Technical Appendix for "Voting, Speechmaking, and the Dimensions of Conflict in the US Senate"

In Song Kim* John Londregan† Marc Ratkovic‡

July 16, 2015

Abstract

We include here several technical appendices. First, we formally derive the expressions for c_l^{vote} and b_p^{vote} that appear in section two of the paper. The next contains some details on the choice theoretic model underlying the standard topic model. The remainder contain the details for implementing the SFA model.

A The Legislator and Proposal Intercepts from the Voting Model

Let's start with the legislator's utility from the "aye":

$$U_l(\{r_{pd}\}_{d=1}^D) = -\frac{1}{2}\sum_{d=1}^{D} a_d(r_{pd} - x_{ld})^2 + \tilde{\xi}_{lp}^{aye} \quad (1)$$

and "nay" alternatives:

$$U_l(\{q_{pd}\}_{d=1}^D) = -\frac{1}{2}\sum_{d=1}^{D} a_d(q_{pd} - x_{ld})^2 + \tilde{\xi}_{lp}^{nay} \quad (2)$$

Next, let's calculate the difference between these expressions to get the legislator's preference intensity for the "aye" outcome. Substituting from expressions (1) and (2) we have:

*Assistant Professor, Department of Political Science, Massachusetts Institute of Technology, Cambridge, MA, 02139. Email: insong@mit.EDU, URL: http://web.mit.edu/insong/www/

†Professor of Politics and International Affairs, Woodrow Wilson School, Princeton University, Princeton NJ 08544. Phone: 609-258-4854, Email: jbl@princeton.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

$$
\begin{aligned}
V^*_{lp} &= U_l(\{r_{pd}\}_{d=1}^D) - U_l(\{q_{pd}\}_{d=1}^D) \\
&= -\frac{1}{2}\sum_{d=1}^D a_d(r_{pd}-x_{ld})^2 + \tilde{\xi}^{aye}_{lp} - \big(-\frac{1}{2}\sum_{d=1}^D a_d(q_{pd}-x_{ld})^2 + \tilde{\xi}^{nay}_{lp}\big) \\
&= \sum_{d=1}^D \frac{a_d}{2}(q_{pd}^2 - r_{pd}^2) + \sum_{d=1}^D \frac{a_d}{2} \times 2x_{pd}\underbrace{(r_{pd}-q_{pd})}_{g^{vote}_{pd}} + \tilde{\xi}^{aye}_{lp} - \tilde{\xi}^{nay}_{lp} \\
&= \underbrace{\Big(\sum_{d=1}^D \frac{a_d}{2}(q_{pd}^2 - r_{pd}^2) + E\{\tilde{\xi}^{aye}_{lp}\} - E\{\tilde{\xi}^{nay}_{lp}\}\Big)}_{c^{vote}_l + b^{vote}_p} + \sum_{d=1}^D \frac{a_d}{2} \times 2x_{pd}\underbrace{(r_{pd}-q_{pd})}_{g^{vote}_{pd}} \\
&\qquad - \underbrace{\Big(\tilde{\xi}^{nay}_{lp} - \tilde{\xi}^{aye}_{lp} + E\{\tilde{\xi}^{aye}_{lp}\} - E\{\tilde{\xi}^{nay}_{lp}\}\Big)}_{\tilde{\epsilon}_{lp}} \qquad (3)
\end{aligned}
$$

Now let:

$$
E\{\tilde{\xi}^{aye}_{lp}\} = \pi^{aye}_l + \varphi^{aye}_p \text{ and } E\{\tilde{\xi}^{nay}_{lp}\} = \pi^{nay}_l + \varphi^{nay}_p
$$

substituting this into our expression for $c^{vote}_l + b^{vote}_p$ we have:

