 2, \ldots, B\}$, where B is some large number, such as 1,000. The bootstrapped sample is generated by fixing the number of rows and sampling $K_{(m)}$ columns for each matrix, with replacement. Uncertainty due to sampling error can be estimated through fitting MD2S to these bootstrapped estimates.

We present a statistical method for estimating the number of dimensions while acknowledging that the first empirical consideration should be substantive interpretability of the estimated subspaces. We recommend separating signal from noise dimensions through the use of a permutation test (e.g. Keele, McConnaughy and White, 2012). A permutation test requires estimating the density of a test statistic on a set of datasets permuted such that under the null hypothesis, there is in-truth no signal in the data, and then the observed value is compared to this simulated null distribution.

In our case, we assume that there is no structure in the data, so the subspace loadings are all zero. Formally, we assume

$$\mathcal{H}^0_{(m)L,q} : L_{(m)q} = 0; \tag{17}$$

$$\mathcal{H}^0_{(m)D,q} : D_{(m)q} = 0 \tag{18}$$

for all (m, q). Under these hypotheses, the observed data is pure noise with no systematic structure, i.e. $Y_{(m)} = \Omega_{(m)}$.

We simulate statistics under these null hypotheses and then compare the statistic under the observed data to the data under the null distribution. To the extent that the statistic is an outlier under the null hypothesis, we can argue that the null hypothesis is not accurate and there is in fact some systematic relationship in the data.

Specifically, we permute the data such that within each column of $Y_{(m)}$, the rows are shuffled. In this case, in truth, there is no systematic relationship in the permuted data. Denote the r^{th} permuted dataset out of R total as $\widetilde{Y}^r_{(m)}$, with R some large number, say 1000. For each permuted dataset, we calculate the dimension weights, $\widehat{L}^r_{(m);q}$ and $\widehat{D}^r_{(m);q}$. These values are then compared to the estimated values on the non-permuted data, $\widehat{L}_{(m);q}$ and $\widehat{D}_{(m);q}$.

Under this formulation, a p-value for dimension q in the shared subspace or idiosyncratic subspace can be estimated as

$$\widehat{p}_{S;q} = \frac{\sum_{r=1}^{R} \mathbf{1}\left(\widehat{L}^r_{S,q} \leqslant \widehat{L}_{S;q}\right)}{R}; \quad \widehat{p}_{(m);q} = \frac{\sum_{r=1}^{R} \mathbf{1}\left(\widehat{D}^r_{(m),q} \leqslant \widehat{D}_{(m);q}\right)}{R}; \tag{19}$$

In our analyses and simulations, we take the standard cut-off of $p = 0.1$.

We adapt the test to our model by noting that the tests are not independent. The dimensions are estimated in order of decreasing loadings, such that more explanatory dimensions are estimated before less explanatory ones. Therefore, we take as our estimated dimensionality the first d dimensions such that each dimension has an estimated p-value below 0.1. We calculate the estimated dimensionality $\widehat{d} = q$ as the largest q such that dimensions 1 to q have estimated p-values below 0.1.[5]

2.5 Extensions

By using a probabilistic model, we are able to augment the model in order to extend the MD2S model to a large class of problems. For example, we can turn the model into a quadratic utility model through utilizing the latent normal representation of a probit model (Clinton, Jackman and Rivers, 2004; Albert and Chib, 1993). We can also utilize scale- and location-mixtures of normals to accommodate ordinal and count data, as in Goplerud (Forthcoming). In this framework, our probabilistic model is the "M"-step of an EM routine, with the "E"-step as an adjustment to the observed data.[6] Our concern here is not with accommodating a particular class of data, such as binary, ordinal, or count data, but instead to develop a framework for integrating multiple sources in a single coherent fashion.

If the researcher is interested in scaling in a geometric, rather than probabilistic, framework, several other possible extensions can be integrated in to MD2S. Our algorithm involves decomposing the similarity matrix $\frac{1}{K_{(m)}} Y_{(m)} Y_{(m)}^\top$, such that the identifying constraints in Equations 10–13 are met. Rather than a correlation matrix, any $N \times N$ similarity or dissimilarity matrix may be used; for an example analyzing a dissimilarity matrix, see Jacoby and Armstrong II (2014).

[5]Formally, $\widehat{d}_S = \text{argmin}_q\{q : \widehat{p}_{S;q} > 0.1\} - 1$; $\widehat{d}_{(m)} = \text{argmin}_q\{q : \widehat{p}_{(m);q} > 0.1\} - 1$

[6]For example, in a latent probit model, this step involves adding to the fitted values the mean of a normal covariate truncated at 0 and centered at the fitted value, with support above zero for observed values of "1" or below zero for observed values of "0," and support over the whole line for missing values. See, Clinton, Jackman and Rivers (2004)

We have also assumed that all of the columns in $Y_{(m)}$ are on the same scale. If an analysis requires combining data on different scales, say a combination of continuous and categorical outcomes, we have two suggestions. First, if all of the data is continuous and approximately normal, each column may be converted to a z-scale by subtracting off the mean and dividing by the sample standard deviation. A recent literature has also suggested placing data on the same scale through an inverse z-transformation of the empirical distribution function,

$$Y^{z}_{(m);ij} = \Phi^{-1}\left(\frac{1}{N+1}\sum_{i'=1}^{N}\mathbf{1}\left(Y_{i'j} \leqslant Y_{ij}\right)\right) \tag{20}$$

where $\Phi(\cdot)$ is the normal distribution function. For more on this and other methods, see Quinn (2004); Hoff (2007); Murray et al. (2013).

3 Validation Exercises

We have conducted three sets of validation exercises to assess the reliability of the proposed method. In the simulation study, we compare estimated with respect to true latent scales. In the next two, we analyze observational datasets in American and comparative politics where we know the main dimensions of interest. Our goal was to conduct a series of exercises to establish that the method performs well across data types, recovering relevant dimensions in different subfields of political science, and relative to existing methods. For the sake of conciseness, a full description of each analysis is available in the Supplemental Appendix. We summarize the results from these exercises below.

