Bayesian Analysis (2013) **8**, Number 1, pp. 233–268

Spatio-Temporal Modeling of Legislation and Votes

Eric Wang *, Esther Salazar †, David Dunson ‡ and Lawrence Carin §

Abstract. A model is presented for analysis of multivariate binary data with spatio-temporal dependencies, and applied to congressional roll call data from the United States House of Representatives and Senate. The model considers each legislator's constituency (location), the congressional session (time) of each vote, and the details (text) of each piece of legislation. The model can predict votes of new legislation from only text, while imposing smooth temporal evolution of legislator latent features, and correlation of legislators with adjacent constituencies. Additionally, the model estimates the number of latent dimensions required to represent the data. A Gibbs sampler is developed for posterior inference. The model is demonstrated as an exploratory tool of legislation and it performs well in quantitative comparisons to a traditional ideal-point model.

Keywords: Factor analysis, Indian buffet process, latent Dirichlet allocation, political science, topic modeling

1 Introduction

Quantitative analysis of legislative processes can help researchers better understand the underlying factors at work in legislative bodies. *Roll call data* collects the votes of legislators on bills and constitutes useful behavioral data political scientists may use to uncover the political tendencies and leanings of politicians. Most approaches to modeling roll call data rely on latent factor models called *ideal point models* (Poole and Rosenthal 1985; Clinton et al. 2004b). Roll call data are attractive for study because the results from model-based analysis can be compared to real-world intuitions. Additionally, time-stamps of votes, spatial location of legislators, and associated text of legislation are often available to augment the roll call data, offering opportunities for interesting spatio-temporal modeling extensions.

Traditionally, the modeling of roll call data has been split into two distinct tasks: exploratory analysis to uncover latent factors, and confirmatory analysis to test certain assumptions and hypotheses about the data given the latent factors (Hahn et al. 2012). However, these models are limited in vote prediction, since they can only fill in randomly missing votes, an ability with limited utility to political scientists. A much more useful ability would be to predict *all* votes of unseen or new legislation, based upon

*Department of Mathematics, United States Naval Academy, Annapolis, MD wang@usna.edu
†Department of Electrical & Computer Engineering, Duke University, Durham, NC esther.salazar@duke.edu
‡Statistical Sciences Department, Duke University, Durham, NC dunson@stats.duke.edu
§Department of Electrical & Computer Engineering, Duke University, Durham, NC lcarin@ece.duke.edu

 DOI:10.1214/13-BA810

the associated text legislation. Recently, Wang et al. (2010) addressed this problem by proposing a joint latent factor topic model that considers both roll call data and its associated legislative text. This powerful new paradigm in roll call analysis opens new possibilities for political scientists, such as making and testing predictions on how entire legislative bodies will vote on a new piece of legislation, and helping to better understand the nature of the latent factors by associating them with topics.

Before Wang et al. (2010), the analyses of text and roll call data were considered two separate lines of research. The most common tool for the analysis of text corpora are *topic models*. While latent factor models and ideal point models have been well established for several decades, the modeling of documents via topic modeling, especially *probabilistic topic models*, is relatively new (Wallach 2008). Early models for text, such as *latent semantic indexing* (LSI) (Deerwester et al. 1990), were primarily used for document retrieval, whereas most modern topic models are probabilistic and are primarily used to analyze document corpora (Steyvers and Griffiths 2007). The reason that probabilistic topic models allow intuitive analysis of documents is that they treat documents as finite mixtures of topics, where topics are discrete distributions over words. Another advantage probabilistic topic models have over models like LSI is that they incur significantly less parameter growth as the number of documents increases. This makes them well suited for the analysis of large datasets (Wallach 2008). Perhaps the best known probabilistic topic model is *latent Dirichlet allocation* (LDA) (Blei et al. 2003). LDA and its many derivatives have been applied to analyze political documents, including studying the press releases of senators to analyze the way they communicate with their constituencies (Grimmer 2010), time-dependent analysis of of Senate speeches (Quinn et al. 2006), and analysis of State of the Union addresses (Pruteanu-Malinici et al. 2010).

Another direction that has received relatively limited interest in roll call analysis via ideal point models is modeling spatio-temporal dependencies. It has been shown that significant gains in predictive performance on unseen data can be realized when model parameters are assumed to be drawn dependent on one another, either temporally (Quinn et al. 2006; Pruteanu-Malinici et al. 2010) or spatially (Dunson and Park 2008). Despite this, most ideal point models assume both spatial independence and temporal stationarity of legislator latent features. However, legislators' views typically evolve gradually over time. Moreover, regardless of party affiliation they are shaped by the region they represent (*e.g.*, in the current U.S. Congress, there are significant differences between the more liberal "Northern" Democrat and a moderate "Southern" or "Blue-Dog" Democrat, even though they caucus together). A model that imposes both spatial correlation and smooth temporal evolution of legislator latent feature information over time has greater utility as an exploratory tool of roll call data. Considering these dependencies can also improve a model's predictive ability on unseen data, since it will tighten the generally diffuse posterior distributions of roll call analysis (Jackman 2001).

In addition to lacking spatial and temporal dependencies, most ideal point models employ a single latent feature (Jackman 2001; Clinton et al. 2004b; Gerrish and Blei 2011). Single-dimensional ideal point models are generally able to capture political leanings (liberal/conservative bias) of the legislators, but are unable to describe local geographic effects unrelated to party affiliation. Some latent factor models have adopted shrinkage priors (Salakhutdinov and Mnih 2008) or max-margin constructions (Srebo et al. 2005) for controlling model complexity; however, they do not yield truly *sparse* solutions. Wang et al. (2010) used the beta process (BP) coupled with the Bernoulli process (Paisley and Carin 2009) to select the active latent feature dimensions via a latent binary vector. This construction simultaneously infers the number of latent feature dimensions while favoring a simple solution.

In this paper a model is proposed for jointly modeling time- and location-stamped text and roll call data. The proposed model assumes that the latent features of legislators are generated from a random-walk process in time, as in Wang et al. (2010), with a graphical prior on the legislator precision matrix (Dawid and Lauritzen 1993); this imposes that legislators from adjoining constituencies are likely dependent. A regression from the parameters of the topic model to the legislation latent feature is used to jointly model the legislation from the perspective of both text and roll calls. The proposed model is applied to time- and location-stamped roll call and corresponding legislative text data from the United States Senate and House of Representatives. The utility of the model is demonstrated not only as an exploratory tool of the Senate and the House (as in Poole and Rosenthal (1985) and Clinton et al. (2004b)) but also as a tool for predicting votes on unseen legislation (as in Wang et al. (2010) and Gerrish and Blei (2011)). Finally, the effects of including spatio-temporal dependencies are studied, comparing the performance of the model in the prediction task to models that assume independence of legislations in time and/or space.

The remainder of the paper is organized as follows. In Section 2, related and past work are reviewed in topic modeling and roll call analysis. Section 3 presents the proposed model, and details its individual components. An inference algorithm is developed in which local conjugacy allows analytic conditional distributions and therefore simple Gibbs sampling, as discussed in Section 4. Section 5 presents experimental results, and conclusions are provided in Section 6.

2 Related Work

2.1 Roll Call Analysis via Ideal Point Models

Most modern statistical analyses of roll call data are related to ideal point models (Poole and Rosenthal 1985; Clinton et al. 2004b). In such models, each legislator i is associated with an *ideal point* $\boldsymbol{x}_i \in \mathbb{R}^K$ and each bill j is associated with a latent position $\boldsymbol{y}_j \in \mathbb{R}^K$, although in practice K is often set to one for computational considerations (Jackman 2001). In order to link the continuous valued random processes $\boldsymbol{x}_i$ and $\boldsymbol{y}_j$ to the observed binary votes, a link function is used. Poole and Rosenthal (1985) used the logistic link function while Clinton et al. (2004b) adopted a probit link function; the probit model is employed in this paper.

The probit ideal point model is summarized as follows. Consider a binary matrix $\mathbf{C} \in \{0,1\}^{N_l \times M}$, where N_l is the number of legislators and M is the number of bills voted on. $\mathbf{C}$ is assumed to be generated from an underlying matrix $\boldsymbol{S} \in \mathbb{R}^{N_l \times M}$ through a probit model (Chib and Greenberg 1998) with random effects β_j for $j \in \{1, \ldots, M\}$

$$c_{ij} = \begin{cases} 1 & \text{if } s_{ij} > 0 \\ -1 & \text{if } s_{ij} \leqslant 0 \end{cases} \tag{1}$$

where $i \in \{1, \ldots, N_l\}$,

$$s_{ij} = \boldsymbol{x}_i^T \boldsymbol{y}_j + \beta_j + \epsilon_{ij}, \tag{2}$$

and $\epsilon_{ij} \sim \mathcal{N}(0, 1)$. The relationship of legislator latent features $\boldsymbol{x}_i \in \mathbb{R}^K$ and legislation latent features $\boldsymbol{y}_j \in \mathbb{R}^K$ determines the voting behavior of legislator i on bill j. If $\boldsymbol{x}_i$ is proximate to $\boldsymbol{y}_j$ in latent space (and both are some distance away from the origin), then s_{ij} will likely be a large positive number, encouraging $c_{ij} = 1$. If $\boldsymbol{x}_i$ and $\boldsymbol{y}_j$ are distant from each other and the origin, s_{ij} will likely be highly negative, encouraging $c_{ij} = -1$.

The random effects β_j are useful in modeling bills with a large fraction of unanimous "Yea" votes (corresponding to large and positive β_j); these votes may mask more partisan characteristics of the legislators. Such votes are generally procedural in nature and noninformative regarding the political leanings of the legislators, and are not well explained by $\boldsymbol{x}_i$ and $\boldsymbol{y}_j$. In these cases β_j must be a relatively large positive value to explain the affirmative votes (β_j is large and negative for votes that are unanimously "Nay", although they are rare). If such popular votes are not pruned either by random effects or manually as in earlier works such as Poole and Rosenthal (1985), they can significantly skew the analysis, resulting in noticeably degraded interpretability of the latent features. In previous work (Wang et al. 2010; Jackman 2001), a zero mean Gaussian prior was placed on β_j. This construction works well in analysis of legislation, but does not allow for effective prediction of random effects of unseen legislation.

2.2 Topic Models

Consider a corpus of documents $\boldsymbol{d}_j$, $j = 1, \ldots, J$ where each $\boldsymbol{d}_j$ is a V_j dimensional vector of order-exchangeable word indices d_{vj}, $v = 1, \ldots, V_j$, and each d_{vj} points to a word in a globally defined dictionary of size W. In topic models, each document is modeled as a mixture of H discrete *topics*, where each topic is a specialized discrete distribution with support on the global dictionary. Latent Dirichlet allocation (LDA) (Blei et al. 2003) is the most commonly used topic model. Hierarchically, LDA is represented as

$$\begin{aligned} d_{vj} &\sim \text{Mult}(1, \boldsymbol{\phi}_{z_{vj}}) \\ z_{vj} &\sim \text{Mult}(1, \boldsymbol{\theta}_j) \\ \boldsymbol{\theta}_j &\sim \text{Dir}(\zeta/H, \ldots, \zeta/H) \\ \boldsymbol{\phi}_h &\sim \text{Dir}(\rho/W, \ldots, \rho/W), \end{aligned} \tag{3}$$

where topic h is characterized by a distribution over words $\boldsymbol{\phi}_h$, and document j has a distribution $\boldsymbol{\theta}_j$ over the topics; ζ and ρ are Dirichlet concentration parameters, and $z_{vj} \in \{1, \ldots, H\}$ is an indicator assigning word v in document j to a topic. In the above,

$\text{Mult}(1, \boldsymbol{\phi})$ is meant to denote a single draw from a multinomial distribution with probability vector $\boldsymbol{\phi}$, and the associated indicator variable so drawn is associated with the non-zero component of this draw. The probability distribution $\text{Dir}(\zeta/H, \ldots, \zeta/H)$ corresponds to a Dirichlet distribution, with homogeneous hyperparameters ζ/H. Compared to previous optimization-based topic models, such as latent semantic analysis (Deerwester et al. 1990), LDA maintains a significantly smaller parameter set, is stochastic and has a locally conjugate construction that facilitates efficient posterior inference through variational approximation (Blei et al. 2003) or Gibbs sampling (Porteous et al. 2008).

2.3 Joint Analysis of Roll Call Data and Text

Wang et al. (2010) were the first to perform joint analysis of roll call data and the associated text legislation. The proposed approach used a mixture model to jointly cluster the legislation latent features $\boldsymbol{y}_j$ and document-dependent probability of topic usage $\boldsymbol{\theta}_j$. Each piece of legislation is associated with a cluster m, and each cluster is jointly characterized by a K dimensional Gaussian distribution on legislation latent features with mean $\boldsymbol{\mu}_m$ and covariance matrix $\boldsymbol{\Sigma}_m^{-1}$, as well as a discrete distribution over topics $\boldsymbol{\tau}_m$. If legislation j is associated with cluster m, then its latent features are drawn as $\boldsymbol{y}_j \sim \mathcal{N}(\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m^{-1})$, and $\boldsymbol{\theta}_j = \boldsymbol{\tau}_m$. A Dirichlet process (DP) (Ferguson 1973) prior was placed on the mixture model in order to infer the appropriate number of clusters. The primary novelty of this model is that votes could be predicted on new legislation based solely on the legislative text by first determining a bill's cluster association via the topic model, and then mapping the bill to a latent feature fixed at the associated cluster center.