$$
\begin{aligned}
\sum_{d=1}^D \frac{a_d}{2}(q_{pd}^2 - r_{pd}^2) + E\{\tilde{\xi}^{aye}_{lp}\} - E\{\tilde{\xi}^{nay}_{lp}\} &= \sum_{d=1}^D \frac{a_d}{2}(q_{pd}^2 - r_{pd}^2) + (\pi^{aye}_l + \varphi^{aye}_p) - (\pi^{nay}_l + \varphi^{nay}_p) \\
&= \underbrace{\pi^{aye}_l - \pi^{nay}_l}_{c^{vote}_l} + \underbrace{\sum_{d=1}^D \frac{a_d}{2}(q_{pd}^2 - r_{pd}^2) + \varphi^{aye}_p - \varphi^{nay}_p}_{b^{vote}_p} \\
&= c^{vote}_l + b^{vote}_p
\end{aligned}
$$

Now let's return to the last line of expression (3) and substitute:

$$
\begin{aligned}
V^*_{lp} &= U_l(\{r_{pd}\}_{d=1}^D) - U_l(\{q_{pd}\}_{d=1}^D) \\
&= \underbrace{\Big(\sum_{d=1}^D \frac{a_d}{2}(q_{pd}^2 - r_{pd}^2) + E\{\tilde{\xi}^{aye}_{lp}\} - E\{\tilde{\xi}^{nay}_{lp}\}\Big)}_{c_l^{vote}+b_p^{vote}} + \sum_{d=1}^D \frac{a_d}{2} \times 2x_{pd} \underbrace{(r_{pd} - q_{pd})}_{g_{pd}^{vote}} \\
&\qquad - \underbrace{\Big(\tilde{\xi}^{nay}_{lp} - \tilde{\xi}^{aye}_{lp} + E\{\tilde{\xi}^{aye}_{lp}\} - E\{\tilde{\xi}^{nay}_{lp}\}\Big)}_{\tilde{\epsilon}_{lp}} \\
&= c_l^{vote} + b_p^{vote} + \sum_{d=1}^D a_d x_{pd} g_{pd}^{vote} - \tilde{\epsilon}_{lp} \qquad (4)
\end{aligned}
$$

Expression (4) matches expression (4) in the text.

B Choice Theoretic Underpinnings of Topic Models

While we opt for SFA, it is useful to consider the behavior that would lead one to adopt a topic model for legislative speech. One way to do this is to suppose that a legislator's speech is generated by the random arrival of opportunities to speak. At each opportunity the legislator must choose one word from a lexicon, which we represent by a $W \times 1$ vector $\boldsymbol{\omega}$, with each entry corresponding to a different word. Each word has a spatial location, which for the moment we place on a single dimension. Legislator $j \in \{1 \dots V\}$ would derive utility $u(\tilde{w}_l|x_j) + \eta_{j,t}$ from uttering word $j \in \{1 \dots W\}$ at time t. Should the opportunity to speak at time t actually arise, the legislator utters the word offering the greatest utility. To keep things simple we assume that $\eta_{j,t}$ and $\eta_{r,s}$ are independent if either $j \neq r$ or $t \neq s$.

Paralleling the development in Maddala (1983), we operationalize our model with a distributional assumption for $\eta_{j,t} \in \mathbb{R}$, which we take to follow a type I extreme value distribution, with probability density:

$$f(\eta) = e^{-(\eta + e^{-\eta})}$$

and by concretizing the utility function $u(\tilde{w}_l|x_j)$:

$$u(\tilde{w}_l|x_j) = -\frac{1}{2}\big(\tilde{w}_l - x_j\big)^2 \tag{5}$$

where x_j is the preferred ideological signal that legislator j would like to convey, and $\tilde{w}_l$ is the ideological connotation of word i.