We have conducted an extensive simulation study in order to assess the method's performance across two different dimensions: first, its ability to identify common and idiosyncratic components; and second, its ability to recover the relationship between scaled locations and covariates. Our simulations range across a number of settings, with the number of observations and number of features in the data. We assess the method in both relative and absolute terms. We first compare the method to existing software implementing inter-batter factor analysis, that recovers latent shared and idiosyncratic subspaces: `plsca` in **R** library `plsdepot` and the method of Klami, Virtanen and Kaski (2013). We show that the performance, in terms of the correlation between the true subspace and its estimate, of the other methods decay as K_2 grows, i.e. if the number of features in one dataset grows while the number of observations is held fixed. In addition, we show that MD2S is immune to this problem, growing more precise in K_2. We also show that MD2S is as good as, and often better, than these other methods in returning estimates of the idiosyncratic dimension. It is important to note that none of the alternative methods recover the spatial location within each estimated dimensions, which is our main output of interest. Next, we assess the permutation test for distinguishing systematic dimensions from noise. We show the permutation test is a powerful means of identifying systematic dimensions and makes few false positives. Lastly, we assess the ability of the method to correctly recover estimates of the effect of covariates. We show that when there is a systematic relationship between scaled locations and covariates, we can recover it accurately, and when there is no relationship, our regressions return point estimates and p-values that generate false positives at the nominal level.

In our first empirical exercise, we reanalyze data from Kim, Londregan and Ratkovic (2018). In this work, the authors developed a latent utility model for combining vote and textual data on the same latent space, and then estimated the model through placing both sets of data on the same latent z-scale. The authors never provided a satisfactory method for balancing the information coming from words and votes, which provided one motivation of the current study. The data we use consist of a rollcall matrix and term-document matrix from the 112th Senate, as was analyzed in the original work. Using MD2S, our shared dimension recovers the same first dimension uncovered by scholars for decades (e.g. Poole and Rosenthal, 1985), as well as a very similar ordering of words uncovered by Kim, Londregan and Ratkovic (2018), with parliamentary words on one end (*meet session, author meet, 10 a.m.*) and budgetary words on the other (*stimulus, trillion, rais tax*). The votes dimension reveals an idiosyncratic dimension that is informed by members' votes, but not their floor speech. This dimension ranges from national security to agricultural concerns.

Our second analysis combines two different characteristics of countries, political attributes and socioeconomic attributes, using MD2S. We find that the POLITY IV measure of Gurr, Marshall and Jaggers (2010) is a cross-cutting measure across our shared and idiosyncratic political dimension. The ordering in the shared dimension is sensible, ranging from war-torn Sierra Leone and Liberia on one end to the Netherlands on the other, and is anchored by measures of female empowerment, infant mortality, egalitarian politics. The second dimension, the idiosyncratic political dimension, ranges along an axis from countries with more corruption than oppression (e.g. Cape Verde) to more oppression than corruption (e.g. Turkmenistan). The idiosyncratic socioeconomic dimension ranges from resource rich countries (e.g. Qatar) to countries with large urban populations (USA, China). Resource income anchors the one end of this dimension, while the other is anchored by urban population. We also find secular improvement over time in the shared dimension, with no systematic change in either idiosyncratic dimension. Again, MD2S is returning a coherent and sensible image of data in a well-studied setting.

These validation exercises are placed in the Supplemental Appendix, as they are not our applied setting of central interest, but they build confidence in the results from our substantive example. We move next to our primary substantive analysis.

4 Scaling FOMC Members

MD2S is particularly well-suited to scale policy makers when their choices are multidimensional and evidenced through multiple types of data. Such is the case with the Federal Open Market Committee (FOMC). The FOMC is charged by statue with a dual mandate of keeping both unemployment and inflation at low levels. Its primary policy mechanism is the overnight rate at which banks can borrow from the Federal Reserve, in effect impacting interest rates for the economy as a whole. The FOMC consists of seven members of the Board of Governors, appointed by the President and confirmed by the Senate, and twelve Presidents of the regional Reserve's Banks. Members meet at eight regularly schedule meetings a year and beyond that as needed. At each meeting, the Governors, the President of the New York Federal Reserve, and a rotating set of four of the remaining Presidents can vote, but all members participate in a policy deliberation process that is recorded in the transcripts. For a description and history of the FOMC, see Meltzer (2004, 2010).

While we are the first to integrate words, votes, and covariates in a scaling model, we build on earlier work on the FOMC. Chappell, Havrilesky and McGregor (1993); Krause (1994); McGregor (1996); Chang (2003); Chappell, McGregor and Vermilyea. (2005); Meade and Stasavage (2008); Eijffinger, Mahieu and Raes (2015) analyzed dissenting votes and policy recommendations, but did not connect the vote data explicitly with speech. Focusing only on votes, especially near the 2008 collapse, ignores information in stated reasons for the vote as well as the discussion of different policy alternatives. Separate work has used automatic content analysis to characterize the speech of members (Bailey and Schonhardt-Bailey., 2008; Schonhardt-Bailey, 2003; Hansen, McMahon and Prat, 2014; Egesdal, Gill and Rotemberg, 2015), yet votes are not included into the statistical model. While interesting patterns in word usage emerge, they are not explicitly connected with policy choices. Baerg et al. (2014) estimate underlying policy preferences from words, under the assumption that members follow a Taylor Rule, balancing inflation/growth tradeoff. Yet a Taylor Rule is not applicable during most of the period we study, since the financial crisis exhausted conventional monetary policy.

Data. Our data come from FOMC meeting transcripts during the publicly available portion of Chairman Bernanke's term (2006–2010). As FOMC members work through an expectation channel, they are only as effective as their public statements serve as a credible commitment to future policy action. This induces a "norm of consensus (Epstein, Segal and Spaeth, 2001)," in their voting record, as FOMC members disagree in the meetings but report a public vote that is unanimous or nearly so (Meade and Stasavage, 2008; Riboni and Ruge-Marcia, 2010). Pressure towards consensus in the official vote also comes from the Chair, who acts as agenda-setter and issues his policy proposal early in deliberations. These constraints are evidenced in two facts: first, in the FOMC's history, the Chairman's policy directive has never lost in the official voting records; and, second, the Chair's directive is statistically indistinguishable from both the mean and median policy recommendation of the whole Committee (Chappell, McGregor and Vermilyea, 2004).

