# Additional Materials: Estimating Preferred Outcomes from Votes and Text

December 15, 2017

#### Abstract

This appendix contains two sets of supplemental results. First, we include a technical appendix outlining estimation for SFA. Second, we include a set of supplemental empirical results.

# **1** Technical Appendix

## 1.1 The Legislator and Proposal Intercepts from the Voting Model

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

$$U_l^{vote}\left(Aye; \{x_{ld}\}_{d=1}^D, \{z_{pd}^{aye}\}_{d=1}^D\right) = -\frac{1}{2}\sum_{d=1}^D a_d (z_{pd}^{aye} - x_{ld})^2 + \tilde{\xi}_{lp}^{aye}$$
(1)

and "Nay" alternatives:

$$U_l^{vote}\left(Nay; \{x_{ld}\}_{d=1}^D, \{z_{pd}^{nay}\}_{d=1}^D\right) = -\frac{1}{2}\sum_{d=1}^D a_d (z_{pd}^{nay} - x_{ld})^2 + \tilde{\xi}_{lp}^{nay}$$
(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:

$$\begin{split} V_{lp}^{*} &= U_{l}^{vote} \left( Aye; \{x_{ld}\}_{d=1}^{D}, \{z_{pd}^{aye}\}_{d=1}^{D} \right) - U_{l}^{vote} \left( Nay; \{x_{ld}\}_{d=1}^{D}, \{z_{pd}^{nay}\}_{d=1}^{D} \right) \\ &= -\frac{1}{2} \sum_{d=1}^{D} a_{d} (z_{pd}^{aye} - x_{ld})^{2} + \tilde{\xi}_{lp}^{aye} - \left( -\frac{1}{2} \sum_{d=1}^{D} a_{d} (z_{pd}^{nay} - x_{ld})^{2} + \tilde{\xi}_{lp}^{nay} \right) \\ &= \sum_{d=1}^{D} \frac{a_{d}}{2} (z_{pd}^{nay^{2}} - z_{pd}^{aye^{2}}) + \sum_{d=1}^{D} \left( \frac{a_{d}}{2} \cdot 2x_{ld} (z_{pd}^{aye} - z_{pd}^{nay}) \right) + \tilde{\xi}_{lp}^{aye} - \tilde{\xi}_{lp}^{nay} \\ &= \left( \sum_{d=1}^{D} \frac{a_{d}}{2} (z_{pd}^{nay^{2}} - z_{pd}^{aye^{2}}) + E\{\tilde{\xi}_{lp}^{aye}\} - E\{\tilde{\xi}_{lp}^{nay}\} \right) + \sum_{d=1}^{D} a_{d}x_{ld} (z_{pd}^{aye} - z_{pd}^{nay}) \\ &- \left( \tilde{\xi}_{lp}^{nay} - \tilde{\xi}_{lp}^{aye} + E\{\tilde{\xi}_{lp}^{aye}\} - E\{\tilde{\xi}_{lp}^{nay}\} \right) \end{split}$$

Now, let:

$$E\{\tilde{\xi}_{lp}^{aye}\} = \pi_l^{aye} + \varphi_p^{aye} \text{ and } E\{\tilde{\xi}_{lp}^{nay}\} = \pi_l^{nay} + \varphi_p^{nay}$$

substituting this into the first part of the equation (3),

$$\begin{split} \sum_{d=1}^{D} \frac{a_d}{2} (z_{pd}^{nay^2} - z_{pd}^{aye^2}) + E\{\tilde{\xi}_{lp}^{aye}\} - E\{\tilde{\xi}_{lp}^{nay}\} &= \sum_{d=1}^{D} \frac{a_d}{2} (z_{pd}^{nay^2} - z_{pd}^{aye^2}) + (\pi_l^{aye} + \varphi_p^{aye}) - (\pi_l^{nay} + \varphi_p^{nay}) \\ &= \underbrace{\pi_l^{aye} - \pi_l^{nay}}_{c_l^{vote}} + \underbrace{\sum_{d=1}^{D} \frac{a_d}{2} (z_{pd}^{nay^2} - z_{pd}^{aye^2}) + \varphi_p^{aye} - \varphi_p^{nay}}_{b_p^{vote}} \\ &= c_l^{vote} + b_p^{vote} \end{split}$$