Again following Maddala (1983) we see that the probability that at a randomly chosen time t legislator j prefers word i to all other elements of the lexicon is:

$$q_{lj} = \frac{e^{u(\tilde{w}_l|x_j)}}{\sum\limits_{k=1}^{W} e^{u(\tilde{w}_k|x_j)}}$$

Let word 1 correspond to a "stop word". We can rewrite the probability j uses word i if she has the opportunity to speak at t as:

$$q_{lj} = \frac{e^{u(\tilde{w}_l|x_j)-u(\tilde{w}_1|x_j)}}{\sum\limits_{k=1}^{W} e^{u(\tilde{w}_k|x_j)-u(\tilde{w}_1|x_j)}}$$

substituting from equation (5) into our expression for q_{lj} we have:

$$q_{lj}(\mathbf{x}, \mathbf{g}, \mathbf{b}) = \frac{e^{x_j g_l + b_l}}{1 + \sum\limits_{k=2}^{W} e^{x_j g_k + b_k}} \tag{6}$$

where $g_k = \tilde{w}_k - \tilde{w}_1$ and $b_k = -\frac{\tilde{w}_k + \tilde{w}_1}{2}$ for $k \in \{2 \ldots W\}$.

The probability of an observed $W \times 1$ vector $\mathbf{c}$ of word counts is:

$$\prod_{w=1}^{W} q_{lj}(\mathbf{x}, \mathbf{g}, \mathbf{b})^{c_w} \tag{7}$$

With the right choice of Dirichlet priors this turns into the latent Dirichlet model of Blei, Ng and Jordan (2003) if we set $x_j = g_l = 0$ for all i and j. In the ideal point setting, though, x_j and

g_i correspond with precisely the preferred outcomes and term ideologies with which we are most interested.

Estimation for these models are not straightforward, requiring a Metropolis algorithm or variational approximations. We favor SFA on theoretical grounds, as it allows legislators to select words as a function of their preferred outcomes. We also favor it because it offers a tractable Gibbs sampling scheme for most of the parameters, which we address in the next section.

C Estimation of SFA

We now shift to a more condensed notation. Hereafter, we reindex the vote and term outcomes using a common index, j, which falls into two sets: J^{terms} and J^{votes} for whether the observed outcome (now a common Y_{lj}) is a term outcome or vote outcome, and $J = |J^{terms}| + |J^{votes}|$. We will also supress the superscript for the θ_{lw}^{terms} and θ_{lp}^{votes} while changing to the joint subscript j. The likelihood is given by:

$$\mathcal{L}\big(\theta_{\cdot\cdot}^{vote}, \theta_{\cdot\cdot}^{term}, \tau.|\tilde{T}_{\cdot\cdot}, \tilde{V}_{\cdot\cdot}\big) = \prod_{l=1}^{L}\Big(\prod_{p=1}^{P} Pr\{V_{lp} = \widetilde{V}_{lp}|\cdot\}^{\frac{W+P}{2P}} \prod_{w=1}^{W} Pr\{T_{lw} = \widetilde{T}_{lw}|\cdot\}^{\frac{W+P}{2W}}\Big). \tag{8}$$

where:

$$Pr\{T_{lw} = \widetilde{T}_{lw}|\cdot\} = \begin{cases} \Phi\left(\theta_{lw}^{terms} - \tau_0\right) & T_{lw} = 0 \\ \Phi\left(\theta_{lw}^{terms} - \tau_{\widetilde{T}_{lw}}\right) - \Phi\left(\theta_{lw}^{terms} - \tau_{\widetilde{T}_{lw}-1}\right) & 0 < T_{wl} \end{cases} \tag{9}$$

$$Pr\{V_{lp} = \widetilde{V}_{lp}|\cdot\} = \Phi\big((2\widetilde{V}_{lp} - 1)\theta_{lp}^{vote}\big) \tag{10}$$

and the prior structure is given by:

$$
\begin{aligned}
c_l^{\cdot}, b_w^{\cdot} &\overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, 1) \\
\mu &\sim \mathcal{N}(0, 1) \\
g_{wd}^{\cdot}, x_{ld} &\overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 4) \\
\log(\beta_1),\ \log(\beta_2) &\overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1) \\
\Pr(a_d) &= \tfrac{1}{2\lambda} e^{-\lambda |a_d|} \\
\Pr(\lambda) &= 1.78 e^{-1.78\lambda}
\end{aligned} \tag{11}
$$