For vote data, we use stated preferences as recorded during informal and preliminary votes in the policy go-round of the deliberation process rather than the publicly reported vote. We gathered the informal vote data from 39 official policy meetings, of which 33 returned non-unanimous policy recommendations. We classify members' desired policy as a binary choice. Before January 2008, we code an expressed recommendation as a "1" if the member's stated preference on interest rate adjustment is equal or higher than the staff median policy and a "0" otherwise.[7] From January

[7]Policy scenarios proposed by the staff of the Board of Governors at each policy meeting are contained in the document known as the *Bluebook*, which is publicly available at `http://www.federalreserve.gov/monetarypolicy`.

2009, with the policy rate rates nearly at the zero lower bound, the primary policy mechanism available to the FOMC shifted from interest rates to asset purchases, mainly of government and private sector securities. We coded members as a "1" if they recommended a level of asset purchases at or above the staff's median proposal and a "0" otherwise. We include more detailed information on how we coded these votes in the Supplemental Appendix.

For text data, we use the verbatim transcripts from the monetary policy meetings during the period under study.[8] We identify all members' statements within any of the two main sections of committee meetings: the economic situation discussion, where board Governors and district Presidents offer their own views about the economic situation both at the regional and national level, and the policy go-around, where members give their own policy recommendations with respect to the policy options presented by the staff. We process the text data through stemming, removing stopwords, and trimming sparse terms. We are left with $18,209$ unique tokens, though we note we receive qualitatively similar results so long as we maintain more than 500 tokens. We analyze the weighting probability matrix given by the proportion of tokens of each unique word spoken by each FOMC member so as to normalize the data and reduce the impact of outliers. Our text processing procedure is standard, so we defer details to the Supplemental Materials.

We also incorporate member-specific covariates to inform the recovered latent dimensions. We include members' age, date of appointment, and two sets of dummy variables. The first is a dummy variable coded 1 if the member is a Regional President and a 0 if they are a Governor. Using the approach from Adolph (2013), we also include a variable coded as 1 if the member spent more than half of their pre-FOMC career in banking and finance, a second coded as 1 if they spent more than half of their pre-FOMC career in the academy, with "other" the omitted category.[9]

[8] The set of transcripts is also publicly available at `www.federalreserve.gov/monetarypolicy`

[9] The career experience measure for each category is the fraction of the FOMC member's career spent in that job

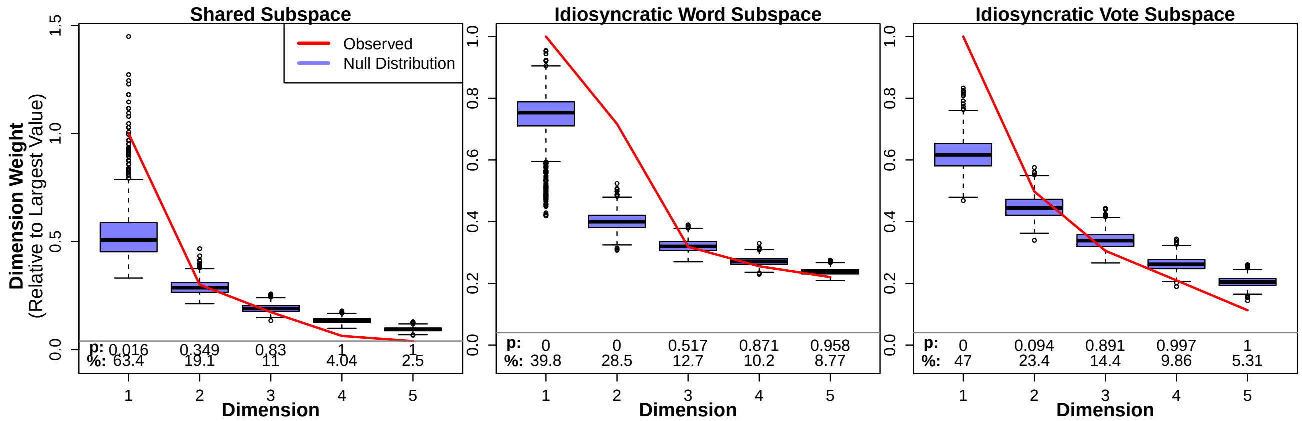


Figure 1: **Results from Permutation Test for Number of Dimensions.** The dimension number is along the x-axis and the y-axis contains the estimated dimension relative to the largest estimated dimension. The red line contains the estimated dimensions; the blue box and whisker plots contain the permuted values estimating the null distribution. The estimated p-value is reported along the top row of the table at the bottom. The percentage of variance explained is in the second row. Using a p-value threshold of 0.1 gives us two dimensions for the shared subspace and idiosyncratic vote subspace and three for the idiosyncratic word space.

4.1 Results

We present the results in four parts. First, we use our permutation test to assess which latent dimensions are the most informative (i.e., are not likely noise). Second, we conduct a leave-one-out cross-validation exercise to demonstrate that adding the text data generates scaled locations with a higher predictive power than those using votes alone. Third, we examine the substance of the scaled locations in the shared subspace, informed by both words and votes. Fourth, we compare results on this subspace from before and after the crisis. Lastly, we discuss insights from the covariates and idiosyncratic subspaces.

Estimating dimensionality. Figure 1 presents the results from the permutation test applied to the FOMC data. The three figures contain the results for the shared subspace and then the two idiosyncratic dimensions. The dimension number is along the x-axis and the y-axis contains the

category up to the date of her most recent appointment as a FOMC member.

dimension weight relative to the largest estimated dimension. The red line contains the estimated dimensions; the blue box and whisker plots contain the permuted values under the null distribution of no systematic structure. The estimated p value is reported along the top row of the table at the bottom. The percentage of variance explained is in the second row.