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

$$V_{lp}^{*} = U_{l}(\{x_{ld}\}_{d=1}^{D}, \{p_{d}\}_{d=1}^{D}) - U_{l}(\{x_{ld}\}_{d=1}^{D}, \{q_{d}\}_{d=1}^{D})$$

$$= \underbrace{\left(\sum_{d=1}^{D} \frac{a_{d}}{2}(z_{pd}^{nay^{2}} - z_{pd}^{aye^{2}}) + E\{\tilde{\xi}_{lp}^{aye}\} - E\{\tilde{\xi}_{lp}^{nay}\}\right)}_{c_{l}^{vote} + b_{p}^{vote}} + E\{\tilde{\xi}_{lp}^{aye} + E\{\tilde{\xi}_{lp}^{aye}\} - E\{\tilde{\xi}_{lp}^{nay}\}\right)}_{\epsilon_{lp}^{vote}} + E\{\tilde{\xi}_{lp}^{aye}\} - E\{\tilde{\xi}_{lp}^{nay}\}\right)}$$

$$= c_{l}^{vote} + b_{p}^{vote} + \sum_{d=1}^{D} a_{d}x_{ld}g_{pd}^{vote} - \epsilon_{lp}^{vote}.$$
(4)

Equation (4) matches equation (??).

## 1.2 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:  $J^{term}$  and  $J^{vote}$  for whether the observed outcome (now a common  $Y_{lj}$ ) is a term outcome or vote outcome, and  $J = |J^{term}| + |J^{vote}|$ . We will denote the systematic components of the vote and term selection as

$$\theta_{lp}^{vote} = c_l^{vote} + b_p^{vote} + \sum_{d=1}^D a_d x_{ld} g_{pd}^{vote}$$

$$\tag{5}$$

$$\theta_{lw}^{term} = c_l^{term} + b_w^{term} + \sum_{d=1}^D a_d x_{ld} g_{wd}^{term}$$
(6)

We will also suppress the superscript for the  $\theta_{lw}^{term}$  and  $\theta_{lp}^{vote}$  while changing to the joint subscript j. The likelihood is given by:

$$\mathcal{L}(\theta_{\cdot\cdot}^{vote}, \theta_{\cdot\cdot}^{term}, \tau, |T_{\cdot\cdot}, V_{\cdot\cdot}) = \prod_{l=1}^{L} \left\{ \left( \prod_{p=1}^{P} Pr\{V_{lp}|\cdot\}^{\frac{W+P}{2P}} \right)^{1-\alpha} \cdot \left( \prod_{w=1}^{W} Pr\{T_{lw} = k|\cdot\}^{\frac{W+P}{2W}} \right)^{\alpha} \right\}$$
(7)

where,

$$Pr\{T_{lw} = k|\cdot\} = \begin{cases} \Phi\left(\theta_{lw}^{term} - \tau_0\right) & T_{lw} = 0\\ \Phi\left(\theta_{lw}^{term} - \tau_k\right) - \Phi\left(\theta_{lw}^{term} - \tau_{k-1}\right) & 0 < T_{lw} \end{cases}$$
(8)

$$Pr\{V_{lp}|\cdot\} = \Phi\left((2V_{lp} - 1)\theta_{lp}^{vote}\right)$$
(9)