Combining the likelihood and prior gives us the posterior:

$$
\begin{aligned}
\Pr(\theta_{lj}, \tau_{\cdot}, \beta_1, \beta_2 | Y_{\cdot\cdot}) = &\prod_{\substack{1 \le l \le L \\ 1 \le j \le J}} \Bigg\{ \left\{ \Phi\left(\theta_{lj}\right)^{Y_{lj}} \left(1 - \Phi\left(\theta_{lj}\right)\right)^{1 - Y_{lj}} \right\}^{\mathbf{1}(j \in \{j_{votes}\})} \\
&\times \left\{ \Phi\left(\tau_{Y_{lj}} - \theta_{lj}\right) - \Phi\left(\theta_{lj} - \tau_{Y_{lj}-1}\right) \right\}^{\mathbf{1}(j \in \{J_{terms}\})} \Bigg\} \\
&\times \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\mu^2} \times \prod_{1 \le l \le L} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(c_l - \mu)^2} \times \prod_{1 \le j \le J} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(b_j - \mu)^2} \\
&\times \prod_{\substack{1 \le d \le D \\ 1 \le l \le L}} \frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{8}(x_{ld})^2} \times \prod_{\substack{1 \le d \le D \\ 1 \le l \le L}} \frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{8}(g_{jd})^2} \times \prod_{1 \le d \le D} \frac{1}{2\lambda} e^{-\lambda |a_d|} \\
&\times \frac{1}{\beta_1 \sqrt{2\pi}} e^{-\frac{1}{2}(\log \beta_1)^2} \times \frac{1}{\beta_2 \sqrt{2\pi}} e^{-\frac{1}{2}(\log \beta_2)^2} \times e^{-1.78\lambda}
\end{aligned} \tag{12}
$$

We implement two forms of data augmentation. In the first, for each observation we introduce a normal random variable Z_{lj}^* as is standard in latent probit models (Albert and Chib, 1993). This transforms the likelihood into a least squares problem, as:

$$
\Pr(Y_{lj} = k | Z_{lj}^*, \theta_{lj}, \tau_{\cdot}, \beta_1, \beta_2) = \prod_{\substack{1 \le l \le L \\ 1 \le j \le J}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(Z_{lj}^* - \theta_{lj})^2} \tag{13}
$$

The second form of augmentation involves representing the double exponential prior for a_d to maintain conjugacy. Following Park and Casella (2008), we introduce latent variables $\widetilde{\tau}_l$, such that:

$$d.|\widetilde{\tau}^2_{\cdot} \sim \mathcal{N}(0_D, \widetilde{D}_{\widetilde{\tau}}) \tag{14}$$

$$\widetilde{D}_{\widetilde{\tau}} = diag(\widetilde{\tau}^2_1, \widetilde{\tau}^2_2, \cdots, \widetilde{\tau}^2_D) \tag{15}$$

$$\widetilde{\tau}^2_1, \widetilde{\tau}^2_2, \ldots, \widetilde{\tau}^2_D \sim \prod_{1 \le d \le D} \frac{\lambda^2}{2} e^{-\lambda^2 \widetilde{\tau}^2_d / 2} d\widetilde{\tau}^2_d \tag{16}$$

where, after integrating out $\widetilde{\tau}^2_l$, we are left with the LASSO prior. The proposed method differs from the presentation in Park and Casella (2008) in that we know $\sigma^2 = 1$, by assumption.

C.1 The Gibbs Sampler

Next, we outline the Gibbs sampler. All conditional posterior densities are conjugate normals except λ, $\widetilde{\tau}^2$, β_1, and β_2. For a derivation of the posterior densities of λ and $\widetilde{\tau}^2$, see Park and Casella (2008). We fit β_1 and β_2, which determine $\tau.$, using a Hamiltonian Monte Carlo algorithm, but first we describe the Gibbs updates.