Using a p-value threshold of 0.1 gives us one strong dimension across the shared subspace and across both the idiosyncratic vote and word subspaces. For instance, the first dimension of the shared subspace explains around 64% of the variance in both votes and words. Thus, we focus our discussion on the first, and largest, of each of the subspaces.

Assessing internal validity. To do so, we conducted a leave-one-out predictive exercise to assess how well the recovered latent locations explain the binary policy recommendations of FOMC members. In particular, we assess the method's predictive performance, as well as the informativeness of members' speech in predicting their recommendations. We focus on the 33 meetings where policy recommendations are non-unanimous and fit the model to all meetings save one. Then, we use the estimated scaled locations to predict the held-out meeting. We also include results from two other scaling methods that only exploit the variation in the vote data: Optimal Classification Poole (2005) and IDEAL (Clinton, Jackman and Rivers, 2004).

Different methods minimize different loss functions, so we entered the estimated scale locations into a least-squares regression (first row, the loss function for MD2S and PCA), a probit regression (second row, the loss function for IDEAL), and a loss function given by misclassified votes (third row, the loss function for OC). The best-performing methods given these metrics are in bold and the next-best is underlined. The first column of the table shows as reference the results of a null model assuming a homogenous probability or recommending the modal policy for each vote.

Since MD2S attempts to minimize squared error (R^2), the improvement in fit from PCA with

	Null (Intercept)	PCA (Votes)	M2DS (Votes and Words)	M2DS (Covariates)	Ideal	OC
R^2	0.000	0.294	**0.371**	0.338	0.287	0.033
log-Lik	-8.243	-5.298	**-4.689**	-5.281	-4.943	-7.946
Class	13.211	**15.211**	14.947	14.632	14.789	13.316

Table 1: **Leave-one-out Predictive Accuracy of Different Methods.** We have 33 votes with at least one observed Yay and Nay recommendation. To assess internal validity, we dropped each of these recommendations one at a time and fit three versions of MD2S: MD2S using just votes, MD2S combining text and votes, and MD2S using text and votes with covariates. For comparison, we also include results from Optimal Classification and IDEAL. We then use the scale locations to estimate the held-out vote. Methods were assessed by R^2 (first row), probit likelihood (second row), and classification rate (third row). The best-performing method is in bold, the next-best is underlined. MD2S tends to outperform other methods, except under a classification loss, where M2DS with votes performs the best and M2DS with words is the second best method.

votes to either MD2S variant is promising. Adding words improves prediction in the held-out votes by over 20%, indicating that information from speech is being incorporated successfully. We find a similar pattern when considering the log-likelihood loss. PCA does outperform both MD2S specifications under classification loss, but the model that adds words outperforms both IDEAL and OC. In terms of predictive power, MD2S with votes and words generally outperforms both IDEAL and OC on this dataset. As these methods are all readily available, and we recommend validation checks along the lines presented here for selecting among the different scaling algorithms. Finally, notice that in this particular application, incorporating covariates does not seem to improve the performance of the model, a point we return to below.

The shared subspace. We present the main results of the scaling model in Figure 2. The righthand panel shows the scaled location of each FOMC member for the entire period 2006–2010. The scaled locations recover two clusters of FOMC members.

In one cluster we find known inflation “hawks” district Presidents Thomas Hoenig (Kansas City Fed), Richard Fisher (Dallas Fed), Jeffrey Lacker (Richmond Fed), and Charles Plosser (Philadel-