The prior structure is given by:

$$c_l^{vote}, b_p^{vote} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, 1)$$
 (10)

$$c_l^{term}, b_w^{term} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, 1)$$
 (11)

$$\mu \sim \mathcal{N}(0,1) \tag{12}$$

$$g_{wd}^{term} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0,4)$$
 (13)

$$g_{pd}^{vote} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0,4) \tag{14}$$

$$\log(\beta_1), \log(\beta_2) \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$$
 (16)

$$\Pr(a_d) = \frac{1}{2\lambda} e^{-\lambda |a_d|} \tag{17}$$

$$\Pr(\lambda) = 1.78e^{-1.78\lambda} \tag{18}$$

Combining the likelihood and prior gives us the posterior:

$$\Pr(\theta_{lj}, \tau, \beta_1, \beta_2 | Y_..) = \left\{ \prod_{\substack{1 \le l \le L \\ 1 \le j \le J}} \Phi\left( (2Y_{lj} - 1)\theta_{lp}^{vote} \right)^{1 \left( j \in \{J^{vote}\} \right) \frac{W + P}{2P} (1 - \alpha)} \\ \times \left\{ \Phi\left( \theta_{lj} - \tau_{Y_{lj}} \right) - \Phi\left( \theta_{lj} - \tau_{Y_{lj} - 1} \right) \right\}^{1 \left( j \in \{J^{term}\} \right) \frac{W + P}{2W} \alpha} \right\} \\ \times \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\mu^2} \times \prod_{1 \le l \le L} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(c_l - \mu)^2} \times \prod_{1 \le j \le J} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(b_j - \mu)^2}$$
(19)  
$$\times \prod_{\substack{1 \le d \le D \\ 1 \le l \le L}} \frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{8}(x_{ld})^2} \times \prod_{\substack{1 \le d \le D \\ 1 \le j \le J}} \frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{8}(g_{jd})^2} \times \prod_{1 \le d \le D} \frac{1}{2\lambda} e^{-\lambda |a_d|} \\ \times \frac{1}{\beta_1 \sqrt{2\pi}} e^{-\frac{1}{2}(\log \beta_1)^2} \times \frac{1}{\beta_2 \sqrt{2\pi}} e^{-\frac{1}{2}(\log \beta_2)^2} \times e^{-1.78\lambda}$$

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:

$$\Pr(Y_{lj} = k | Z_{lj}^*, \theta_{lj}, \tau, \beta_1, \beta_2) = \prod_{\substack{1 \le l \le L \\ 1 \le j \le J}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(Z_{lj}^* - \theta_{lj})^2}$$
(20)

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  $\tilde{\tau}_l$ , such that:

$$d_{\cdot}|\tilde{\tau}_{\cdot}^2 \sim \mathcal{N}(0_D, \tilde{D}_{\tilde{\tau}}) \tag{21}$$

$$\widetilde{D}_{\widetilde{\tau}} = diag(\widetilde{\tau}_1^2, \widetilde{\tau}_2^2, \cdots, \widetilde{\tau}_D^2)$$
(22)

$$\widetilde{\tau}_1^2, \widetilde{\tau}_2^2, \dots, \widetilde{\tau}_D^2 \sim \prod_{1 \le d \le D} \frac{\lambda^2}{2} e^{-\lambda^2 \widetilde{\tau}_d^2/2} d\widetilde{\tau}_d^2$$
(23)

where, after integrating out  $\tilde{\tau}_l^2$ , 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.

#### 1.2.1 Sampling from the Posterior

Next, we outline the MCMC sampler. All conditional posterior densities are conjugate normals except  $\lambda$ ,  $\tilde{\tau}^2$ ,  $\beta_1$ , and  $\beta_2$ . For a derivation of the posterior densities of  $\lambda$  and  $\tilde{\tau}^2$ , see Park and Casella (2008). We fit  $\beta_1$  and  $\beta_2$ , which determine  $\tau$ , using a Hamiltonian Monte Carlo algorithm, but first we describe the Gibbs updates.