In addition, Wang et al. (2010) also introduced a simple time-series construction that allowed the latent features of the legislators to "drift" smoothly in latent space over time, and inferred the dimension of the latent space through the beta process (BP) coupled with the Bernoulli process (Paisley and Carin 2009). Let $t = \{1, \ldots, T\}$ denote the time stamps, where here t indicates a congressional session. Hierarchically, their construction for the legislator latent features is

$$\begin{aligned} \boldsymbol{x}_{it} &= \boldsymbol{b} \odot \hat{\boldsymbol{x}}_{it} \\ \hat{\boldsymbol{x}}_{it} &\sim \mathcal{N}(\hat{\boldsymbol{x}}_{i,t-1}, \sigma^{-1}\mathbf{I}_K) \\ b_k &\sim \text{Bernoulli}(\pi_k) \\ \pi_k &\sim \text{Beta}(e/K, f(K-1)/K), \end{aligned} \tag{4}$$

where $b_k \in \{0, 1\}$ and $\boldsymbol{b} = (b_1, \ldots, b_K)^T$ is a K-dimensional sparse binary latent vector, $\hat{\boldsymbol{x}}_{it} \in \mathbb{R}^K$ is a latent position of legislator i in session t, σ is a precision term, $\odot$ denotes a Hadamard or point-wise product and $\mathbf{I}_K$ is a $K \times K$ identity matrix. Notice that the mean of $\hat{\boldsymbol{x}}_{it}$ is legislator i's latent feature in the previous session $t-1$, and that $\boldsymbol{b}$ controls the dimensionality of the latent space by setting $b_k = 0$ for unused dimensions.

The model proposed in Wang et al. (2010) worked well and was novel in its application, but the link between the factor model and topic model (via the joint clustering of legislation latent features $\boldsymbol{y}_j$ and the document-dependent probability of topic usage $\boldsymbol{\theta}_j$) was weak, and the factor model numerically dominated the topic model. This resulted in a model that was primarily driven by the roll call data and not particularly influenced by the text documents. Moreover, the clustering structure imposed that all legislation sharing the same cluster m had the *same* distribution over topics $\boldsymbol{\tau}_m$, reducing model flexibility. Finally, since new legislation was mapped only to its associated cluster mean in latent space, careful attention had to be paid to the setting of the hyperparameters to favor a mixture model with many compact Gaussian clusters in latent space rather than a single large Gaussian.

Gerrish and Blei (2011) replaced the mixture model construction by proposing several models that incorporated text regression to join the topic and factor models. Recall that Gerrish and Blei (2011) only used a single factor, and therefore vector $\boldsymbol{y}_j$ becomes scalar y_j; they modeled the latent features of bill j as

$$y_j \sim \mathcal{N}(\boldsymbol{\eta}^T \bar{\boldsymbol{z}}_j, \gamma_{(\hat{y})}^{-1}), \tag{5}$$

where $\boldsymbol{\eta} \in \mathbb{R}^H$ are regression weights, $\gamma_{(\hat{\boldsymbol{y}})}$ is a precision term, and

$$\bar{z}_{jh} = \frac{\hat{z}_{jh}}{\sum_{h'=1}^{H} \hat{z}_{jh'}} \tag{6}$$

is the *fraction* of words in bill j drawn from topic h, where $\hat{z}_{jh} = \sum_v \mathbf{1}(z_{vj} = h)$ defines the *total* number of words in bill j drawn from topic h, and $\mathbf{1}(\cdot) = 1$ if the argument is true, and it is zero otherwise. The regression weights $\boldsymbol{\eta}$ were fit without regularization. In this paper (5) is extended to multiple latent dimensions, and the appropriate number of latent dimensions is inferred.

Besides being able to predict votes for unseen bills from text, another important aspect of Wang et al. (2010) and Gerrish and Blei (2011) is the inclusion of *random effects* terms that act as intercepts and give the model significantly improved robustness to unanimous votes. Previous work such as Poole and Rosenthal (1985) and Clinton et al. (2004b) required the pruning of bills with unanimous votes from the dataset. Wang et al. (2010) assigned zero mean Gaussian priors to the random effects, while Gerrish and Blei (2011) proposed regressing to β_j in a manner similar to how it was done for y_j, giving their paper the unique ability to predict the contentiousness of new legislation.

We note that, to the authors' knowledge, there has not been previous work within the political science statistics literature on utilizing the text of the legislation when performing roll-call analysis. However, Jackman (2009) shows that a hierarchical model over legislator preferences does tighten up the inferences with respect to the ideal points; he also makes the point that because of the bilinear form of the model, prior information about bill parameters are informative about ideal points, and vice-versa. Therefore, while the text of legislation has not been used previously by political scientists in roll-call studies, the potential value of such has been understood. It appears that the principal reason the traditional ideal-point model (Poole and Rosenthal 1985; Clinton et al. 2004b) has been retained (without modeling text) was not because of a lack of recognition of the potential value of incorporating information like text/legislation, but rather because of a desire to stay close to the "rational actor" nature of such traditional models. There has recently been increasing interest within the political science community on topic models (Grimmer 2010; Grimmer and Stewart to appear in 2013).

3 Model Specification

3.1 General Structure

Suppose that now the binary matrices $\mathbf{C}_t$ are time-stamped by $t \in \{1, \ldots, T\}$, and for each time t we also have an $N_l \times N_l$ binary adjacency matrix $\boldsymbol{G}_t$ that encodes the spatial adjacency of the legislators. The legislators and edges in $\boldsymbol{G}_t$ define a graph where two legislators i and i' are connected ($g_{ii't} = 1$) if their congressional districts (House of Representatives) or states (Senate) share a common border. All elements along the main diagonal of $\boldsymbol{G}_t$ are one. Finally, for each piece of legislation j in session t, we have a document $\boldsymbol{d}_{jt}$ defined as in (3).

The underlying matrix $\boldsymbol{S}_t$ is generated as in (2)

$$s_{ijt} = \boldsymbol{x}_{it}^T \boldsymbol{y}_{jt} + \alpha_{it} + \beta_{jt} + \epsilon_{ijt}, \tag{7}$$

but a time indicator t is added and an additional per-legislator random effect α_{it} is employed for symmetry (although in practice this term can be eliminated). The following sections discuss each term in (7) in detail.

3.2 Imposing Spatio-Temporal Dependencies on the Legislators

Let $\hat{\boldsymbol{x}}_{it} \in \mathbb{R}^K$ correspond to the latent features of legislator i at time t, and let $\hat{\boldsymbol{x}}_{(k)t} \in \mathbb{R}^{N_l}$ represent the kth feature vector for all N_l legislators at time t. In order to account for both temporal and spatial dependencies in the legislator latent features $\hat{\boldsymbol{x}}_{it}$, the second expression of (4) is modified as

$$\hat{\boldsymbol{x}}_{(k)t} \sim \mathcal{N}(\hat{\boldsymbol{x}}_{(k),t-1}, \boldsymbol{\Omega}_t), \tag{8}$$

where the mean of $\hat{\boldsymbol{x}}_{(k)t}$, $\hat{\boldsymbol{x}}_{(k),t-1}$, is the kth latent feature of all legislators from time $t-1$, $\hat{\boldsymbol{x}}_{(k),0}$ is a vector of zeros with dimension N_l and $\boldsymbol{\Omega}_t \in \mathbb{R}^{N_l \times N_l}$ is a covariance matrix drawn from a Hyper-Inverse-Wishart (HIW) distribution (Dawid and Lauritzen 1993). A key property of $\boldsymbol{\Omega}_t$ is that its inverse, $\boldsymbol{\Omega}_t^{-1}$, is a sparse precision matrix with non-zero elements at the positions where links exist in $\boldsymbol{G}_t$. Specifically,

$$\boldsymbol{\Omega}_t \sim \mathrm{HIW}_{\boldsymbol{G}_t}(\kappa, \boldsymbol{\Omega}_0), \tag{9}$$

where $\boldsymbol{G}_t$ is the adjacency matrix described in Section 3.1, $\boldsymbol{\Omega}_0^{-1} = \mathbf{I}_{N_l}$, and κ is a scale term set to $10 + N_l$. The model has been found to be insensitive to the setting of κ (many "reasonable" values of κ yielded similar results). The HIW provides a computationally convenient method of generating sparse precision matrices, such that two entries are *conditionally* independent if they are not connected in $\boldsymbol{G}_t$ (however, while the precision matrix is sparse, the covariance matrix is generally not, implying the opportunity for correlation between legislators from distant regions, as desired). Note that in this construction the absolute locations of the legislators are discarded. The legislators take the role of nodes in a graph whose edges are defined by the adjacency of their constituencies. For inference, the efficient sampling scheme proposed in Carvalho et al. (2007) is employed.

It is desirable to impose identifiability in the latent factor model. Previous work fixed the ideal points of several well known legislators to impose identifiability (Poole and Rosenthal 1985; Clinton et al. 2004b; Gerrish and Blei 2011). However, assigning ideal points in higher dimensions can still lead to rotationally non-identified solutions (Jackman 2001). Here a block lower-triangular construction is imposed as,

$$\begin{aligned} \hat{x}_{ik't} &= 0 \text{ for } k' > i \\ \hat{x}_{iit} &= 1 \text{ for } i = 1, \ldots, K. \end{aligned} \tag{10}$$

Imposing identifiability is important in obtaining interpretable latent features. Note that when $K = 1$, this approach to imposing identifiability is equivalent to setting the first legislator's ideal point to 1. As the number of dimensions increases, the lower block-triangular structure of our model imposes rotational identifiability. Note that our experiments also indicated insensitivity to the ordering of the legislators, embedded in index i.

3.3 Controlling Model Complexity

In Clinton et al. (2004b) and Gerrish and Blei (2011), the dimension of latent vectors $\boldsymbol{x}_{it}$ and $\boldsymbol{y}_{jt}$ was set to one. However, the existence of additional dimensions has received mention even in early literature (Poole and Rosenthal 1985). A generative process is developed that simultaneously infers the number of latent features and imposes a parsimonious model:

$$\begin{aligned} \boldsymbol{x}_{it} &= \boldsymbol{b} \odot \hat{\boldsymbol{x}}_{it} \\ \boldsymbol{y}_{jt} &= \boldsymbol{b} \odot \hat{\boldsymbol{y}}_{jt} \end{aligned} \tag{11}$$

where, as in (4), $\hat{\boldsymbol{x}}_{it} \in \mathbb{R}^K$ and $\hat{\boldsymbol{y}}_{jt} \in \mathbb{R}^K$ are real vectors, $\boldsymbol{b}$ is a sparse binary vector, and $\odot$ denotes a Hadamard or point-wise product. In (11), $\boldsymbol{b}$ is used to define both $\boldsymbol{x}_{it}$ and $\boldsymbol{y}_{jt}$ for clarity, while in practice it is only necessary to attach $\boldsymbol{b}$ to either $\boldsymbol{x}_{it}$ or $\boldsymbol{y}_{jt}$. This construction is attractive because unlike the maximum margin optimization of Srebo et al. (2005) or shrinkage priors of Salakhutdinov and Mnih (2008), it assumes a sparse generative process on $\boldsymbol{x}_{it}$ and $\boldsymbol{y}_{jt}$ that sets unused dimensions to *exactly* zero.

In order to generate a sparse binary vector $\boldsymbol{b}$, a truncated beta process (BP) prior is placed on $\boldsymbol{b}$ in a manner similar to Paisley and Carin (2009) as in (4) where we set K to be a large integer. For most "reasonable" settings of e and f, this prior favors a sparse $\boldsymbol{b}$ (most of its values set to zero). By integrating out the π_k, $k = 1, \ldots, K$, one can show that the expected number of ones in $\boldsymbol{b}$ is $eK/[e + f(K-1)]$, and that the total number of non-zero components in $\boldsymbol{b}$ is drawn from Binomial$(K, eK/[e + f(K-1)])$. This construction represents a finite approximation to the Indian Buffet Process (Griffiths and Ghahramani 2005), but has the advantage that the samples $\boldsymbol{b}$ can be block-sampled from its posterior.