The Gibbs updates occur in two steps. First, we place all data on the latent z scale. Second, we update all of the remaining parameters. For the first step, we sample as:

$$Z^*_{lj}|\cdot \sim \begin{cases} \mathcal{TN}\left(\theta_{lj}, 1, 0, \infty\right);\ Y_{lj} = 1,\ j \in j_{votes} \\ \mathcal{TN}\left(\theta_{lj}, 1, -\infty, 0\right);\ Y_{lj} = 0,\ j \in j_{votes} \\ \mathcal{TN}\left(\theta_{lj}, 1, \tau_{k-1}, \tau_k\right);\ Y_{lj} = k,\ j \in J_{terms} \\ \mathcal{N}\left(\theta_{lj}, 1\right);\ Y_{lj} \text{ missing} \end{cases} \tag{17}$$

Note that we have ignored missing values up to this point. In the Bayesian framework used here, imputing is straightforward: the truncated normal is replaced with a standard normal, whether term or vote data.

Next, we update all of θ_{lj} except for $\tau.$ using a Gibbs sampler, as:

$$\mu|\cdot \sim \mathcal{N}\left(\frac{\sum_{l=1}^{L}\sum_{j=1}^{J} Z_{lj}^{*}}{LJ+1}, \frac{1}{L^2J^2+1}\right) \tag{18}$$

$$c_l|\cdot \sim \mathcal{N}\left(\frac{\sum_{j=1}^{J} Z_{lj}^{*}}{J+1}, \frac{1}{J^2+1}\right) \tag{19}$$

$$b_j|\cdot \sim \mathcal{N}\left(\frac{\sum_{l=1}^{L} Z_{lj}^{*}}{L+1}, \frac{1}{L^2+1}\right) \tag{20}$$

$$Z_{lj}^{**} = Z_{lj}^{*} - c_l - b_j + \mu \tag{21}$$

$$\text{Update } x_{\cdot\cdot}, w_{\cdot}, v_{\cdot\cdot} \text{ from SVD of } Z^{**} \tag{22}$$

$$a_{\cdot}|\cdot \sim \mathcal{N}\left(A^{-1}\widetilde{X}^{\top} vec(Z^{**}), A^{-1}\right) \text{ where}$$

$$\widetilde{X} = \left[vec\left(x_{\cdot 1}g_{\cdot 1}^{\top}\right) : vec\left(x_{\cdot 2}g_{\cdot 2}^{\top}\right) : \ldots : vec\left(x_{\cdot L}g_{\cdot L}^{\top}\right)\right] \text{ and} \tag{23}$$

$$A = \widetilde{X}^{\top}\widetilde{X} + T^{-1} \text{ with } T = diag(\tau_l^2)$$

$$x_{\widetilde{l}\widetilde{d}}|\cdot \sim \mathcal{N}\left(\frac{\sum_{j=1}^{J} Z_{\widetilde{l}j,-\widetilde{d}}^{**} a_{\widetilde{d}} g_{j\widetilde{l}}}{\sqrt{\sum_{j=1}^{J}\left(a_{\widetilde{d}}^2 g_{j\widetilde{d}}^2 + \frac{1}{4J}\right)}}, \frac{1}{\sum_{j=1}^{J}\left(a_{\widetilde{d}}^2 g_{j\widetilde{d}}^2 + \frac{1}{4J}\right)}\right) \tag{24}$$

$$g_{\widetilde{j}\widetilde{d}}|\cdot \sim \mathcal{N}\left(\frac{\sum_{l=1}^{L} Z_{l\widetilde{j},-\widetilde{d}}^{**} a_{\widetilde{d}} x_{l\widetilde{d}}}{\sqrt{\sum_{l=1}^{L}\left(a_{\widetilde{d}}^2 x_{l\widetilde{d}}^2 + \frac{1}{4L}\right)}}, \frac{1}{\sum_{l=1}^{L}\left(a_{\widetilde{d}}^2 x_{l\widetilde{d}}^2 + \frac{1}{4L}\right)}\right) \tag{25}$$