The 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:

$$Z_{lj}^{*}| \sim \begin{cases} \mathcal{TN}\left(\theta_{lj}, 1, 0, \infty\right); \ Y_{lj} = 1, \ j \in J^{vote} \\ \mathcal{TN}\left(\theta_{lj}, 1, -\infty, 0\right); \ Y_{lj} = 0, \ j \in J^{vote} \\ \mathcal{TN}\left(\theta_{lj}, 1, \tau_{k-1}, \tau_{k}\right); \ Y_{lj} = k, \ j \in J^{term} \\ \mathcal{N}\left(\theta_{lj}, 1\right); \ Y_{lj} \text{ missing} \end{cases}$$
(24)

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  $\theta_{lj}$  except for  $\tau$ . using a Gibbs sampler, as:

$$\mu | \cdot \sim \mathcal{N}\left(\frac{\sum_{l=1}^{L} \sum_{j=1}^{J} Z_{lj}^{*}}{LJ+1}, \frac{1}{L^{2}J^{2}+1}\right)$$
(25)

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

$$(26)$$

$$b_j | \cdot \sim \mathcal{N}\left(\frac{\sum_{l=1}^L Z_{lj}^*}{L+1}, \frac{1}{L^2+1}\right)$$
 (27)

$$Z_{lj}^{**} = Z_{lj}^* - c_l - b_j + \mu \tag{28}$$

Now, we update  $x_{...}, a_{...}, v_{...}$  from a scaled SVD of  $Z^{**}$ . By a scaled SVD, we mean that the left and right singular vectors are scaled to have sample standard deviation, rather than length, one. Doing so is a standard means of scaling the latent space, and result in a latent scale that is on the same scale as L, J grows.

$$a.|\cdot \sim \mathcal{N}\left(A^{-1}\widetilde{X}^{\top}vec(Z^{**}), A^{-1}\right) \text{ where}$$
$$\widetilde{X} = \left[vec\left(x_{.1}g_{.1}^{\top}\right) : vec\left(x_{.2}g_{.2}^{\top}\right) : \dots : vec\left(x_{.L}g_{.L}^{\top}\right)\right] \text{ and}$$
$$A = \widetilde{X}^{\top}\widetilde{X} + T^{-1} \text{ with } T = diag(\tau_{l}^{2})$$

$$x_{\tilde{l}\tilde{d}}| \sim \mathcal{N}\left(\frac{\sum_{j=1}^{J} Z_{\tilde{l}j,-\tilde{d}}^{**} a_{\tilde{d}} g_{j\tilde{l}}}{\sqrt{\sum_{j=1}^{J} \left(a_{\tilde{d}}^{2} g_{j\tilde{d}}^{2} + \frac{1}{4J}\right)}}, \frac{1}{\sum_{j=1}^{J} \left(a_{\tilde{d}}^{2} g_{j\tilde{d}}^{2} + \frac{1}{4J}\right)}\right)$$
(30)

$$g_{\tilde{j}\tilde{d}}| \sim \mathcal{N}\left(\frac{\sum_{l=1}^{L} Z_{l\tilde{j},-\tilde{d}}^{**} a_{\tilde{d}} x_{l\tilde{d}}}{\sqrt{\sum_{l=1}^{L} \left(a_{\tilde{d}}^{2} x_{l\tilde{d}}^{2} + \frac{1}{4L}\right)}}, \frac{1}{\sum_{l=1}^{L} \left(a_{\tilde{d}}^{2} x_{l\tilde{d}}^{2} + \frac{1}{4L}\right)}\right)$$
(31)

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

$$\tag{32}$$

$$\widetilde{\tau}_d^2 | \cdot \sim InvGauss\left(\sqrt{\frac{\lambda^2}{a_d^2}}, \lambda^2\right)$$
(33)

$$\lambda^2 | \cdot \sim Gamma\left(D+1, \sum_{d=1}^D \widetilde{\tau}_d^2/2 + 1.78\right)$$
(34)

We fix throughout this and the EM implementation

$$\beta_0 = \Phi^{-1} \left( \frac{\sum_{l=1}^L \sum_{j \in J^{term}} \mathbf{1} \left( Y_{lj} = 0 \right)}{L J^{term}} \right)$$
(35)

and we sample  $\beta_1, \beta_2$  using the Hamiltonian Monte Carlo method described below.