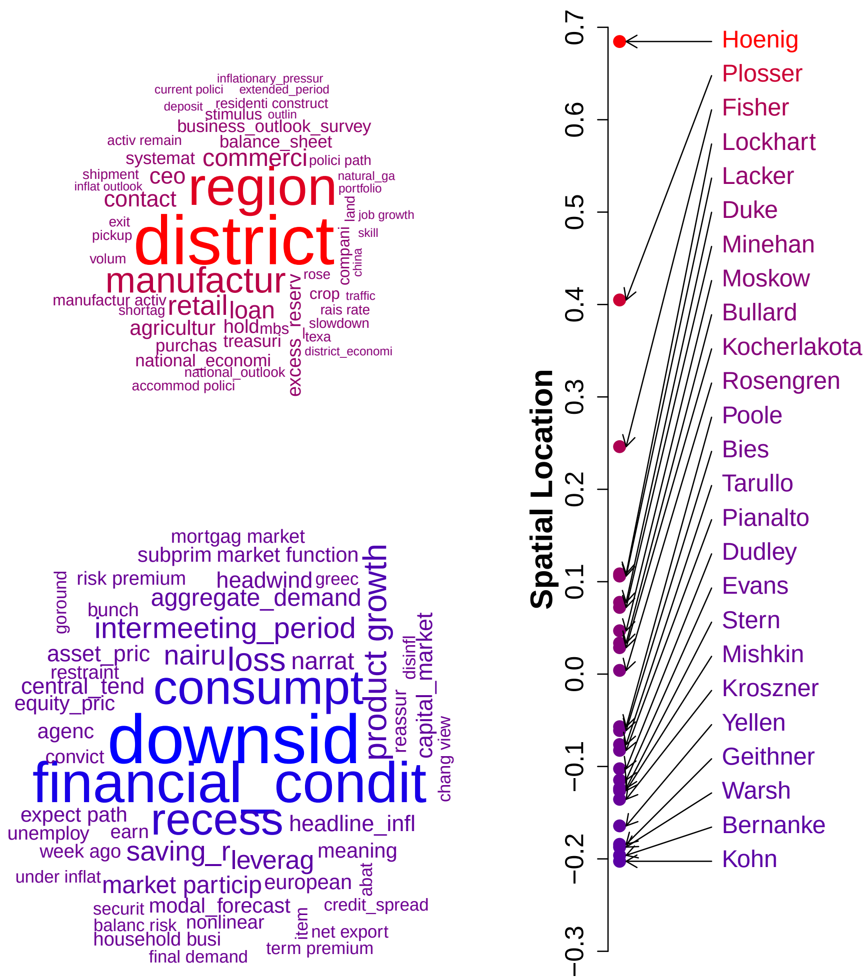
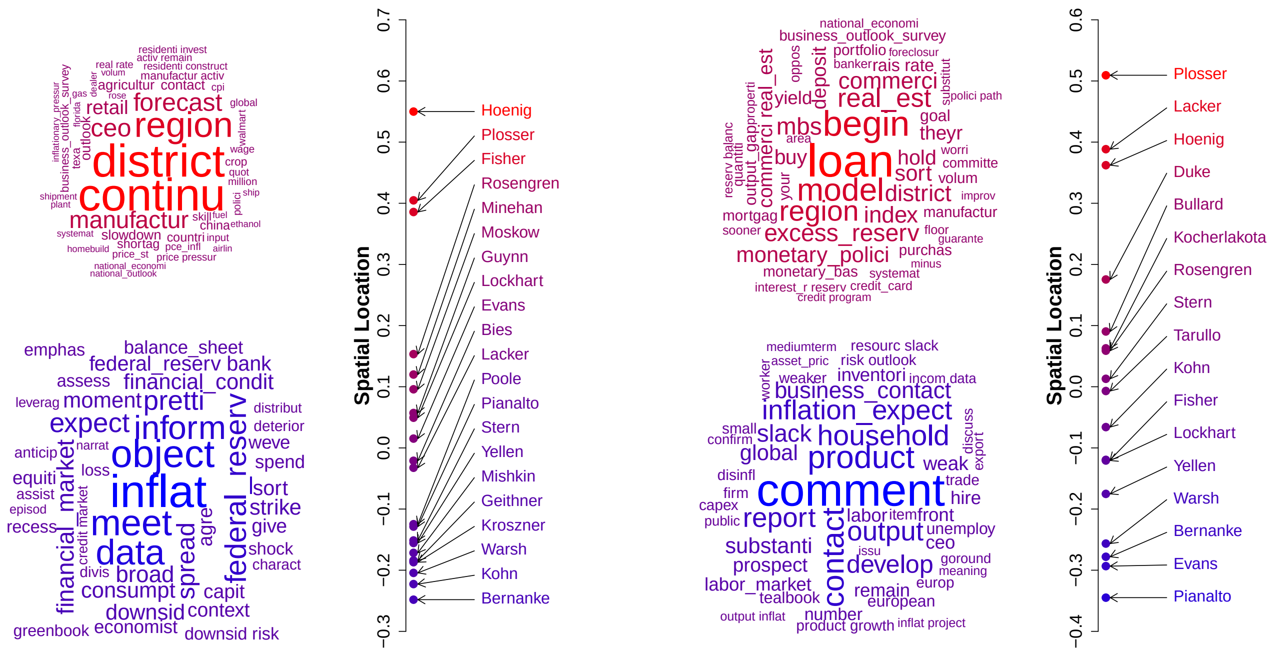


Figure 2: **Spatial Locations of Words and FOMC Members for the Shared Subspace of Speech and Policy Recommendations (The Entire Bernanke Period).**

phia Fed) who, during their terms, prioritized low inflation and were wary of governmental asset purchases after the crisis. On the other side of the spectrum, we find the known "doves" Governors Janet Yellen, Randall Kroszner, Frederic Miskin, and Kevin Warsh, who were strong supporters of keeping the policy rate pegged at zero for an extended period of time after the crisis and favored strong intervention by the Federal Reserve after the crisis.

MD2S not only orders the members, but helps us identify the words associated with each dimension. This is shown in the leftside panel of Figure 2 in the form of "wordclouds", where larger words denote higher term loadings. On the upper side of the scale we see a set of terms associated with economic conditions at Fed presidents' regions such as *district*, *region*, and *contact*, that Hoenig, Plosser, Fisher and Lockhart, among other Fed Presidents, systematically employed at the economic go-round when explaining the status of their regional economies. When the FOMC still had some room to maneuver on the policy rate, "hawks" emphasized terms associated with inflation risks such as *inflation outlook*, *raise rate*, *inflationary pressures*, while known doves were more likely to use terms associated with real output and unemployment like *recession*, *product growth*, *nairu* (natural rate of unemployment), *disinflation*, *unemployment*.

The words anchoring each dimension can help guide our interpretation of the text, as they indicate words that correspond with the extremes in vote choice. This contrasts with topic model approaches, where the topics in the text are estimated in a manner unconnected to the votes (Hansen, McMahon and Prat, 2014; Lauderdale and Clark, 2014). For example, a key area of disagreement after the crisis was whether to expand asset and bond purchases and whether to maintain the overnight rate near zero for an extended period, with hawks opposing and doves in favor. MD2S associates the tokens *downsid* and *financial_condit* with the "dovish" side of the spectrum. In the March, 2009 meeting Governor Kohn argued in favor of expanding the asset purchase program: *"It really concerned me over the intermeeting period that there was this generalized tightening of* **financial conditions** [...] *at exactly the wrong time. So I think we need to lean against that tendency in a very visible, transparent way. I think buying longer-term Treasuries is the way to do that. I think it will work in the sense of lowering interest rates."* This contrasts with statements from President Hoenig during the same meeting, who expressed concern about his district but also about the size of the asset purchases: "*I would say that the Tenth* **District** *is* [...] *doing better than the nation as a whole but is now quickly deteriorating into the national average* [...] *I think*



Before crisis After crisis

Figure 3: **Spatial Locations of Words and FOMC Members for the Votes and Words' Subspaces of Speech and Policy Recommendations (Before and After).**

it's sadly ironic that, after we've doubled the size of our balance sheet, we're having a conversation about being worried that it may shrink a little bit [. . .] *I don't think we can turn it into a monetary policy instrument in term of how much we expect the balance sheet to grow at this point.*"

Before and after the crisis Turnpoint. In Figure 3, we repeat the previous analysis on two subsets of our data: before and after the financial crisis turnpoint given by the demise of Lehman Brothers in September 2008. We see little movement in scaled locations; most members that were on the extremes in the shared dimension before 2009 remained there after (Hoenig, Plosser, Bernanke, Warsh, Yellen). Additionally, MD2S is able to capture changes in the location of Fed Presidents Evans and Pianalto who, before 2009, are ranked as moderate FOMC members, but are estimated as more extreme after the crisis under the new policy regime of credit easing. Both Pianalto and

Evans increasingly turned more supportive of the credit expansion policies of the Fed, allying with Bernanke in every single policy recommendation during 2009 and 2010.