3.4 Topic Regression to Latent Space

The legislation latent features $\hat{\boldsymbol{y}}_{jt}$ are given an informative prior whose mean is constructed via a regression that maps from the normalized topic frequency to the latent features. Since there are multiple dimensions, a regression model similar to (5) is needed for each latent dimension, and is written as

$$\hat{y}_{jkt} \sim \mathcal{N}(\boldsymbol{\eta}_k^T \bar{\boldsymbol{z}}_{jt} + \eta_{k0}, \gamma^{-1}_{(\hat{\boldsymbol{y}}_k)}), \tag{12}$$

where intercept term η_{k0} is introduced, and $\bar{\boldsymbol{z}}_{jt}$ is defined as in (6), except with the addition of the time indicator t.

The regression weights $\boldsymbol{\eta}_k \in \mathbb{R}^H$ and the corresponding offset term η_{k0} have their own respective Gaussian priors

$$\begin{aligned} \boldsymbol{\eta}_k &\sim \mathcal{N}(0, \gamma^{-1}_{(\boldsymbol{\eta}_k)} \mathbf{I}_H) \\ \eta_{k0} &\sim \mathcal{N}(0, \gamma^{-1}_{(\eta_{k0})}). \end{aligned} \tag{13}$$

Note that the parameters for topic regression for latent dimension k — $\boldsymbol{\eta}_k$, η_{k0} and the predictor $\hat{y}_{jkt}$ — are constructed independently from the other latent dimensions. Diffuse Gamma$(10^{-3}, 10^{-3})$ priors are assigned for $\gamma_{(\boldsymbol{y})}$, $\gamma_{(\boldsymbol{\eta}_k)}$ and $\gamma_{(\eta_{k0})}$.

A drawback of linking the topic model and the factor model through this construction is significantly increased computational complexity in the inference of the per-word topic indicators z_{vjt}. This is especially pronounced when modeling large datasets such as the House of Representatives. To ease the computational burden of the model in these situations, it is also possible to perform topic modeling on the legislative text as a separate stand-alone step before the vote matrices (after which each document is characterized by a distribution over topics, with which the aforementioned regression is performed; the distribution over topics may be viewed as a covariate for the matrix analysis). This can be done by first running LDA as described in (3) and fixing the topic assignments z_{vjt} (and correspondingly $\hat{\boldsymbol{z}}_{jt}$ and $\bar{\boldsymbol{z}}_{jt}$) for example based on the maximum-likelihood sample, before running the rest of the model as discussed in this section. Doing so decouples the inference of the topic model from the factor model. Inference of the topic model parameters can then be performed via high-speed methods (Porteous et al. 2008). The model that estimates the z_{vjt} *in situ* with the rest of the model is referred as the "one-step" model and the model that performs topic modeling before the factor model is termed the "two-step" model.

3.5 Legislation and Legislator Random Effects

Two random effect terms are considered in the model, α_{it} and β_{jt}: α_{it} is a legislator-specific random effect that impacts the probability of legislator i votes "Yea" or "Nay", and β_{jt} is a legislation specific random effect that reflects the "difficulty" of the vote. If $|\beta_{jt}|$ is relatively large, then all legislators are more apt to vote one way or the other, while if $|\beta_{jt}|$ is small, then the inner product of the latent features $\boldsymbol{x}_{it}^T \boldsymbol{y}_{jt}$ strongly impacts the votes. Random effects allow the factor model (defined by $\boldsymbol{x}_{it}$ and $\boldsymbol{y}_{jt}$) to impact only those contentious pieces of legislation that are highly informative. We find that the random effects for legislation plays an important role in predicting votes, while the random effects for legislators are relatively small, since typically no legislator votes "Yea" or "Nay" constantly.

The prior for the legislator random effect α_{it} is zero mean Gaussian

$$\alpha_{it} \sim \mathcal{N}(0, \gamma_{(\alpha)}^{-1}); \tag{14}$$

however, because β_{jt} is connected to the legislation, it is drawn from a Gaussian distribution whose mean is constructed as a regression with $\bar{\boldsymbol{z}}_{jt}$ as covariates,

$$\begin{aligned} \beta_{jt} &\sim \mathcal{N}(\boldsymbol{\lambda}^T \bar{\boldsymbol{z}}_{jt} + \lambda_0, \gamma_{(\beta)}^{-1}) \\ \boldsymbol{\lambda} &\sim \mathcal{N}(0, \gamma_{(\boldsymbol{\lambda})}^{-1} \mathbf{I}_H) \\ \lambda_0 &\sim \mathcal{N}(0, \gamma_{(\lambda_0)}^{-1}), \end{aligned} \tag{15}$$

where $\gamma_{(\beta_{jt})}$, $\gamma_{(\boldsymbol{\lambda})}$ and $\gamma_{(\lambda_0)}$ are all given diffuse Gamma$(10^{-3}, 10^{-3})$ priors and $\boldsymbol{\lambda} \in \mathbb{R}^H$ and λ_0 denote the random effect regression weights and intercept, respectively. This construction of β_{jt} is attractive because it allows for accurate prediction of the legislation difficulty from text.

An advantage of the regression construction for β_{jt} and $\hat{\boldsymbol{y}}_{jt}$ is the ease of interpretability of the results for the legislation latent features, particularly with respect to β_{jt} and the primary dimension of $\hat{\boldsymbol{y}}_{jt}$. This is illustrated with two examples. In the following let h and h' denote different topics and k denote the dominant latent dimension (largest variance). Consider a topic with index h that is generally popular. This type of legislation is generally strongly approved by Congress, and will have a relatively positive associated random effect regression weight λ_h. A document that uses topic h heavily will be encouraged to have a large positive random effect *a priori*. By contrast, a contentious topic h' will have a $\lambda_{h'}$ that is relatively close to 0, and an $\eta_{h'k}$ that is distant from 0. Since the dominant dimension generally captures liberal/conservative affiliation (Wang et al. 2010), a large (positive or negative) $\eta_{h'k}$ can indicate a topic is strongly affiliated with one political ideology (conservative or liberal) and a document that has significant mass in topic h' will then tend to have a small random effect and significant displacement in the primary dimension. Notice that while the topics themselves do not reside in latent space, studying how they map into latent space via their regression weights can give useful insight into the meaning of the latent dimensions.

3.6 Model Summary

The proposed model has several components, and therefore it is desirable to provide a concise summary, as depicted in Figure 1. In this diagram V_{jt} represents the number of words in document/legislation j in year t, J_t represents the number of pieces of legislation in year t, and N_l represents the number of legislators. In this "plate" diagram, the number of draws of a given type is indicated by the number associated with a given rectangle/plate. The words, denoted by d, and the votes, denoted by c, are observed, and the remaining parameters are random variables, with statistics to be inferred.

At left in Figure 1 is a relatively standard Bayesian topic model (Blei et al. 2003). The topics employed, defined by indicator variable z, are used via regression to represent the legislation-dependent random effect β as well as the legislation-dependent feature vector $\boldsymbol{y}$. Each legislator has a latent feature vector $\boldsymbol{x}$, which is connected to his/her ideal point. Via $\boldsymbol{x}$ and $\boldsymbol{y}$, and random effects α and β, a probit model is employed to represent the generative process for the vote. Each legislator also has an associated random effect α (for model symmetry, although in most cases this will be negligible). The dimension of the latent space in which legislators and legislation reside is defined by the number of non-zero components in the binary feature vector $\boldsymbol{b}$.

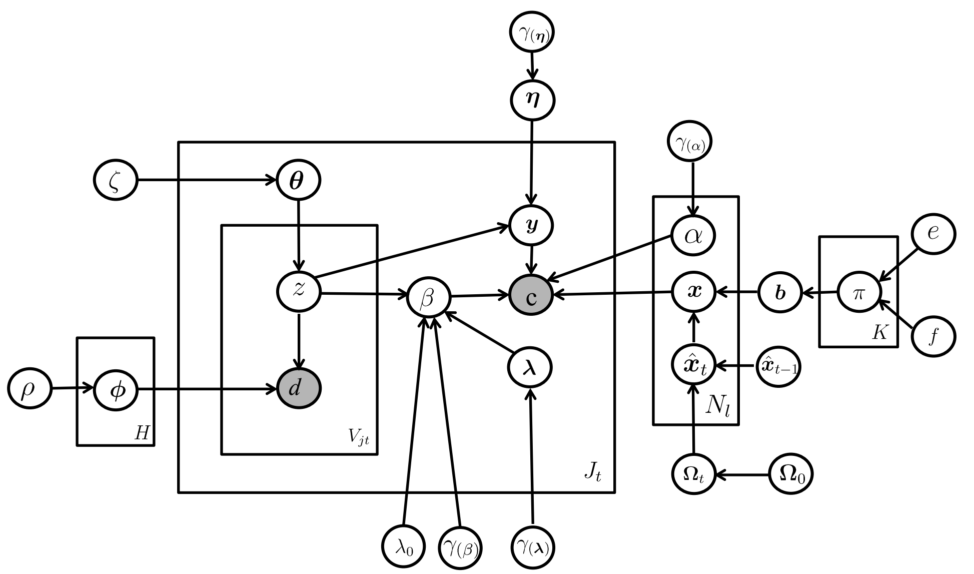


Figure 1: Graphical representation of the model.

3.7 Special Cases

In terms of constructing the latent features of the legislators $\hat{\boldsymbol{x}}_{it}$ and legislation $\hat{\boldsymbol{y}}_{jt}$, the model can be shown to be a generalization of previous Bayesian matrix factorization models. Specifically, it can be shown that the inclusion of the topic model via regression can collapse back to Gaussian priors when the text documents are non-informative or when no text is available. Additionally, the spatial effects imposed by the HIW distribution on the covariance matrix of the legislators can be shown to also be a generalization of the construction found in Wang et al. (2010)

In the hypothetical case of a completely non-informative topic model for legislation j, assume that $\bar{\boldsymbol{z}}_{jt}$ is a uniform discrete distribution of dimension H. In the limit as the number of topics $H \to \infty$, the elements of $\bar{\boldsymbol{z}}_{jt}$ approach 0. Thus, regardless of the regression weight vectors $\boldsymbol{\eta}_k$ or $\boldsymbol{\lambda}$, the contribution of the topic model to the mean of $\hat{\boldsymbol{y}}_{jt}$ is zero. This leads to the prior on $\hat{\boldsymbol{y}}_{jt}$ to be

$$\hat{y}_{jkt} \sim \mathcal{N}(\eta_{k0}, \gamma^{-1}_{(\hat{\boldsymbol{y}}_k)}), \tag{16}$$

for all k, where η_{k0} is the intercept term for latent dimension k. Likewise, the prior on the random effect β_{jt} in this situation is

$$\beta_{jt} \sim \mathcal{N}(\lambda_0, \gamma^{-1}_{(\beta)}). \tag{17}$$

In the case where a piece of legislation has no text associated with it, (6) is indeterminate due to division by 0 in its denominator. It is therefore defined that when no text is present $\bar{\boldsymbol{z}}_{jh} = 0 \ \forall \ h$. This definition also leads to the prior on $\hat{\boldsymbol{y}}_{jt}$ and β_{jt} shown in (16) and (17), respectively.

The proposed model is a generalization of many previous Bayesian matrix factorization methods but with a non-zero mean prior. Notice, however, that if $\eta_{k0} = \lambda_0 = 0$ for all k, then we arrive at the widely used zero mean Gaussian priors on both $\hat{y}_{jkt}$ and β_{jt} (Clinton et al. 2004b; Salakhutdinov and Mnih 2008).

Furthermore, if the graph of the legislators $\boldsymbol{G}_t$ is set to $\mathbf{I}_{N_l}$, then both the precision matrix $\boldsymbol{\Omega}_t^{-1}$ and thus the covariance matrix $\boldsymbol{\Omega}_t$ of the legislator latent features become diagonal. This allows the construction on the kth latent feature of legislator i at time t, $\hat{x}_{ikt}$, to be written as

$$\hat{x}_{ikt} = \mathcal{N}(\hat{x}_{ik,(t-1)}, \omega_{iit}) \tag{18}$$

where ω_{iit} is the ith diagonal element of $\boldsymbol{\Omega}_t$; (18) is identical to the simple random-walk process on legislator latent features in Wang et al. (2010).