#### 1.2.2 EM Implementation of SFA

As we have derived the conditional posterior densities for all of our parameters, above, an EM implementation is straightforward. We treat the spatial locations,  $x_{ld}$  and  $g_{jd}$  as the parameters to be estimated and the remainder as "missing." To begin, we initialize the votes at the inverse Mills ratio and terms after a log transform:

$$\theta_{lj}^{0} = \begin{cases} \frac{\phi(0)}{1 - \Phi(0)} \times (-1)^{Y_{lj}}; & j \in J^{vote}; Y_{lj} \text{ not missing} \\ \log(1 + Y_{lj}); & j \in J^{term} \\ 0; & Y_{lj} \text{ missing} \end{cases}$$
(36)

We also initialize  $\lambda = .5$ ;  $x_{ld}, g_{jd}, a_d$  off a scaled SVD of the double-centered  $\theta$ ;  $\tau_d = 1 \ \forall d$ .

We can now begin the EM algorithm. Note that it is implied that the right hand side of each update below contains the most current update of estimates of all parameters and that all updates occur in turn over l, j, d.

- 1. First E-Step: Updating Z.
  - $Z_{lj}^* \leftarrow \theta_{lj} + \frac{\phi(low_{lj}) \phi(high_{lj})}{\Phi(high_{lj}) \Phi(low_{lj})}$  were  $high_{lj}$  and  $low_{lj}$  are the upper and lower truncation points as given in 24

• 
$$\mu \leftarrow \frac{\sum_{l=1}^{L} \sum_{j=1}^{J} Z_{lj}^{*}}{LJ+1}$$
  
•  $c_{l} \leftarrow \frac{\sum_{j=1}^{J} Z_{lj}^{*}}{J+1}$   
•  $b_{j} \leftarrow \frac{\sum_{l=1}^{L} Z_{lj}^{*}}{L+1}$   
•  $Z_{lj}^{**} \leftarrow Z_{lj}^{*} - c_{l} - b_{j} + \mu$ 

2. *M* step: Updating legislator, term, and vote locations over l, j, d:

$$\begin{array}{l} \bullet \quad \text{Calculate} \ Z_{lj,-\widetilde{d}}^{**} := Z_{lj}^{**} - \sum_{d \neq \widetilde{d}} x_{l\widetilde{d}} g_{jd} a_d \\ \bullet \quad x_{\widetilde{l}\widetilde{d}} \leftarrow \frac{\sum_{j=1}^{J} Z_{\widetilde{l}j,-\widetilde{d}}^{**} a_{\widetilde{d}} g_{j\widetilde{l}}}{\sqrt{\sum_{j=1}^{J} \left(a_{\widetilde{d}}^2 g_{j\widetilde{d}}^2 + \frac{1}{4J}\right)}} \\ \bullet \quad g_{\widetilde{j}\widetilde{d}} \leftarrow \frac{\sum_{l=1}^{L} Z_{l\widetilde{j},-\widetilde{d}}^{**} a_{\widetilde{d}} x_{l\widetilde{d}}}{\sqrt{\sum_{l=1}^{L} \left(a_{\widetilde{d}}^2 x_{l\widetilde{d}}^2 + \frac{1}{4L}\right)}} \end{array}$$