The spatial location of the words used before and after the crisis evidence a drastic change in the concerns and discussion within the FOMC meetings. Before the crisis, we see the hawks focused more on their district's economic situation (*district*, *region*, *contact*). After the crisis, inflation hawks put more emphasis on addressing the failures leading to the crisis (*loan*, *realestate*, *mortgages*, *foreclosures*) and on the unconventional policies being pursued (*monetarypolicy*, *policypath*, *creditprogram*, *excessreserves*, *mbs*). Inflation doves before 2009 emphasized the financial risks of the crisis (*financialmarkets*, *financialconditions*, *downsiderisk*, *balancesheet*). Once the crisis turned into a recession, inflation doves focused more on the macroeconomic impact on output, unemployment and inflation (*slack*, *riskoutlook*, *productgrowth*, *output*, *unemployment*, *labor*, *worker*, *inflationexpectations*, *disinflation*).

Idiosyncratic subspaces and covariates. Figure 4 shows the votes-only subspace, a scaling based only of information in the votes and uncorrelated with the shared subspace. In the shared subspace, Hoenig, Plosser, Fisher, and Lacker all clustered at one end of the shared subspace (see Figure 2). The idiosyncratic votes-space differentiates these four, placing Hoenig and Fisher on one end and Plosser and Lacker on the other. Even though all four voted for less accommodative policies with respect to the rest of the committee, they did so in different circumstances. Lacker and Plosser supported a relatively higher policy rate during 2006 and 2007 where conventional monetary policy was still feasible, while Hoenig and Fisher supported less expansionary policies than the rest of the FOMC in terms of the Fed's asset purchases program during 2009 and 2010, when the policy rate was at the zero lower bound. Similarly, the words-only subspace distinguishes a cluster of moderate FOMC members, with Poole and Minehan from Dudley and Duke, despite all four clustering in the

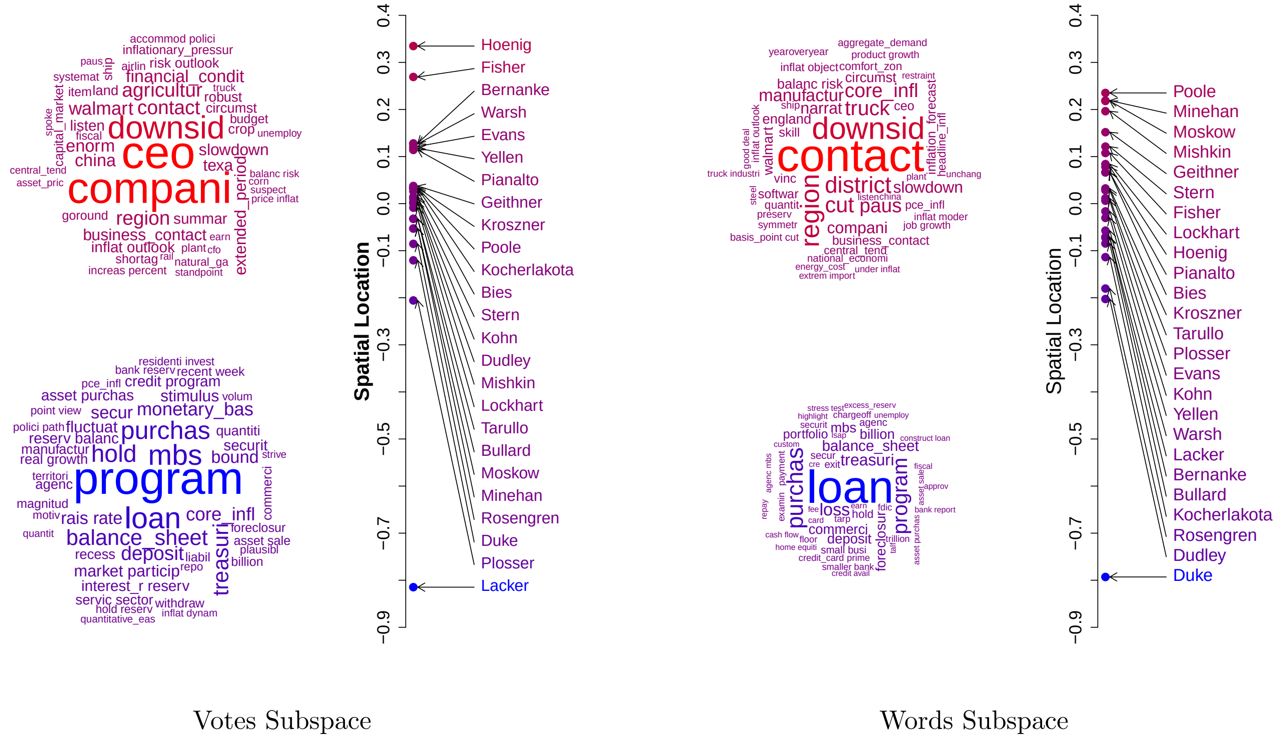


Figure 4: **Spatial Locations of Words and FOMC Members for the Votes and Words' Subspaces of Speech and Policy Recommendations.**

center of the shared subspace. We see a similar spread between the two sides, with the positive side focusing on local issues and regional economies (*contact*, *region*), and the other side more putting more emphasis on the national financial situation (*loan*, *purchase*, *balance_sheet*).

Regression results giving the relationship between the scaled locations and the member covariates in Table 2. We note these regressions have little power, with only 29 observations and 5 covariates, so our results are simply suggestive. We find some evidence that, as a group Regional Presidents tend to be on one side of the latent votes-only subspaces, systematically recommending relative higher policy rates than Board Governors, who in addition use more terms associated to the financial situation than Presidents (*downsid*, *financial_condit*).