where

$$Z_{lj,-\widetilde{d}}^{**} = Z_{lj}^{**} - \sum_{d\neq\widetilde{d}} x_{l\widetilde{d}} g_{jd} a_d$$

$$\widetilde{\tau}_l^2|\cdot \sim InvGauss\left(\sqrt{\frac{\lambda^2}{a_d^2}}, \lambda^2\right) \tag{26}$$

$$\lambda^2|\cdot \sim Gamma\left(L+1, \sum_{l=1}^{L}\widetilde{\tau}_l^2/2 + 1.78\right) \tag{27}$$

C.2 The Hamiltonian Monte Carlo Sampler

We have no closed form estimates for the conditional posterior densities of β_1 and β_2. To estimate these, we implement a Hamiltonian Monte Carlo scheme adapted directly from Neal (2011). We adapt the algorithm in one important manner: rather than taking a negative gradient step, we calculate the numerical Hessian and take a fraction (α) of a Newton-Raphson step at each. We select α so that the acceptance ratio of proposed (β_1, β_2) is about .4.

Specifically, let $\widehat{dev}(\beta_1, \beta_2)$ denote the estimate deviance at the point (β_1, β_2). Define the numerical gradients, $\widehat{\nabla}_1 dev(\beta_1, \beta_2)$ and $\widehat{\nabla}_2 dev(\beta_1, \beta_2)$ as the estimated gradient at (β_1, β_2) and $\widehat{\nabla}_{11} dev(\beta_1, \beta_2)$, $\widehat{\nabla}_{22} dev(\beta_1, \beta_2)$, and $\widehat{\nabla}_{12} dev(\beta_1, \beta_2)$ as the cross derivative. Next, define the empirical Hessian as:

$$\widehat{H}(\beta_1, \beta_2) = \begin{pmatrix} \widehat{\nabla}_{11} dev(\beta_1, \beta_2) & \widehat{\nabla}_{12} dev(\beta_1, \beta_2) \\ \widehat{\nabla}_{12} dev(\beta_1, \beta_2) & \widehat{\nabla}_{22} dev(\beta_1, \beta_2) \end{pmatrix} \tag{28}$$

We implement the algorithm in Neal (2011) exactly, except instead taking updates of the form:

$$\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}^{+} := \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}^{-} - \alpha \begin{pmatrix} \widehat{\nabla}_1(\beta_1, \beta_2) \\ \widehat{\nabla}_2(\beta_1, \beta_2) \end{pmatrix} \tag{29}$$

we instead do updates of the form:

$$\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}^{+} := \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}^{-} - \alpha \times \left\{ \widehat{H}(\beta_1, \beta_2) \right\}^{-1} \begin{pmatrix} \widehat{\nabla}_1(\beta_1, \beta_2) \\ \widehat{\nabla}_2(\beta_1, \beta_2) \end{pmatrix} \tag{30}$$

where the Hessian and gradients are updated every third update of the parameters. The step length parameter α is adjust every 50 iterations to by a factor of 4/5 if the acceptance rate is below 10%, 5/4 if the acceptance rate is above 90%, and left the same otherwise. After the burn-in period, the acceptance rate levels off around 45%. We implement twenty steps in order to produce a proposal.

C.3 Numerical Approximation of the Deviance

Calculating the gradient and Hessian terms, and assessing the proposal, in the Hamiltonian Monte Carlo scheme requires evaluating functions of the form $l(a,b) = \log(\Phi(a) - \Phi(b))$. Unfortunately, for values of a and b much larger in magnitude than 5.3 produces returns values of 1 or 0, leaving it impossible to evaluate the logarithm.