- 3. Second E step: Updating the rest.
  - $a \leftarrow A^{-1} \widetilde{X}^{\top} vec(Z^{**}); A$  defined above

$$\bullet \ \widetilde{\tau}_d^2 \leftarrow \lambda/a_d$$

• 
$$\lambda \leftarrow \frac{L+1}{\sum_{d=1}^{D} \tilde{\tau}_d^2 + 1.78}$$

• Numerically integrate  $\beta_1$ ,  $\beta_2$  to calculate  $\mathbb{E}(\beta_1|\cdot)$ ;  $\mathbb{E}(\beta_2|\cdot)$  using kernel

$$\Pr(\beta_1|\cdot) \propto \prod_{j \in J^{term}} \left\{ \Phi\left(\theta_{lj} - \tau_{Y_{lj}}\right) - \Phi\left(\theta_{lj} - \tau_{Y_{lj}-1}\right) \right\} \times \frac{1}{\beta_1} e^{-\frac{1}{2}(\log\beta_1)^2}$$
(37)

$$\Pr(\beta_2|\cdot) \propto \prod_{j \in J^{term}} \left\{ \Phi\left(\theta_{lj} - \tau_{Y_{lj}}\right) - \Phi\left(\theta_{lj} - \tau_{Y_{lj}-1}\right) \right\} \times \frac{1}{\beta_2} e^{-\frac{1}{2}(\log \beta_2)^2}$$
(38)

#### 1.2.3 The Hamiltonian Monte Carlo Sampler

We have no closed form estimates for the conditional posterior densities of  $\beta_1$  and  $\beta_2$ . 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 ( $\alpha$ ) of a Newton-Raphson step at each. We select  $\alpha$  during the burnin so that the acceptance ratio of proposed ( $\beta_1, \beta_2$ ) is about .4.

Specifically, let  $\widehat{dev}(\beta_1, \beta_2)$  denote the estimate deviance at the point  $(\beta_1, \beta_2)$ . 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  $(\beta_1, \beta_2)$  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:

$$\widehat{H}(\beta_1, \beta_2) = \begin{pmatrix} \widehat{\nabla}_{11} dev(\beta_1, \beta_2) & \widehat{\nabla}_{12} dev(\beta_1, \beta_2) \\ \widehat{\nabla}_{12} dev(\beta_1, \beta_2) & \widehat{\nabla}_{22} dev(\beta_1, \beta_2) \end{pmatrix}$$
(39)

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

$$\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}^+ := \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}^- - \alpha \begin{pmatrix} \widehat{\nabla}_1(\beta_1, \beta_2) \\ \widehat{\nabla}_2(\beta_1, \beta_2) \end{pmatrix}$$
(40)

we instead do updates of the form:

$$\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}^+ := \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}^- - \alpha \times \left\{ \hat{H}(\beta_1, \beta_2) \right\}^{-1} \begin{pmatrix} \hat{\nabla}_1(\beta_1, \beta_2) \\ \hat{\nabla}_2(\beta_1, \beta_2) \end{pmatrix}$$
(41)

where the Hessian is updated every third update of the parameters during the and burnin period and afterwards fixed at the average of the last ten steps. The step length parameter  $\alpha$  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 during the burnin-in. After the burn-in period, the acceptance rate levels off around 45%. We implement twenty steps in order to produce a proposal.

#### 1.2.4 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, large values of a and b in magnitude eturns values of 1 or 0, leaving it impossible to evaluate the logarithm.

Extrapolating from the observed values yields the linear approximation:

$$l(a,b) = \begin{pmatrix} 1 \\ a \\ b \\ a^{2} \\ b^{2} \\ \log(|a-b|) \\ \{\log(|a-b|)\}^{2} \\ ab \end{pmatrix}^{\top} \gamma$$
(42)

where

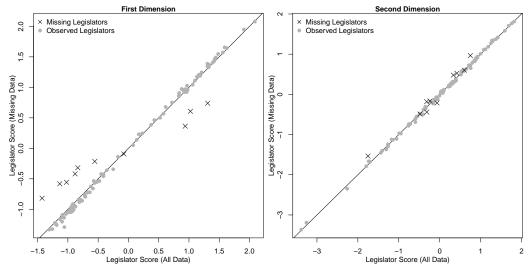


Figure 1: Estimated Ideal Points for Ten Legislators Missing at Random. The lefthand panel compares the censored and uncensored estimates (marked by X's) of the preferred outcomes for the ten randomly censored legislators on the first dimension, while the righthand panel makes the analogous comparison for dimension two.