	Shared	Votes	Words
Intercept	7.67	10.71	5.47
	(12.18)	(11.94)	(12.55)
Year Born	0.18	0.33	0.19
	(0.57)	(0.56)	(0.59)
Year Appointed	−0.55	−0.85	−0.46
	(0.69)	(0.68)	(0.71)
Regional President	17.32	17.39	−4.24
	(8.70)	(8.53)	(8.97)
Banking/Financial	−1.83	−0.80	10.24
	(8.66)	(8.49)	(8.93)
Academic/Government	5.72	0.42	−10.58
	(16.53)	(16.20)	(17.04)
N	29	29	29
R^2	0.16	0.19	0.11
adj. R^2	−0.02	0.02	−0.09
F	0.87	1.09	0.55
p-value	0.52	0.39	0.74
Resid. sd	0.19	0.19	0.20

Standard errors in parentheses

Table 2: **Estimated Impact of Covariates on Each Scaled Location.** We do not find any significant effects using the covariates in the FOMC example. We cannot reject the global null hypothesis that all coefficients are zero Considering marginal significance, we find suggestive evidence that regional Presidents behave differently in both the shared and votes dimensions .

5 Conclusion

As we enter a period of "big data," we encourage political scientists to think not just of analyzing large datasets but also how to combine data from disparate sources. We present such a method here, for scaling data from two separate datasets. The method, MD2S, successfully incorporates information from two different data sources, generating scaled locations with a higher internal validity than analyzing the two datasets separately. We include methods for checking validity, separating systematic dimensions from noise, and a way to relate scaled locations to covariates, all fit using an efficient statistical algorithm.

The method also allows the user to use the scaled locations from both datasets to help infer the meaning of the latent dimensions. In our example, FOMC scaled locations were also associated with words that let us better interpret the meaning of the latent scale. The idiosyncratic subspaces also offer a new mode of analysis, allowing us to identify ways in which the members at the extremes of the shared subspace differed. Though our regressions for connecting covariates to scaled locations were underpowered, we believe that such a tool will prove useful in other settings.

We anticipate several ways in which this project can be moved forwards. First, we have presented the method in a geometric, least squares framework. Placing the method in a probabilistic framework will allow for an extension to commonly used Bayesian techniques (Hare et al., 2015; Tipping and Bishop, 1999). We also plan to extend the method to allow for cross-time comparisons, so as to place multiple observations in the same space over time.

References

Adolph, Christopher. 2013. *Bankers, Bureaucrats,and Central Bank Politics: The Myth of Neutrality.* Cambridge Studies in Comparative Politics Cambridge University Press.

Albert, James H. and Siddhartha Chib. 1993. "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association* 88:669–679.

Aldrich, John and Richard McKelvey. 1977. "A Method of Scaling with Applications to the 1968 and 1972 Presidential Elections." *American Political Science Review* 71(1):111–130.

Anderson, T.W. 1989. "Linear Latent Variable Models and Covariance Structures." *Journal of Econometrics* 41:91–119.

Bach, F. and M. Jordan. 2005. A probabilistic interpretation of canonical correlation analysis. Technical Report 688 Department of Statistics, University of California at Berkeley.

Baerg, N. R., W. Lowe, S. Ponzetto, H. Stuckenschmidt and C. Zirn. 2014. Estimating central bank preferences. In *Workshop on Language Technologies and Computational Social Science*, ed. Maryland. ACL 2014, Baltimore.

Bafumi, Joseph and Michael Herron. 2010. "Leapfrog Representation and Extremism: A Study of American Voters and Their Members in Congress." *American Political Science Review* 104(3):519–542.

Bailey, A. and C. Schonhardt-Bailey. 2008. "Does deliberation matter in FOMC monetary policymaking? The Volcker Revolution of 1979." *Political Analysis* 16(4):404–427.

Barbera, Pablo. 2016. “Birds of the Same Feather Tweet Together. Bayesian Ideal Point Estimation Using Twitter Data.” *Political Analysis* Forthcoming.

Bonica, Adam. 2014. “Mapping the Ideological Marketplace.” *American Journal of Political Science* 58(2):367–386.

Browne, M. W. 1979. “The maximum-likelihood solution in inter-battery factor analysis.” *British Journal ofMathematical and Statistical Psychology* 32(75–86).

Chang, Kelly H. 2003. *Appointing central bankers: The politics of monetary policy in the United States and the European monetary union.* Cambridge University Press.

Chappell, Henry W., Rob R. McGregor and Thomas M. Vermilyea. 2005. *Committee decisions on monetary policy: Evidence from historical records of the Federal Open Market Committee.* The MIT Press.

Chappell, Henry W., Rob Roy McGregor and Todd Vermilyea. 2004. *Committee Decisions on Monetary Policy.* MIT Press.

Chappell, Henry W., Thomas M. Havrilesky and Rob R. McGregor. 1993. “Partisan monetary policies: Presidential influence through the power of appointment.” *The Quarterly Journal of Economics* 108(1):185–218.

Clinton, Joshua, Simon Jackman and Doughlas Rivers. 2004. “The Statistical Analysis of Roll Call Data.” *American Political Science Review* 98:355–370.

Coppedge, Michael, John Gerring, Staffan I. Lindberg, Svend-Erik Skaaning, Jan Teorell, with David Altman, Michael Bernhard, M. Steven Fish, Adam Glynn, Allen Hicken, Carl Henrik Knutsen, Kelly McMann, Pamela Paxton, Daniel Pemstein, Jeffrey Staton, Brigitte Zimmerman, Rachel Sigman, Frida Andersson, Valeriya Mechkova and Farhad Miri. 2015. "V-Dem Codebook v5." Varieties of Democracy (V- Dem) Project.

Egesdal, Michael, Michael Gill and Martin Rotemberg. 2015. "How Federal Reserve Discus- sions Respond to Increased Transparency." Available at SSRN.

Eijffinger, Sylvester CW, Ronald Mahieu and Louis. Raes. 2015. "Hawks and Doves at the FOMC.".