4 Posterior Inference

A Gibbs sampler is employed for posterior inference (Wang et al. 2010; Jackman 2001; Clinton et al. 2004b). Let

$$\boldsymbol{\Psi}_t = \{\boldsymbol{S}_t, \hat{\boldsymbol{X}}_t, \boldsymbol{\Omega}_t, \hat{\boldsymbol{Y}}_t, \boldsymbol{\alpha}_t, \boldsymbol{\beta}_t, \boldsymbol{z}_{jt}, \boldsymbol{\theta}_{jt}, \} \tag{19}$$

be the set of model parameters of the model specific to session t and let $\mathbf{\Gamma}$ denote the set of all specified Gaussian precision parameters for all sessions. The goal of posterior inference is to find the posterior

$$P(\{\mathbf{\Psi}_t\}_{t=1:T}, \{\boldsymbol{\eta}_k, \eta_{0k}\}_{k=1:K}, \{\boldsymbol{\phi}_h\}_{h=1:H}, \boldsymbol{\lambda}, \lambda_0, \boldsymbol{b}, \boldsymbol{\pi}, \mathbf{\Gamma} | \{\boldsymbol{C}_t, \boldsymbol{G}_t, \boldsymbol{D}_t\}_{t=1:T}), \tag{20}$$

where $\boldsymbol{C}_t$ is the vote matrix, $\boldsymbol{G}_t$ is the graph of the legislators, and $\boldsymbol{D}_t$ is the collection of documents, all for session t. The topic model is as described hierarchically in (3), with the addition of the time indicator t on d_{vjt}, z_{vjt}, and $\boldsymbol{\theta}_{jt}$. The hierarchical representation of the factor model is shown below:

$$\begin{aligned}
c_{ijt} &= \begin{cases} 1 & \text{if } s_{ijt} > 0 \\ -1 & \text{if } s_{ijt} \leqslant 0 \end{cases} \\
s_{ijt} &= \boldsymbol{x}_{it}^T \boldsymbol{y}_{jt} + \alpha_{it} + \beta_{jt} + \epsilon_{ijt} \\
\boldsymbol{x}_{it} &= \boldsymbol{b} \odot \hat{\boldsymbol{x}}_{it}, \qquad \hat{\boldsymbol{x}}_{it} = \{\hat{x}_{ikt}\}_{k=1:K} \\
\boldsymbol{y}_{jt} &= \boldsymbol{b} \odot \hat{\boldsymbol{y}}_{jt}, \qquad \hat{\boldsymbol{y}}_{jt} = \{\hat{y}_{jkt}\}_{k=1:K} \\
b_k | \pi_k &\sim \text{Bernoulli}(\pi_k), \qquad \pi_k \sim \text{Beta}(e/K, f(K-1)/K) \\
\hat{\boldsymbol{x}}_{(k)t} | \hat{\boldsymbol{x}}_{(k),t-1}, \mathbf{\Omega}_t &\sim \mathcal{N}(\hat{\boldsymbol{x}}_{(k),t-1}, \mathbf{\Omega}_t), \qquad \mathbf{\Omega}_t \sim \text{HIW}_{\boldsymbol{G}_t}(\kappa, \mathbf{\Omega}_0) \\
\hat{y}_{jkt} | \boldsymbol{\eta}_k, \eta_{k0}, \gamma_{(\hat{\boldsymbol{y}}_k)}, \bar{\boldsymbol{z}}_{jt} &\sim \mathcal{N}(\boldsymbol{\eta}_k^T \bar{\boldsymbol{z}}_{jt} + \eta_{k0}, \gamma_{(\hat{\boldsymbol{y}}_k)}^{-1}) \\
\beta_{jt} | \boldsymbol{\lambda}, \lambda_0, \gamma_{(\beta)}, \bar{\boldsymbol{z}}_{jt} &\sim \mathcal{N}(\boldsymbol{\lambda}^T \bar{\boldsymbol{z}}_{jt} + \lambda_0, \gamma_{(\beta)}^{-1}) \\
\boldsymbol{\eta}_k | \gamma_{(\boldsymbol{\eta}_k)} &\sim \mathcal{N}(0, \gamma_{(\boldsymbol{\eta}_k)}^{-1} \mathbf{I}_H), \qquad \eta_{k0} | \gamma_{(\eta_{k0})} \sim \mathcal{N}(0, \gamma_{(\eta_{k0})}^{-1}) \\
\boldsymbol{\lambda} | \gamma_{(\boldsymbol{\lambda})} &\sim \mathcal{N}(0, \gamma_{(\boldsymbol{\lambda})}^{-1} \mathbf{I}_H), \qquad \lambda_0 | \eta_{(\lambda_0)} \sim \mathcal{N}(0, \gamma_{(\lambda_0)}^{-1}) \\
\alpha_{it} | \gamma_{(\alpha)} &\sim \mathcal{N}(0, \gamma_{(\alpha)}^{-1}) \\
\epsilon_{ijt} &\sim \mathcal{N}(0, 1).
\end{aligned} \tag{21}$$

Gamma hyperpriors are placed on all the gamma precision terms in the normal distributions (*e.g.*, on $\gamma_{(\beta)}$), and the $\bar{\boldsymbol{z}}_{jt}$ are linked to the topic model, as in (3). The most interesting (not widely used) Gibbs parameter updates are presented in the Appendix, specifically those having to do with $\hat{\boldsymbol{x}}_{(k)t}$, $\mathbf{\Omega}_t$, $\hat{\boldsymbol{y}}_{jt}$, β_{jt}, and z_{vjt}.

5 Experimental Results

The model is applied for joint analysis of votes and documents in both the House of Representatives and Senate. The 103rd to 107th House of Representatives (1993 to 2003) and the 106th to 111th Senate (1999 to 2011) are considered. The two houses were considered independently, with the "one-step" model used to model the Senate dataset, and the "two-step" model used to model the significantly larger House of Representatives dataset. For both the House and Senate, bills with multiple rounds of voting were pruned to a single bill since the additional rounds are usually on minor changes. The Senate dataset contains 416 bills, and the House dataset contains 1260 bills. For each congressional session t, there is a graph $\boldsymbol{G}_t$ that encodes the adjacency of districts or states.

Each bill's text was processed into n-grams, resulting in a vocabulary size of 4743 n-grams, with n ranging from 1 to 5. This means that the individual tokens in the vocabulary can be more than single words; they can be informative phrases of up to 5 words. For more information on the specific processing of the text, see Gerrish and Blei (2011). Unless specified otherwise, all Gaussian precision parameters have $\text{Gamma}(10^{-3}, 10^{-3})$ priors. We initialized each experiment by setting the latent position in all dimensions of Republican legislators to -1, Democrats to 1 and Independents to 0. The model was run with this initialization and with random initialization, and the effect on the quantitative prediction results were negligible; however, the specified initialization yielded slightly improved and more interpretable qualitative results. Each experiment was run for 5000 iterations for each experiment, with the first 2500 iterations discarded as burn-in. We set the BP parameters $e = f = 1$, and truncated the number of latent dimensions to $K = 20$. The number of topics H was truncated as an upper bound to 32, and both Dirichlet concentration parameters were set $\xi = \zeta = 1$.

5.1 Legislators in Latent Space

The model learned $||\boldsymbol{b}||_0 = 3$ dimensions for the Senate and $||\boldsymbol{b}||_0 = 4$ for the House of Representatives. In both cases, a single dimension had significantly higher variance than the others. This dimension is termed the dominant dimension, and the dimension with the second largest variance is termed the secondary dimension. In Figure 2 the Senators are shown in the first 2 latent dimensions, from 1999 to 2011, with Democrats in blue, Republicans in red, and Independents in black.

The dominant dimension in latent space seems to encode the party affiliation of the senators. This result agrees with Gerrish and Blei (2011) and Poole and Rosenthal (1985), although unlike them the model does not require strong party members fixed to opposite ends of latent space. From 2001 to 2005, in the wake of the September 11, 2001 attacks, the Senate was apparently more bipartisan than in later years. From 2005 onward, there is generally greater party divide between the Republicans and Democrats. This is consistent with the increased partisanship that has been observed by political researchers (Hahn et al. 2012).

In Figure 3, a selection of prominent senators is highlighted and their position in the dominant latent dimension is examined over time. We separate the senators by party affiliation with Democrats (Clinton, Obama, Kennedy, Kerry, Edwards, Feingold) in blue, Independents (Lieberman) in black and Republicans (McCain, McConnell, Santorum) in red. The darker colored lines show the party mean for the Democrats (blue) and Republicans (red) and the error bars denote the standard deviation. These senators are examined because they are prominent enough to have made serious runs at the presidency, or — as in the case of McConnell, Feingold and Kennedy — are well known party stalwarts.

In addition to providing interesting insight into congressional voting patterns, the model offers a temporal evolving analysis of legislators. For example, considering John McCain, due to his opposition to some early President Bush era policies (after 2000), his

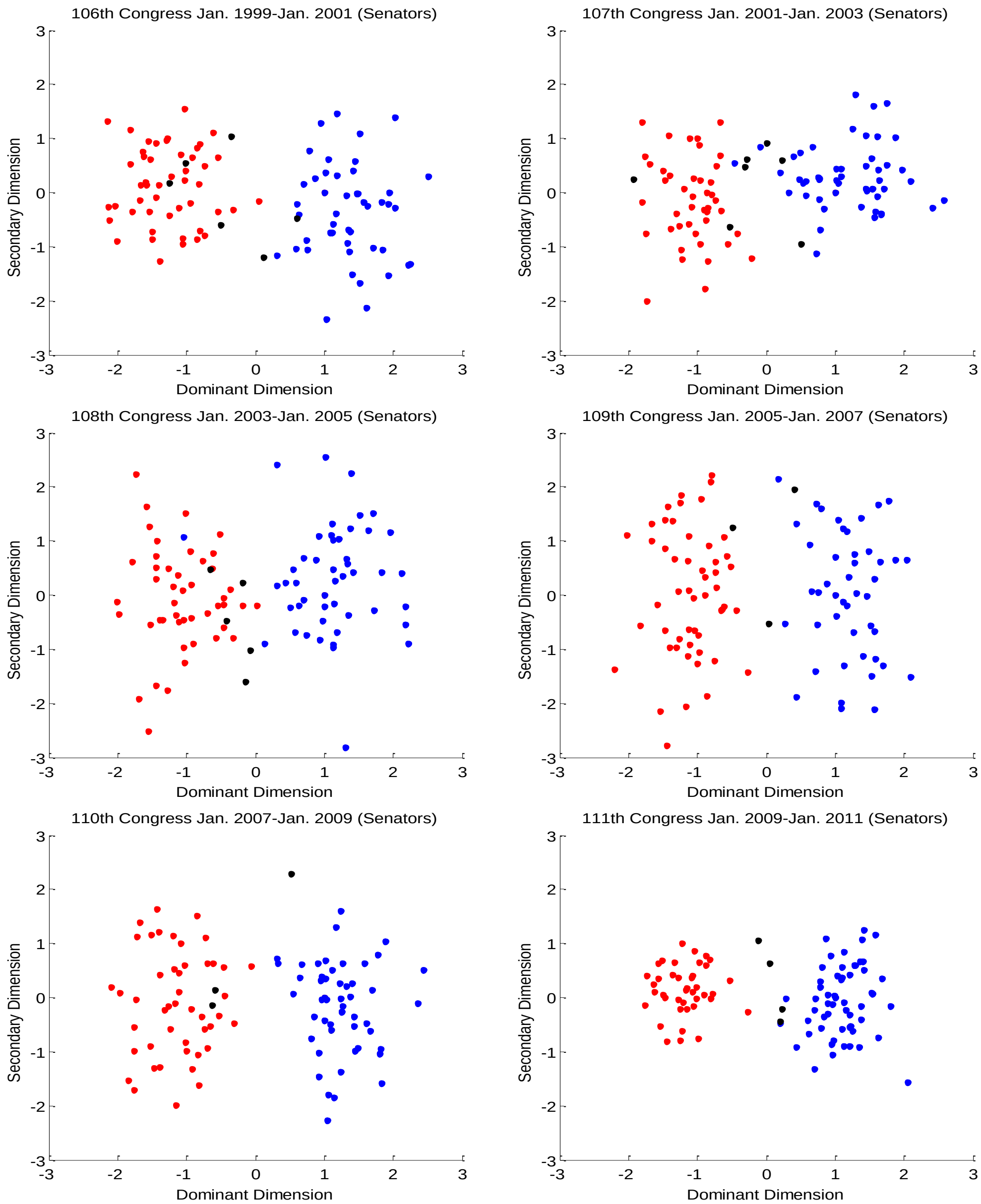


Figure 2: Latent positions of senators in the dominant and secondary latent dimensions from 1999 to 2011. Republicans are shown in red, Democrats are shown in blue and Independents are shown in black.