Extrapolating from the observed values yields the linear approximation:

$$l(a,b) = \begin{pmatrix} 1 \\ a \\ b \\ a^2 \\ b^2 \\ \log(|a-b|) \\ \{\log(|a-b|)\}^2 \\ ab \end{pmatrix}^{\top} \gamma \tag{31}$$

where

$$\gamma = \begin{pmatrix} -1.82517672 \\ 0.51283415 \\ -0.81377290 \\ -0.02699400 \\ -0.49642787 \\ -0.33379312 \\ -0.24176661 \\ 0.03776971 \end{pmatrix} \tag{32}$$

We derived the values for γ from fitting a model over the range $4 \leq b < a \leq 8$. We get a mean absolute error of 0.0165, or 0.08% error as a fraction of the value returned by **R**. We use this approximation in order to extrapolate to values where **R** returns values of NA or Inf for $f(a,b)$.

D How do Dimensions and Topics Differ?

As we noted in the introduction, there is an apparent inconsistency between Congress voting across a broad array of topics (national defense, agriculture, social insurance, etc.) while roll call voting over this range of substantive topics arranges into a single dimensional ideological space. This seeming paradox is closely tied to the distinction between topics, as estimated by topic models, and dimensions, as estimated by SFA. There is no actual contradiction. Which rendition of Congress a researcher finds more useful: a collection of substantive legislative topics subject to Congressional dispensation, or a low dimensional "radiography" of the ideological space that organizes Congress, depends on the researcher's objectives.

To differentiate topics from dimensions, consider the distinction between "surface" and "source" traits. Factor analysis pioneer Raymond Cattell characterized a surface trait as "... an obvious 'going together' of this and that" while he defined the source trait as the "underlying contributor or

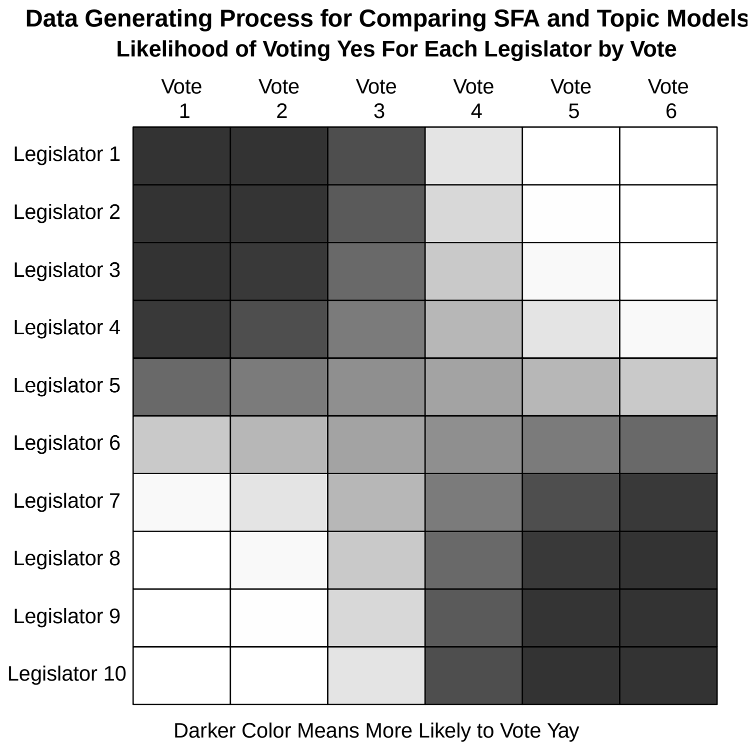

Figure 1: **Simulated data setup.** Legislators are arrayed across rows and votes across columns. The darker the square, the more likely the legislator to vote Aye on that particular vote.

determiner" of the observed surface traits (**?**, p. 45). As an example, he offered common symptoms of schizophrenia as surface traits, while the underlying syndrome itself is the source trait.