Epstein, Lee, Jeffrey Segal and Harold Spaeth. 2001. "The Norm of Consensus on the US Supreme Court." *American Journal of Political Science* 45(2):362–377.

Goplerud, Max. Forthcoming. "A Multinomial Framework for Ideal Point Estimation." *Political Analysis* .

Gupta, Sunil Kumar, Dinh Phung, Brett Adams and Svetha Venkatesh. 2011. *Advances in Knowledge Discovery and Data Mining*. Lecture Notes in Computer Science Springer-Verlag chapter "A Bayesian Framework for Learning Shared and Individual Subspaces from Multiple Data Sources", pp. 136–147.

Gurr, Ted R, Monty Marshall and Keith Jaggers. 2010. "Polity IV Project: Political Regime Characteristics and Transitions, 1800-2009.".

Hansen, Stephen, Michael McMahon and Andrea Prat. 2014. "Transparency and Deliberation within the FOMC: A Computational Linquistics Approach." Working Paper. Available at http://eprints.lse.ac.uk/58072/.

Hare, Christopher, David A. Armstrong II, Ryan Bakker Royce Carroll and Keith T. Poole. 2015. "Using Bayesian Aldrich-McKelvey Scaling to Study Citizens' Ideological Preferences and Perceptions." *American Journal of Political Science* 59(3):759–774.

Hastie, Trevor, Robert Tibshirani and Jerome Friedman. 2013. *The Elements of Statistical Learning.* 10 ed. New York: Springer-Verlag.

Hobbs, William. 2017. "Pivoted Text Scaling for Open-Ended Survey Responses." Unpublished manuscript.

Hoff, Peter D. 2007. "Extending the Rank Likelihood for Semiparametric Copula Estimation." *The Annals of Applied Statistics* 1(1):265–283.

Jackman, Simon and Shawn Trier. 2008. "Democracy as a Latent Variable." *American Journal of Political Science* 52(1):201–17.

Jacoby, William G. and David A. Armstrong II. 2014. "Bootstrap Confidence Regions for Multidimensional Scaling Solutions." *American Journal of Political Science* 58(1):264–278.

Jessee, Stephen. 2016. "(How) Can We Estimate the Ideology of Citizens and Political Elites on the Same Scale?" *American Journal of Political Science* Forthcoming.

Keele, Luke, Corrine McConnaughy and Ismail White. 2012. "Strengthening the Experimenter's Toolbox: Statistical Estimation of Internal Validity." *American Journal of Political Science* 56(2):484–499.

Kim, In Song, John Londregan and Marc Ratkovic. 2018. "Estimating Spatial Preferences from Votes and Text." *Political Analysis.* .

Klami, Arto, Seppo Virtanen and Samuel Kaski. 2013. "Bayesian Canonical Correlation Analysis." *Journal of Machine Learning Research* 14(965–1003).

Krause, George A. 1994. "Federal Reserve policy decision making: Political and bureaucratic influences." *American Journal of Political Science* pp. 124–144.

Ladha, Krishna. 1991. "A Spatial Model of Leglslative Voting with Perceptual Error." *Public Choice* 68:151–74.

Lauderdale, Benjamin and Tom Clark. 2014. "Scaling Politically Meaningful Dimensions Using Texts and Votes." *American Journal of Political Science* Forthcoming.

Lipset, Seymour Martin. 1959. "Some Social Requisites of Democracy: Economic Development and Political Legitimacy." *American Political Science Review* 53(1):69–105.

McGregor, R.R. 1996. "FOMC voting behavior and electoral cycles: Partisan ideology and partisan loyalty." *Economics and Politics* 8(1):17–32.

Meade, Ellen and David Stasavage. 2008. "Publicity of Debate and the Incentive to Dissent: Evidence from the US Federal Reserve." *Economic Journal* 118:695–717.

Meltzer, Allan. 2004. *A History of the Federal Reserve, Volume 1*. University of Chicago Press.

Meltzer, Allan. 2010. *A History of the Federal Reserve, Volume 2*. University of Chicago Press.

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

Murray, Jared S., David B. Dunson, Lawrence Carin and Joseph E. Lucas. 2013. "Bayesian Gaussian Copula Factor Models for Mixed Data." *Journal of the American Statistical Association* 108(502).

Poole, Keith and Howard Rosenthal. 1997. *Congress: A Political Economic History of Roll Call Voting*. New York: Oxford University Press.

Poole, Keith T. 2005. *Spatial Models of Parliamentary Voting*. Analytical Methods for Social Research Cambridge: Cambridge University Press.

Poole, Keith T. and Howard Rosenthal. 1985. "A Spatial Model for Legislative Roll Call Analysis." *American Journal of Political Science* 29:357–84.

Quinn, Kevin M. 2004. "Bayesian Factor Analysis for Mixed Ordinal and Continuous Responses." *Political Analysis* 12(4):338–353.

Riboni, A. and F. Ruge-Marcia. 2010. "Monetary Policy by Committee: Consensus, Chairman Dominance or Simple Majority?" *Quarterly Journal of Economics* 125(1):363–416.

Roberts, Molly, Brandon Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Gadarian, Bethany Albertson and David Rand. 2014. "Structural topic models for open ended survey responses." *American Journal of Political Science* 58:1064–1082.

Schonhardt-Bailey, Cheryl. 2003. *Deliberating American monetary policy: a textual analysis.* The MIT Press.

Shor, Boris and Nolan McCarty. 2011. "The Ideological Mapping of American Legislatures." *American Political Science Review* 105(03):530–551.

Tausanovitch, Chris and Christopher Warshaw. 2013. "Measuring Constituent Policy Preferences in Congress, State Legislatures, and Cities." *The Journal of Politics* 75(02):330–342.

Tipping, Michael E. and Christopher M. Bishop. 1999. "Probabilistic Prncipal Component Analysis." *Journal of the Royal Statistcal Society, Series B* 61(3):611–622.

Tucker, Ledyard R. 1958. "An Inter-Battery Method of Factor Analysis." *Psychometrika* 23(2):111–136.

Abstract: 171 Words

Body of Paper: 6645 Words