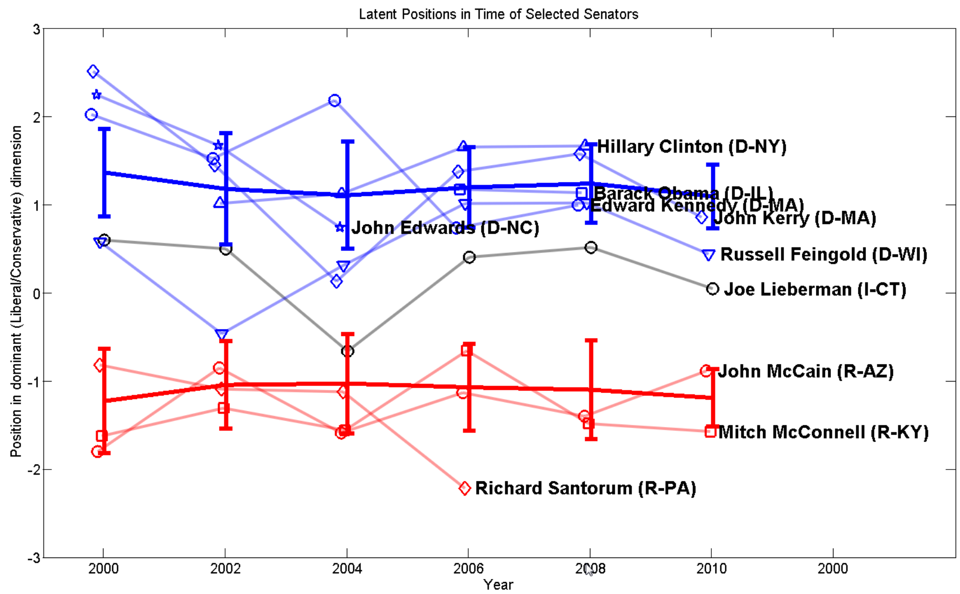


Figure 3: Trajectories in the dominant dimension of prominent senators. Republicans are shown in red, Democrats are shown in blue and Independents are shown in black. Party means for the Democrats and Republicans are shown in darker lines.

latent position became more liberal in 2002. In 2004 he supported the war in Iraq and made a distinct movement toward the right that session. One Senate session earlier, in 2002, McCain and Feingold co-sponsored the McCain-Feingold campaign finance reform act (also known as the *Bipartisan Campaign Finance Reform Act*), and this is represented primarily by Feingold's movement toward conservatism and by McCain's small shift toward the center.

The model also captured interesting movement during the 2004 Presidential election. Joe Lieberman ran for the Democratic nomination on a "hawkish" platform, reflected in his conservative position in that year. Also in 2004, both John Kerry and John Edwards made dramatic centrist shifts from previously highly liberal positions, presumably in order to garner a wider swath of voters. Kerry moved back to his previous highly liberal position after the 2004 election.

In Figure 4, the latent positions of Representatives is shown in the dominant and secondary dimensions over time (1993-2003). The House of Representatives is generally considered more partisan than the Senate, and Representatives usually vote along party lines. This is manifested as a distinct separation along party lines in the dominant dimension. Additionally, Representatives are here characterized by greater tightness in latent feature space (relative to senators) of their latent positions within their party clusters.

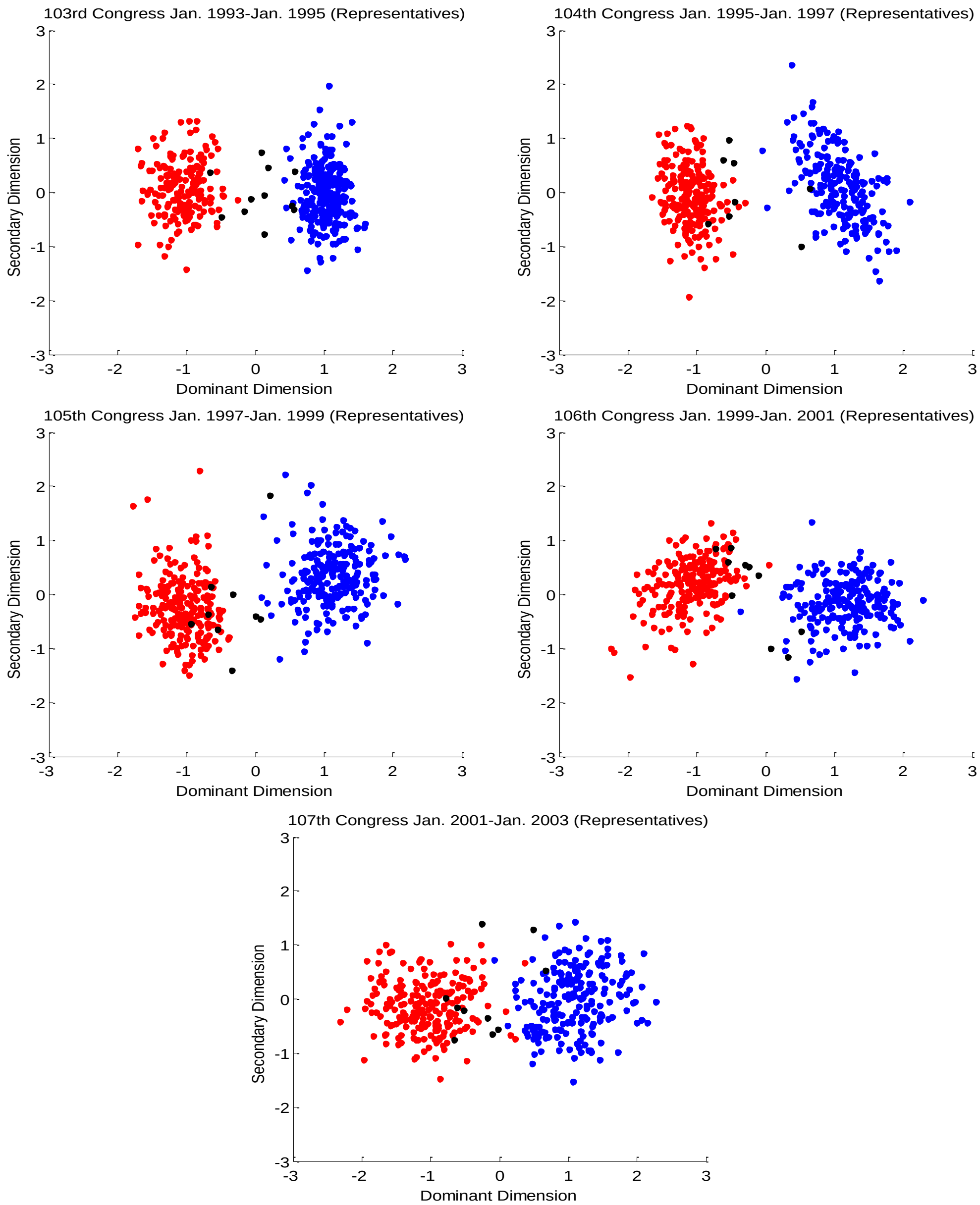


Figure 4: Latent positions of Representatives in the dominant and secondary latent dimensions. Republicans are shown in red, Democrats are shown in blue and Independents are shown in black.

5.2 Topics and Latent Space

The relationship between topics and latent space is examined. As described in Section 3.5, topic h is associated with K regression coefficients for latent space $\boldsymbol{\eta}_h = \{\eta_{h1}, \ldots, \eta_{hK}\}$, and a regression coefficient for the random effect η_{h0}. Topics with large regression weights in a latent dimension are highly influential in that dimension. Similarly, topics with large positive random effect weights tend to be those for which all legislators vote "Yea". Analyzing the regression weights of the topics along with the topics themselves can provide useful insight to better interpret the latent space.

In Figures 5 and 6 are shown the topics for the Senate and their regression weights in latent space plotted against their difficulty regression weights, respectively. Figures 7 and 8 show the topics and regression weights for the House of Representatives, respectively. The sizes of the dots represent the relative usage of the topics and shown are only those topics containing more than 1% of the corpus. In the dominant dimensions of both Figures 6 and 8 the Republicans are to the left and the Democrats are to the right.

Several topics are highlighted in the Senate latent space. Senate Topic 2 is by far the most popular topic in the Senate dataset and describes Department of Defense (DoD) acquisitions. The years we considered (2000 to 2010) saw Congress allocate significant increases in military expenditures due to the war in the Middle East and the fight against terrorism. The increases in military spending generally had bipartisan support during this period, thus explaining the relatively large positive difficult regression weight. This topic is slightly conservative, although the large random effect regression weight makes its position in latent space relatively inconsequential.

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
debtor	dod	student	election	juvenile
bankruptcy	defense	school	iraq	school
chapter	subtitle	crop	ats	offense
debt	military	eligible	people	crime
court	defense and appropriation	teacher	americans	firearm
district of columbia	army	loan	judge	attorney
creditor	officer	producer	voting	criminal
trustee	contract	institution	attorney	court
claim	personnel	indian	voter	child
property	acquisition	technology	justice	youth

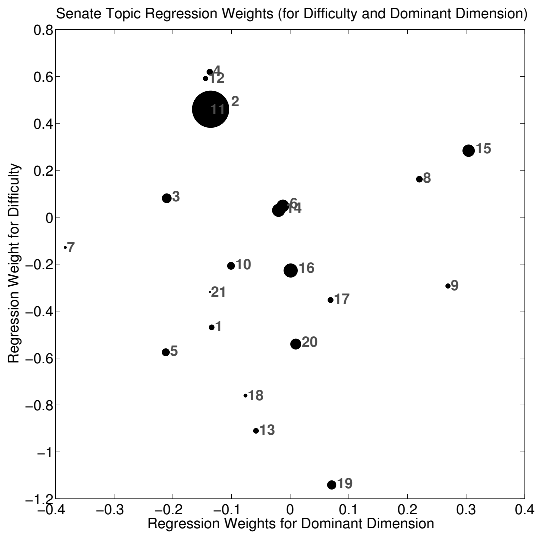
Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
transfer	surveillance	drug	energy	medicare
defense	court	patent	pipeline	coverage
intelligence	electronic	response	price	beneficiary
expense	foreign	site	commodity	prescription drug
head	procedure	prescription drug	market	benefit
september	terrorist	application	foreign	drug
account	record	subparagraph	oil	premium
remain available	foreign intelligence	clause	trading	offer
construction	intelligence	pediatric	trade	provider
remain available until expend	disclosure	applicant	bills	eligible

Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
international	board	class	transportation	food
country	loan	defendant	highway	bills
export	corporation	iraq	safety	agent
foreign	loan guarantee	bills	rail	audit
palestinian	air carrier	propose	motor vehicle	food safety
israel	flood	court	homeland security	agriculture
foreign service	applicant	plaintiff	vehicle	auditor
read	matter	settlement	fuel	foreign
aid	travel	claim	develop	fee
united nations	financial	responsibility	passenger	article

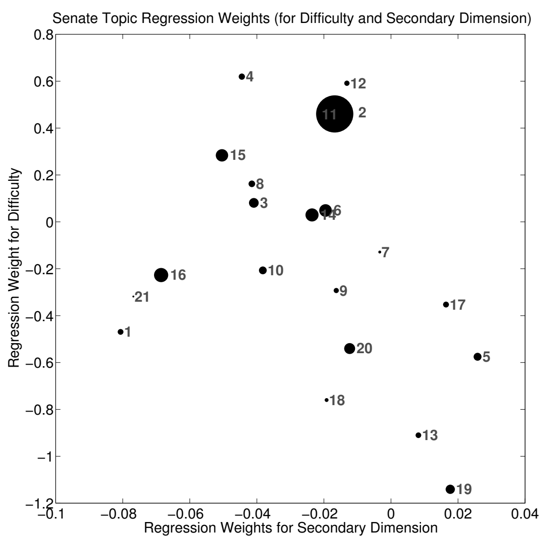
Topic 16	Topic 17	Topic 18	Topic 19	Topic 20
tax	relief	legislation	health care	alien
credit	article	effect	damage	status
income	import	budget	claimant	immigration
property	duty	budgetary	lawsuit	visa
qualified	material	outlay	liability	application
corporation	country	budget authority	product	removal
foreign	rate	chairman	claim	child
taxpayer	determination	adjustment	party	border
fuel	territory	billion	loss	employer
tax credit	apparel	estimate	injury	employment

Topic 21
insured
loss
chapter
terrorism
insurance
participate
insurance company
property
read
term mean

Figure 5: Topics for the Senate, 106th to 111th Congress (principal words depicted).



(a)



(b)

Figure 6: Regression weights for the Senate, 106th to 111th Congress. Each topic h plotted with the (a) regression weights for difficulty λ_h (vertical axis) vs. regression weights for dominant latent feature k, η_{hk} (horizontal axis) (b) Regression weights for difficulty λ_h (vertical axis) vs. regression weights for secondary latent feature k', $\eta_{hk'}$ (horizontal axis). The number labels correspond to Senate topics and the size of the dots indicates the mass of the topic listed in Figure 5.

Another interesting topic with bipartisan support is Topic 4, pertaining to the Iraq conflict. Although this topic could potentially be controversial and contentious, closer examination reveals that most of the bills using this topic are primarily bills dealing with honoring fallen American soldiers, and condemning the treatment of American prisoners of war.

By contrast, Topic 7, dealing with foreign surveillance and sensitive documents, is much more contentious. Most of the bills using this topic seek to amend the *Foreign Surveillance Act* of 1978. This topic is the most conservative topic in the Senate. One of the most liberal topics is Topic 8, describing prescription drug benefits. Interestingly, Topic 19, tort reform, receives little support from both parties. Topic 9 is the second most liberal topic we found and is primarily used by bills seeking to amend the *Commodity Exchange Act* to prevent excessive price speculation.