We provide an example to illustrate the point. Assume ten legislators facing six votes. For simplicity, the true underlying probability of voting Aye comes from an underlying process with one ideological dimension, as presented in Figure 1. Legislators are arrayed across rows and votes across columns. The darker the square, the more likely the legislator to vote Aye on that particular proposal. Legislators 1 – 5 are more likely to vote Aye on the first 3 votes and more likely to vote Nay on the last 3. Legislators 6–10 are more likely to vote Nay on the first 3 votes, and Aye on the last 3. Legislators 5 and 6 are relative moderates, while bills 3 and 4 are relatively noncontroversial.[1]

We fit both SFA and a topic model to a draw of the vote data. In order to fit a topic model, we assume each legislator uttered six "terms" representing their vote and the bill number. For example,

[1] Specifically, let $s_i = \{-4.5, -3.5, \dots, 4.5\}$ and $w_j = \{-2.5, -1.5, \dots, 2.5\}$. We drew $Y_{lj} \sim Bern\left(\Phi(s_i w_j/2)\right)$.

SFA Results			Topic Model Results		
Dimension Displacements	Most Preferred Outcomes	Bill Weights	Topic 1	Topic 2	Topic 3
0.24	0.87	1.19	Nay on 2	Nay on 4	Aye on 1
0	1.02	0.84	Nay on 1	Nay on 6	Nay on 5
0	1.02	0.57	Aye on 4	Aye on 2	Aye on 2
0	1.02	-0.60	Aye on 6	Aye on 3	Aye on 3
0	0.36	-1.06			
0	0.14	-1.03			
	-1.02				
	-0.95				
	-1.11				
	-1.35				

Table 1: **Results from SFA and a Topic Model on the Simulated Dataset.**

if the legislator voted Nay on vote 4, we assume that they said "Nay on 4," and enter that into a topic model as a unique word. Each legislator then uttered six elements from the set {"Aye on 1", "Nay on 1", "Aye on 2", "Nay on 2", …, "Aye on 6", "Nay on 6"}, one of either "Aye" or "Nay" on votes 1-6. We implemented the EM version of SFA and also gave the same data to the a Structural Topic Model, as implemented in `stm`. We fit a three-topic model to the data. Four-, five-, and six-topic models returned qualitatively similar results.

The left three columns of Table 1 contain the results from SFA. The first column contains the estimated posterior mode, and only the first dimension has a non-zero mode. The next two columns contain each legislator's ideal points and the bill estimates. SFA returns estimates of the underlying structure, correctly recovering the unidimensional structure of the data generating process, and identifying legislators 1-4 and 7-10 as relative extremists at opposite ends of the spectrum. SFA also successfully identifies the relatively moderate legislators, 5 and 6, and correctly notes which proposals will draw support from which legislators.

The rightmost three columns of Table 1 report the topic model estimates, presenting the first four terms of the three fitted topics. Consider the first topic. Legislators that vote Nay on votes 1 and 2 are likely to vote "Aye" on votes 4 and 6. Similarly, considering the second topic, legislators who vote "Nay" on votes 4 and 6 are likely to vote "Aye" on votes 2 and 3. The topic model is returning surface traits–"symptoms" of an underlying ideological dimension, but not the dimension itself.

Topic models provide an excellent tool for summarizing word co-occurrence if the goal is primarily descriptive (e.g., **?**). If, on the other hand, the researcher seeks to identify the underlying structure of the source of the observed behavior, whether it is speech or voting, then SFA provides a clearer picture.

References

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

Blei, David M., Andrew Y. Ng and Michael I. Jordan. 2003. "Latent dirichlet allocation." *J. Mach. Learn. Res.* 3:993–1022.

Maddala, Gangadharrao Soundalyarao. 1983. *Limited Dependent and Qualitative Variables in Econometrics.* New York: Cambridge.

Neal, Radford. 2011. MCMC Using Hamiltonian Dynamics. In *Handbook of Markov Chain Monte Carlo*, ed. Steve Brooks, Andrew Gelman, Galin Jones and Xiao-Li Meng. CRC Handbooks of Modern Statistical Method Chapman and Hall.

Park, Trevor and George Casella. 2008. "The bayesian lasso." *Journal of the American Statistical Association* 103(482):681–686.