The House of Representatives corpus used more topics than the Senate, reflecting the larger variety of bills considered by the lower assembly. The most popular topic in the House of Representatives is Topic 4, having to do with Department of Defense acquisitions. Like its counterpart in the Senate, it has a large random effect regression weight, and generally enjoyed bipartisan support. Other noncontroversial issues that pass almost unanimously include Topic 24 (supporting and commending humanitarian and peacekeeping missions abroad), Topic 16 (Coast Guard activities and regulation), Topic 6 (National Science Foundation funding), Topic 20 (military construction funding), and Topic 9 (hydroelectric power funding allocation).

Perhaps the most striking topic in our analysis of the House of Representatives is also one of the smallest. Topic 18 is primarily used to describe the *Partial-Birth Abortion Ban Act* and its many amendments. While the bill itself had near unanimous Republican and limited Democratic support, it was the subsequent Democrat supported initiatives to remove or repeal certain parts of the act that dominate the topic's use and cause the topic to display significant liberal bias

House Topic 14 is primarily used to describe the *Medicare Prescription Drug, Improvement, and Modernization Act* of 2003. This Republican backed bill was very hard fought in both the Senate and the House, with multiple members of Congress changing their votes at the last minute. The bill had opposition and support from both parties, explaining its small random effect weight and its slightly conservative position in the dominant dimension. However, this topic has a relatively large movement away from the origin in the secondary dimension, a sign that factors other than party affiliation are at work.

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
following	officer	risk	defense	loan	research
subparagraph	expense	regulatory	subtitle	agriculture	technology
add	library	analysis	dod	agricultural	director
clause	appoint	document	military	food	network
read	flight	permit	transfer	research	science
chapter	congressional	major	army	rural	sea
mean	center	risk assessment	procurement	subtitle	organ
thereof	travel	significant	navy	producer	computer
notice	legislative branch	propose	personnel	purchase	develop
recovery	board	administrator	contract	crop	national science foundation

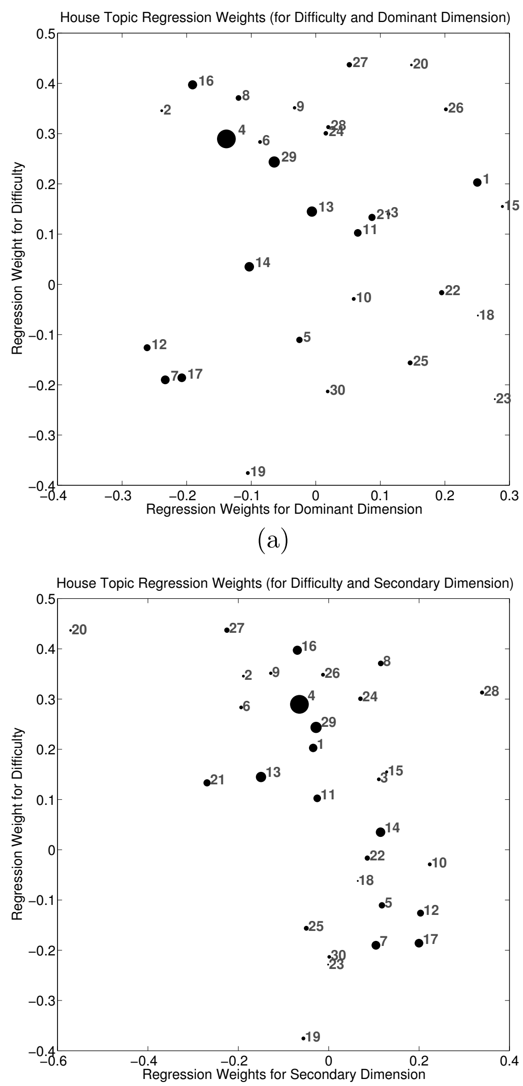
Topic 7	Topic 8	Topic 9	Topic 10	Topic 11	Topic 12
election	bank	energy	debtor	school	tax
contribution	housing	water	bankruptcy	student	subtitle
intelligence	subtitle	expense	debt	teacher	contribution
foreign	financial	county	property	educational	credit
candidate	public housing	construction	chapter	child	property
transfer	insurance	river	consumer	eligible	qualified
officer	board	district	trustee	local educational	income
immigration	consumer	flood control	claim	charter school	account
official	mortgage	remain available until expend	contract	performance	corporation
personnel	transaction	power	court	application	trust

Topic 13	Topic 14	Topic 15	Topic 16	Topic 17	Topic 18
land	medicare	loan	airport	district	abortion
national park	benefit	business	transportation	district of columbia	chapter
resource	subtitle	board	vessel	chapter	violation
park	coverage	company	safety	emergency	protection
property	care	loan guarantee	coast guard	treasury	minor
county	hospital	guarantee	commercial	limitation	procedure
boundary	ssa title	labor	international	requires	implement
river	health care	administrator	administrator	services	life
national forest	child	party	highway	corporation	perform
secretary of the interior	family	rate	aircraft	transfer	employment

Topic 19	Topic 20	Topic 21	Topic 22	Topic 23	Topic 24
budget	defense	country	claim	june	people
budget authority	construction	international	property	medal	nato
level	military construction	china	court	joint resolution	afghanistan
outlay	transfer	foreign	class	honor	taiwan
legislation	september	human	loss	constitution	effort
outlays	appropriations	export	defendant	hundred	sudan
revenue	family housing	transfer	party	session	peace
deficit	acts	aid	damage	american	united nations
reform	operation and maintenance	people s republic of china	product	calendar no congress session	international
surplus	acquisition	debt	application	memorial	force

Topic 25	Topic 26	Topic 27	Topic 28	Topic 29	Topic 30
trade	tax	child	carrier	offense	print
country	return	drug	rate	attorney	motion
import	taxpayer	adoption	joint resolution	crime	question
duty	rate	family	satellite	court	read
africa	benefit	substance abuse	station	juvenile	clause
export	subparagraph	parent	passenger	victim	chairman
product	disclosure	treatment	rail	criminal	debate
article	account	coalition	television	subtitle	waive
customs service	railroad retirement	research	network	alien	vote
trade agreement	december	training	access	child	adopt

Figure 7: Topics for the House of Representatives, 103rd to 107th Congress (principal words depicted).



(a)

(b)

Figure 8: Regression weights for the House of Representatives, 103rd to 107th Congress. Each topic h plotted with the (a) regression weights for difficulty λ_h (vertical axis) vs. regression weights for dominant latent feature k, η_{hk} (horizontal axis). (b) Regression weights for difficulty λ_h (vertical axis) vs. regression weights for secondary latent feature k', $\eta_{hk'}$ (horizontal axis). The number labels correspond to House topics and the size of the dots indicates the mass of the topic listed in Figure 7.

It should be noted that the direct association of topics with ideological positions must be interpreted carefully. Above, Senate Topic 7 (relating to defense and intelligence) is deemed the most conservative topic; however, it is not difficult to imagine bills that relate to defense and intelligence from either a liberal or conservative perspective. This is true for any of the topics listed in above: one might create or eliminate restrictions on abortion (House Topic 18), raise or lower taxes (House Topic 26), etc. The fact that the defense and intelligence topic is associated with bills supported by conservative legislators is a contingent fact about the political climate of the mid-2000s; it might be completely different in another era (like the post-Watergate 1970s). Other work on ideology combines content-based topics with additional word distributions covering ideological perspectives (Ahmed and Xing 2010). Such an approach might allow one to maintain the part of the model which associates controversiality with topics, while modeling ideology separately.

5.3 Inferred Random Effects

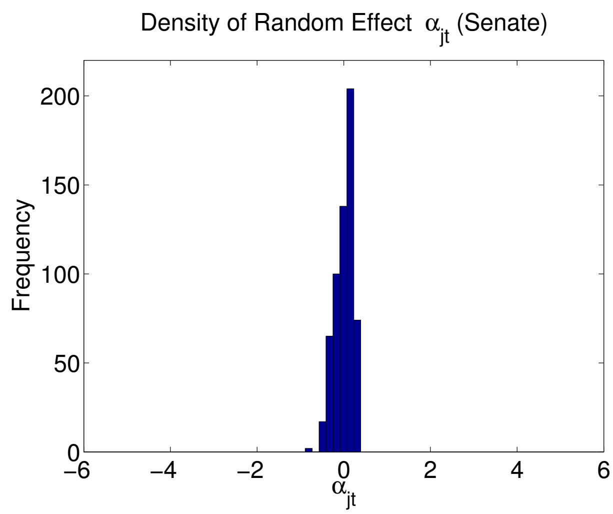
The random effects in the model play a vital role by helping to explain unanimous votes. Although the per-legislator random effect α_{it} is not strictly necessary, and is often not included (see *e.g.*, Gerrish and Blei (2011) and Jackman (2001)), we include it for symmetry.

In Figure 9 are shown the histograms of the random effects for the Senators and the Senate bills. Figure 9(a) shows that the Senator random effects are closely clustered around 0 with small variance. This is repeated for the Representatives' random effects histogram shown in Figure 10(a).

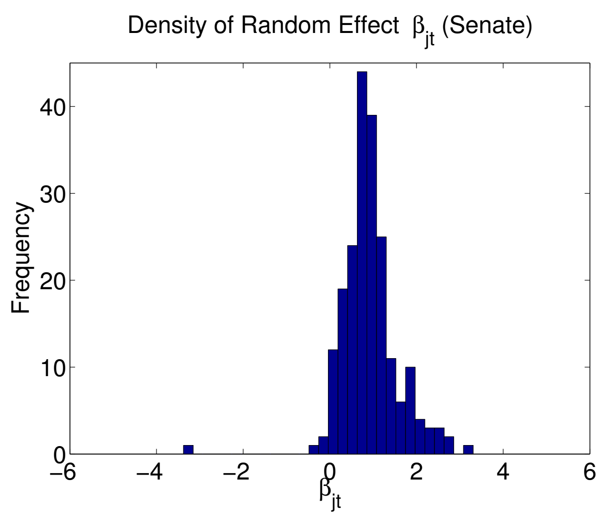
The Senate legislation random effects histogram in Figure 9(b) has a peak near 0.7 and is significantly tailed toward positive values, from the large number of near-unanimous "Yea" votes on procedural or non-contentious bills. In Figure 10(b) is shown the histogram of House bill random effects. An interesting feature of the House random effects is the bi-modal nature of the histogram with a peak near 0 and a peak near 1.8. The bi-modal nature of this histogram suggests that a significant portion of the bills in the House are procedural and non-contentious. Recall that a large per-bill random effect means that the latent features $\boldsymbol{x}_{it}$ and $\boldsymbol{y}_{jt}$ have little influence in the votes of the legislators, and indeed this appears to be true. In the House, representatives voted "Yea" 91% of the time on bills with $\beta_{jt} > 1$, and that number drops to 66% of the time for $0 < \beta_{jt} < 1$, showing that β_{jt} is small for contentious votes (where $\boldsymbol{x}_{it}$ and $\boldsymbol{y}_{jt}$ have significant impact). For reference, the percent of "Yea" votes across the House dataset is 74%.

5.4 Prediction of Held-Out Legislation

Like Gerrish and Blei (2011) and Wang et al. (2010), the model can predict votes on held out legislation. To do this, following Gerrish and Blei (2011), the data are partitioned



(a)

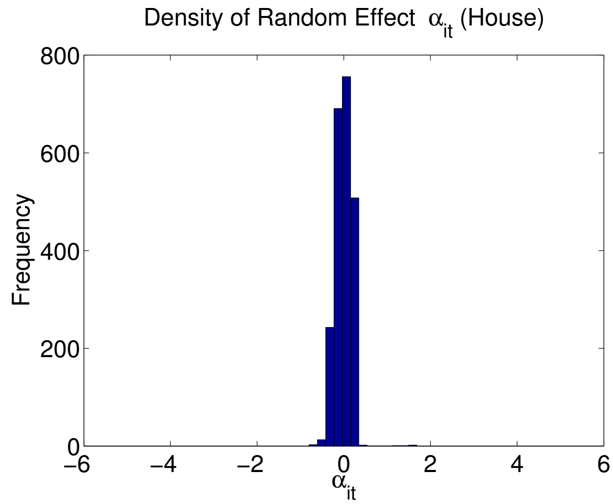


(b)

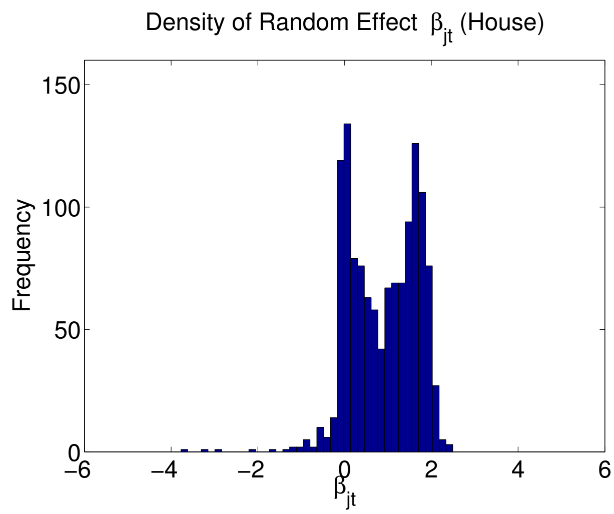
Figure 9: (a) Histogram of all Senator random effects α_{it}, (b) Histogram of all Senate bill random effects β_{jt}.

into 6 folds, removing whole columns of votes. Training is performed on the five folds and tested on the sixth, rotating through until all six folds are tested, predicting each piece of legislation once. The results are aggregated over all six folds.

The full model as presented in Section 3 is considered, as well as limited versions of the model with only spatial or temporal dependency and a model that considered no



(a)



(b)

Figure 10: (a) Histogram of all Representative random effects α_{it}, (b) Histogram of all House bill random effects β_{jt}.

spatio-temporal dependencies. To remove the spatial model the graph $\boldsymbol{G}_t$ is set to $\mathbf{I}_{N_l}$. To remove temporal dependencies, (4) is modified to

$$\hat{\boldsymbol{x}}_{it} \sim \mathcal{N}(0, \boldsymbol{\Omega}_t), \tag{22}$$

where $\boldsymbol{\Omega}_t$ is drawn as in (9). The full model is termed "Temporal+Graphical"; the model with only the random-walk process on the legislators is termed "Temporal"; the model with only the graphical model is termed "Graphical"; and the model with neither the random-walk nor graphical model is termed "No Dependencies".

For the Senate, a comparison is made to a traditional ideal point version of our model without spatio-temporal dependencies and the random effect α_{it} removed. To do this, the number of latent dimensions K is set to 1, and the latent position of Ted Kennedy is set to 1 and that of Mitch McConnell to -1. This ideal point model is similar to that proposed by Gerrish and Blei (2011), except that the model inference is done using a Gibbs sampler (rather than a variational Bayes approximation), and a probit link is used to generate the votes. We note that adopting a probit ideal point model is equivalent to using quadratic utility functions as in Clinton et al. (2004b). This model is denoted "Ideal Point Probit".

Probabilistic link functions, of which the probit model is a member, are able to yield a prediction confidence with the predicted vote. For each held-out vote by legislator i on legislation j, the model yields a probabilitiy of "Yea", $p(c_{ijt} = 1|s_{ijt})$, and a probability of "Nay", $1-p(c_{ijt} = 1|s_{ijt})$. Model confidence can be assessed by collecting $\max\{p(c_{ijt} = 1|s_{ijt}), 1 - p(c_{ijt} = 1|s_{ijt})\}$ for each held out vote, irrespective of whether the actual prediction is "Yea" or "Nay", and placing them into one of 5 probability bins, ranging from "highly unsure" $[0.5, 0.6)$ to "highly sure" $[0.9, 1)$.

Prediction confidence is a highly useful metric to compare different models, but it is even more important to examine whether the model performance actually matches the model confidence, *e.g.*, the votes in the $[0.6, 0.7)$ probability bin should ideally have an empirical probability of being correct between 0.6 and 0.7. In other words, if the model prediction confidence is in the bin $[0.6, 0.7)$, we expect that from 30 to 40% of the time the prediction will be wrong; we wish to examine this empirically. Any model with probabilistic predictions whose behavior follows such a fashion is, in this sense, correct; however, the *best* model is the one that puts the most predicted votes into the "highly confident" $[0.9, 1)$ bin.

Figure 11(a) shows a comparison of the models on the Senate data. The full "Temporal+Graphical" model performs the best of all the models studied, with the most votes in the [0.9,1) confidence bin. All models perform well, in that the empirical prediction performance matches the model prediction; the "Temporal+Graphical" is deemed best because it places the most votes in the highest-confidence bin. Both the "Graphical" and "Temporal" models outperform the "No Dependencies" model, although the advantage of the spatial information in the "Graphical" model is much smaller than that of the "Temporal" model. This is due to the relatively coarse state-level spatial dependency in the Senate data. The "Temporal" model performs almost equal to the full "Temporal+Graphical" model. This suggests that most of the gains of the full model are from smoothing the temporal evolution of the legislators. However, all of our models outperform the "Ideal Point Probit" model. This is primarily due to the additional information captured by considering multiple latent dimensions.

In Figure 11(b) the models are compared on the House of Representatives dataset. The "Graphical" model performs the best due to the fine-grain district-level spatial

	No Dependencies		Graphical		Temporal		Temporal+Graphical		Ideal Point Probit	
Probit Probability Bins	Votes in Confidence Bin	Empirical % Correct	Votes in Confidence Bin	Empirical % Correct	Votes in Confidence Bin	Empirical % Correct	Votes in Confidence Bin	Empirical % Correct	Votes in Confidence Bin	Empirical % Correct
0.5-0.6	1828	0.53	1610	0.56	1594	0.55	1508	0.54	2213	0.53
0.6-0.7	2234	0.65	1919	0.66	1711	0.66	1598	0.64	2290	0.64
0.7-0.8	2352	0.79	2356	0.79	1974	0.78	1902	0.77	2062	0.72
0.8-0.9	3074	0.88	3361	0.88	2871	0.88	2683	0.87	3692	0.81
0.9-1	7401	0.93	7654	0.93	8750	0.93	8920	0.93	6624	0.92

(a)

	No Dependencies		Graphical		Temporal		Temporal+Graphical	
Probit Probability Bins	Votes in Confidence Bin	Empirical % Correct	Votes in Confidence Bin	Empirical % Correct	Votes in Confidence Bin	Empirical % Correct	Votes in Confidence Bin	Empirical % Correct
0.5-0.6	76386	0.56	71292	0.56	72737	0.57	72533	0.56
0.6-0.7	77164	0.65	72354	0.66	73639	0.66	73897	0.66
0.7-0.8	86779	0.75	83387	0.77	83855	0.77	84010	0.77
0.8-0.9	129457	0.85	132453	0.86	133027	0.86	131305	0.86
0.9-1	189420	0.92	199719	0.93	195950	0.93	197461	0.93

(b)

Figure 11: Comparison of various models in prediction confidence of new legislation based on the associated text for the (a) Senate and (b) House of Representatives. For each model, the left column denotes the number of votes in each probability bin, and the right column is the empirical probability of correct prediction for that bin.

structure present in the House data. The "Temporal" model resulted in improved performance over the "No Dependencies" model but performed worse than the "Graphical" and the "Temporal+Graphical" model. Additionally, the "Temporal+Graphical" model performs slightly worse than the "Graphical" model. Our results suggest that the temporal evolution structure in the model may not be flexible enough to account for the rapidly evolving latent space, but that strong spatial effects are at work between the representatives.

6 Conclusions

A model has been developed for the joint analysis of multivariate binary data and text with spatio-temporal dependencies, and it has been applied to model congressional roll call data. The model can predict all votes of new legislation from only text, via a regression construction from a topic model to the factor model. Random effects are included to help the model handle non-informative unanimous (or near-unanimous) votes, and the legislation random effect can also be predicted from text. The appropriate number of latent features is inferred, and a temporal evolution of the legislators is learned, allowing for fine-grained temporal analysis of their political leanings. The model has utility as an exploratory tool of legislative space, being able to assign meaning to the latent dimensions, inferring the liberal/conservative bias of topics. Experimental results demonstrated the value of spatio-temporal dependencies in improving vote prediction performance compared to models that did not incorporate them.

While the model provides a flexible and powerful framework for the analysis of roll call data with text, its versatile framework allows additional opportunities for extensions. Dynamic topic models such as Pruteanu-Malinici et al. (2010) or Blei and Lafferty (2006) have been shown to offer gains in per-word prediction over the LDA used in our model. Incorporating such a topic model could offer improvements in vote prediction. Party effects could be incorporated into our model to improve the estimation of latent space, and a Gaussian Process (GP) could be used to replace the graphical model construction of the legislator latent feature covariance matrix $\mathbf{\Sigma}_t^{-1}$, with district/state locations as covariates, at the cost of increased computation complexity. A different direction would be to extend the random effects by assuming that they are drawn from different clusters with varying precisions, allowing such a model to cluster legislators or bills into different groups with different "volatilities".

Legislative data offers a rich body of textual data beyond legislative text. For example, transcripts of floor debates, committee meetings, and written communications (Grimmer 2010) can be incorporated to provide textual data for the legislators. Bill co-sponsorship may also offer meaningful improvements in prediction performance. The model is also applicable to data outside of roll calls; applications include the analysis and prediction of weather based features across large areas based on textual weather observations, joint analysis of stocks and business news, or predicting possible terrorist activity patterns in conflict areas based on intelligence gathered on the ground.

The time dependence investigated here concerned the time evolution of the legislator features (related to their ideal point) as a function of time. However, this analysis has not been placed within the context of political theory, which may be of interest to political scientists. For example, in Clinton and Meirowitz (2003) the authors discussed how the sequence of votes on legislation should often be viewed from a strategic perspective (the votes on a sequence of pieces of legislation may not be independent). This is a form of temporal dependence that has not been considered here, and is worthy of future study within a model of the type developed here.

One of the key aspects of the model concerns its predictive power, particularly for cases in which the prediction of votes is performed only based upon the text of the legislation. Further, temporal dependence of the legislator feature vectors, and spatial locations of the congressional districts (or states) has been accounted for. While this significantly improves prediction, it may undermine the ability to interpret the results. For example, the topics are linked to prediction of the feature vectors of the legislation. This may aid model predictive power, but it may make the inferred topics less interpretable (interpretability is of course not explicitly accounted for in the model). This is meant to emphasize and acknowledge that the increased predictive power of a model like that developed here is well within the capability of many statistically inclined political scientists (to cite a few, see Jackman (2009); Clinton and Meirowitz (2003); Clinton et al. (2004b); Poole and Rosenthal (1985); Jackman (2001); Poole and Rosenthal (1997); Quinn et al. (2006); Grimmer (2010); Clinton et al. (2004a)). However, often in political science studies, model simplicity is a strength, aiding interpretation and grounding the model in a "rational actor" setting. Going forward, a challenge for the statistics and political science communities is to strike a proper balance between model interpretability and prediction power.

References

Ahmed, A. and Xing, E. P. (2010). "Staying Informed: Supervised and Semi-Supervised Multi-view Topical Analysis of Ideological Perspective." In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 1140–1150. 256

Blei, D. and Lafferty, J. (2006). "Dynamic Topic Models." In *Proceedings of the 23rd International Conference on Machine Learning*, 113–120. 261

Blei, D., Ng, A., and Jordan, M. (2003). "Latent Dirichlet Allocation." *Journal of Machine Learning Research*, 3: 993–1022. 234, 236, 237, 243

Carvalho, C. M., Massam, H., and West, M. (2007). "Simulation of hyper-inverse Wishart distributions in graphical models." *Biometrika*, 94: 647–659. 240, 266

Chib, S. and Greenberg, E. (1998). "Analysis of multivariate probit models." *Biometrika*, 85(2): 347–361. 236

Clinton, J., Jackman, D., and Rivers, S. (2004a). "The most liberal senator? Analyzing and interpreting Congressional Roll Calls." *Political Science and Politics*, 37(4): 805–811. 261

Clinton, J., Jackman, S., and Rivers, D. (2004b). "The Statistical Analysis of Roll Call Voting: A Unified Approach." *American Political Science Review*, 98(2): 355–370. 233, 234, 235, 238, 240, 244, 259, 261

Clinton, J. and Meirowitz, A. (2003). "Integrating voting theory and roll call analysis: A framework." *Political Analysis*, 11: 381–396. 261

Dawid, A. and Lauritzen, S. (1993). "Hyper-Markov laws in the statistical analysis of decomposable graphical models." *The Annals of Statistics*, 21(3): 1272–1317. 235, 239

Deerwester, S., Dumais, S., Furnas, G., Landauer, T., and Harshman, R. (1990). "Indexing by Latent Semantic Analysis." *Journal of the American Society for Information Science*, 41(6): 391–407. 234, 237

Dunson, D. and Park, J. (2008). "Kernel stick-breaking processes." *Biometrika*, 95(2): 307–323. 234

Ferguson, T. (1973). "A Bayesian analysis of some nonparametric problems." *The Annals of Statistics*, 1(2): 209–230. 237

Gerrish, S. and Blei, D. (2011). "Predicting legislative roll calls from text." In *Proceedings of the 28th Annual International Conference on Machine Learning*. 234, 235, 238, 240, 246, 256, 259

Griffiths, T. and Ghahramani, Z. (2005). "Infinite latent feature models and the Indian buffet process." In *Advances in Neural Information Processing Systems 18*, 475–482. MIT Press. 241

Grimmer, J. (2010). "A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases." *Political Analysis*, 18(1): 1–35. 234, 239, 261

Grimmer, J. and Stewart, B. (to appear in 2013). "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis*. 239

Hahn, P., Carvalho, C., and Scott, J. (2012). "A sparse factor-analytic probit model for Congressional voting patterns." *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 61(4): 619–635. 233, 246

Jackman, S. (2001). "Multidimensional analysis of roll call data via Bayesian simulation: Identification, estimation, inference, and model checking." *Political Analysis*, 9(3): 227–241. 234, 235, 236, 240, 244, 256, 261, 265

— (2009). *Bayesian Analysis for the Social Sciences*. Wiley Series in Probability and Statistics. Wiley. 238, 261

Paisley, J. and Carin, L. (2009). "Nonparametric factor analysis with beta process priors." In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, 777–784. 235, 237, 240

Poole, K. and Rosenthal, H. (1985). "A spatial model for legislative roll call analysis." *American Journal of Political Science*, 29(2): 357–384. 233, 235, 236, 238, 240, 246, 261

— (1997). *Congress: A Political-Economic History of Roll Call Voting*. New York: Oxford University Press. 261

Porteous, I., Newman, D., Ihler, A., Asuncion, A., Smyth, P., and Welling, M. (2008). "Fast Collapsed Gibbs Sampling for Latent Dirichlet Allocation." In *KDD08: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 569–577. 237, 241

Pruteanu-Malinici, I., Ren, L., Paisley, J., Wang, E., and Carin, L. (2010). "Hierarchical Bayesian Modeling of Topics in Time-Stamped Documents." *IEEE Transactions on Pattern Analysis Machine Intelligence*, 32(6): 996–1011. 234, 261

Quinn, K., Monroe, B., Colaresi, M., Crespin, M., and Radev, D. (2006). "An automated method of topic-coding legislative speech over time with application to the 105th-108th U.S. senate." In *Midwest Political Science Association Meeting*, 1–61. 234, 261

Salakhutdinov, R. and Mnih, A. (2008). "Probabilistic Matrix Factorization." In *Advances in Neural Information Processing Systems 21*, volume 20, 1257–1264. MIT Press. 235, 240, 244

Srebo, N., Rennie, J., and Jaakkola, T. (2005). "Maximum-Margin Matrix Factorization." In *Advances in Neural Information Processing Systems 18*, 1329–1336. MIT Press. 235, 240

Steyvers, M. and Griffiths, T. (2007). "Probabilistic topic models." *Handbook of latent semantic analysis*, 427(7): 424–440. 234

Wallach, H. (2008). "Structured Topic Models for Language." Ph.D. thesis, University of Cambridge. 234

Wang, E., Liu, D., Silva, J., Dunson, D., and Carin, L. (2010). "Joint Analysis of Time-Evolving Binary Matrices and Associated Documents." In *Advances in Neural Information Processing Systems 23*, 2370–2378. MIT Press. 234, 235, 236, 237, 238, 242, 244, 256

Acknowledgments

The authors wish to thank the reviewers and editors for their comments, which have substantially improved the paper. The research reported here was supported by the Army Research Office (Dr. Liyi Dai) and the Office of Naval Research (Dr. Wen Masters).

Appendix

• Update for $\hat{\boldsymbol{x}}_{(k)t}$

The latent features of the legislators, encoded in the matrix $\hat{\boldsymbol{X}}_t$, are tasked with imposing both spatio-temporal dependence and identifiability. Jackman (2001) discussed the difficulty in imposing identifiably via ideal point models to multiple dimensions. A lower-block triangular structure is imposed on $\hat{\boldsymbol{X}}_t$, enforcing its main diagonal to be 1 as in (10). The posterior of the latent features of *all* legislators in dimension k and session t, $\hat{\boldsymbol{x}}_{(k)t}$, is a Gaussian

$$P(\hat{\boldsymbol{x}}_{(k)t}|-) = \mathcal{N}(\mu^*_{(\hat{\boldsymbol{x}}_{(k)t})}, \boldsymbol{\Sigma}^{*-1}_{(\hat{\boldsymbol{x}}_{(k)t})}) \tag{23}$$

where the posterior mean $\mu^*_{(\hat{\boldsymbol{x}}_{(k)t})}$ and precision matrix $\boldsymbol{\Sigma}^*_{(\hat{\boldsymbol{x}}_{(k)t})}$ can take different forms depending on the status of the legislator at time t. Specifically, a legislator may be serving his/her only term, first term, last term, or a term that is neither the first nor last. In each case, the temporal dependence of the legislator is different. Defining $\rho_{jkt} = (s_{ijt} - \boldsymbol{x}_{it}^T\boldsymbol{y}_{jt} - \alpha_{it} - \beta_{jt} + x_{ikt}y_{jkt})/y_{jkt}$, and $\boldsymbol{\Omega}_t$ as the prior covariance matrix of $\hat{\boldsymbol{x}}_{(k)t}$, the posterior updates for $\hat{\boldsymbol{x}}_{(k)t}$ are as follows.

If the legislator is in their first and last (only) term, then there are no temporal dependencies

$$\begin{aligned}\boldsymbol{\Sigma}^*_{(\hat{\boldsymbol{x}}_{(k)t})} &= \sum_j \sqrt{|y_{jkt}|} \\ \mu^*_{(\hat{\boldsymbol{x}}_{(k)t})} &= \boldsymbol{\Sigma}^{*-1}_{(\hat{\boldsymbol{x}}_{(k)t})}(\sum_j \rho_{jkt}\sqrt{|y_{jkt}|}).\end{aligned} \tag{24}$$

If the legislator is in the first of multiple terms, then the posterior is dependent on the legislator's latent features of the following session $t+1$

$$\begin{aligned}\boldsymbol{\Sigma}^*_{(\hat{\boldsymbol{x}}_{(k)t})} &= \boldsymbol{\Omega}^{-1}_{t+1} + \sum_j \sqrt{|y_{jkt}|} \\ \mu^*_{(\hat{\boldsymbol{x}}_{(k)t})} &= \boldsymbol{\Sigma}^{*-1}_{(\hat{\boldsymbol{x}}_{(k)t})}(\boldsymbol{\Omega}^{-1}_{t+1}\hat{\boldsymbol{x}}_{(k),t+1} + \sum_j \rho_{jkt}\sqrt{|y_{jkt}|}).\end{aligned} \tag{25}$$

If the legislator is in the last of multiple terms, then the posterior is dependent on the legislator's latent features of the previous session $t-1$

$$\begin{aligned}\boldsymbol{\Sigma}^*_{(\hat{\boldsymbol{x}}_{(k)t})} &= \boldsymbol{\Omega}^{-1}_{t-1} + \sum_j \sqrt{|y_{jkt}|} \\ \mu^*_{(\hat{\boldsymbol{x}}_{(k)t})} &= \boldsymbol{\Sigma}^{*-1}_{(\hat{\boldsymbol{x}}_{(k)t})}(\boldsymbol{\Omega}^{-1}_{t-1}\hat{\boldsymbol{x}}_{(k),t-1} + \sum_j \rho_{jkt}\sqrt{|y_{jkt}|}).\end{aligned} \tag{26}$$

If the legislator is in the middle of multiple terms, then the posterior is dependent on the legislator's latent features at both the previous and following sessions

$$\begin{aligned}
\boldsymbol{\Sigma}^{*}_{(\hat{\boldsymbol{x}}_{(k)t})} &= \boldsymbol{\Omega}_{t-1}^{-1} + \boldsymbol{\Omega}_{t+1}^{-1} + \sum_{j} \sqrt{|y_{jkt}|} \\
\mu^{*}_{(\hat{\boldsymbol{x}}_{(k)t})} &= \boldsymbol{\Sigma}^{*-1}_{(\hat{\boldsymbol{x}}_{(k)t})} (\boldsymbol{\Omega}_{t-1}^{-1} \hat{\boldsymbol{x}}_{(k),t-1} + \boldsymbol{\Omega}_{t+1}^{-1} \hat{\boldsymbol{x}}_{(k),t+1} + \sum_{j} \rho_{jkt} \sqrt{|y_{jkt}|}).
\end{aligned} \tag{27}$$

• Update for $\boldsymbol{\Omega}_t$
The posterior for covariance matrix $\boldsymbol{\Omega}_t$ is HIW,

$$P(\boldsymbol{\Omega}_t|-) = \mathrm{HIW}_{\boldsymbol{G}_t}(\kappa + N_l, \boldsymbol{\Omega}_0 + (\hat{\boldsymbol{x}}_{(k)t} - \mu^{*}_{(\hat{\boldsymbol{x}}_{(k)t})})(\hat{\boldsymbol{x}}_{(k)t} - \mu^{*}_{(\hat{\boldsymbol{x}}_{(k)t})})^T), \tag{28}$$

where $\kappa = 10 + n_t$, $\boldsymbol{\Omega}_0 = \mathbf{I}_{N_l}$ and $\boldsymbol{\Omega}_t$ is sampled using the scheme proposed by Carvalho et al. (2007).

• Update for $\hat{\boldsymbol{y}}_{jt}$
The posterior for the legislation latent features $\hat{\boldsymbol{y}}_{jt}$ is Gaussian and is influenced by the regression from the topic model parameters,

$$P(\hat{\boldsymbol{y}}_{jt}|-) = \mathcal{N}(\mu^{*}_{(\hat{\boldsymbol{y}}_{jt})}, \boldsymbol{\Sigma}^{*-1}_{(\hat{\boldsymbol{y}}_{jt})}) \tag{29}$$

wtih posterior hyperparameters

$$\begin{aligned}
\boldsymbol{\Sigma}^{*}_{(\hat{\boldsymbol{y}}_{jt})} &= \gamma_{(\hat{\boldsymbol{y}})} + \sum_{i} \boldsymbol{x}_{it} \boldsymbol{x}_{it}{}^T \\
\mu^{*}_{(\hat{\boldsymbol{y}}_{jt})} &= \boldsymbol{\Sigma}^{*-1}_{(\hat{\boldsymbol{y}}_{jt})} (\gamma_{(\hat{\boldsymbol{y}})} (\hat{\boldsymbol{\eta}}^T \bar{\boldsymbol{z}}_{jt} + \eta_0) + \sum_{i} s^{(-\boldsymbol{y})}_{ijt} \boldsymbol{x}_i),
\end{aligned} \tag{30}$$

where $s^{(-\boldsymbol{y})}_{ijt} = s_{ijt} - \alpha_{it} - \beta_{jt}$ and $\hat{\boldsymbol{\eta}} \in \mathbb{R}^{H \times K}$ is a matrix of mixing weights whose kth column contains the regression weights for the kth latent dimension.

• Update for β_{jt}
The posterior update for the random effect β_{jt} is very similar to that of the legislation latent features

$$P(\beta_{jt}|-) = \mathcal{N}(\mu^{*}_{(\beta_{jt})}, \boldsymbol{\Sigma}^{*-1}_{(\beta_{jt})}) \tag{31}$$

where

$$\begin{aligned}
\boldsymbol{\Sigma}^{*}_{(\hat{\boldsymbol{y}}_{jt})} &= \gamma_{(\beta)} + N_l \\
\mu^{*}_{(\beta_{jt})} &= \boldsymbol{\Sigma}^{*-1}_{(\beta_{jt})} (\gamma_{\hat{\beta}} (\boldsymbol{\lambda}^T \bar{\boldsymbol{z}}_{jt} + \lambda_0) + \sum_{i} s^{(-\beta)}_{ijt} \boldsymbol{x}_i),
\end{aligned} \tag{32}$$

and $s^{(-\beta)}_{ijt} = s_{ij} - \boldsymbol{x}_{it}^T \boldsymbol{y}_{jt} - \alpha_{it}$ is the contribution of β_{jt} to s_{ijt}.

• Update for z_{vjt}
The posterior update for the per-word topic indicator z_{vjt} is identical to that in LDA in the two-step model. However, the additional dependencies induced by the one-step

model increase the complexity of this update. The posterior probability of word v in bill j and session t being drawn from topic h is

$$P(z_{vjt} = h|-) = \frac{\tau_{vjth}}{\sum_{h'=1}^{H} \tau_{vjth}}, \tag{33}$$

where

$$\tau_{vjth} = P(d_{vjt}|\phi_h)P(z_{vjt} = h|\boldsymbol{s}_{jt}^{(-v)}, \boldsymbol{\eta}, \boldsymbol{\eta}_0, \gamma_{(\hat{\boldsymbol{y}}_k)}, \boldsymbol{b})P(z_{vjt} = h|\boldsymbol{s}_{jt}^{(-v)}, \boldsymbol{\lambda}, \lambda_0, \gamma_{(\beta_k)}, \boldsymbol{b})P(\phi_h)$$

and $\boldsymbol{z}_{jt}^{(-v)}$ is the counts of topics for each document with the vth word's topic assignment removed. Compared to the standard LDA topic assignment update, equation (34) includes two additional likelihood terms that deal with the effect the topic assignment of word d_{vjt} has on the regression covariates $\boldsymbol{z}_{jt}$.