$$\gamma = \begin{pmatrix} -1.82517672\\ 0.51283415\\ -0.81377290\\ -0.02699400\\ -0.49642787\\ -0.33379312\\ -0.24176661\\ 0.03776971 \end{pmatrix}$$
(43)

We derived the values for  $\gamma$  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).

## 2 Supplemental Materials

We offer tests for the internal and external validity of SFA. The leadership imputation exercise appears in the main body as well, but we include it in here for completeness.

### 2.1 Internal Validity

**Imputing estimates for legislators missing completely at random.** First, we randomly discard the votes cast by ten legislators selected completely at random, coding *all* of their votes as "missing,", while we maintain all of their speech data. The left and right panels of Figure 1 plot the imputed versus fitted values (X) for the dropped legislators, for the first (left) and second (right) dimension. SFA recovers

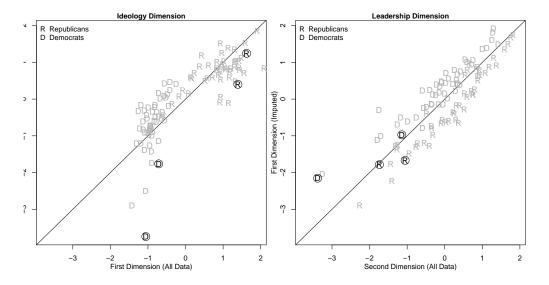


Figure 2: Estimated Ideology when Only Leaders Votes are Informative. The voting dimension estimates appear in the left panel, with the censored estimates measured on the vertical (y-axis) while the uncensored ones appear on the horizontal (x-axis). In the censored data the salience of the voting dimension drops, so that it becomes the second dimension. The righthand panel exhibits the leadership dimension, again the censored estimates correspond with the vertical (y-axis) and the uncensored ones coincide with the horizontal (x-axis).

reliable first-dimension preferred outcomes well, except for some expected attenuation bias. The second dimension ideal points are recovered almost exactly. We remind the reader that the first SFA dimension coincides closely with the dimension that emerges from an analysis of the votes alone, and so we might expect it to be more affected by the loss of voting data, while the accuracy of our second dimension estimates, which are dominated by speech data, would be expected to suffer less from the censorship of the votes.

Imputing estimates for members' given only votes from leadership. We next offer a more challenging test of internal validity. For this analysis, we coded all vote data except for the party leaders and whips as missing, while maintaining all speech data. This left a vote record for less than 4% of the Senate. We then compared the SFA ideal point estimates to the SFA estimates using everyone's speech, but only leaders' votes. Essentially our exercise in censorship diminishes the importance of the dimension related to voting. Figure 2 shows that we again recover two dimensions based on the censored data, although their order is now reversed. Specifically, as the voting dimension becomes noisier, it falls into second place, while the leadership dimension, the evidence for which comes almost entirely through legislative speech, earns the higher dimension weight. The left panel of Figure 2 compares estimates for the voting dimension, which is the second dimension estimated with the heavily censored data (plotted along the vertical y-axis) while it corresponds with the first dimension of the uncensored estimates (graphed relative to the horizontal x-axis). Observations are labeled by party, and leaders' locations are in bold and circled. As one would expect, with less than 1/25 of the voting data, recovery of the first dimension is far from perfect, but remarkably the imputed scores correlate highly, at more than 0.85, with the estimates

based on the full data set. The right hand panel compares estimates for the "leadership" dimension, which coincides with the first dimension based on the censored data, but with the second dimension based on the uncensored data set. In contrast with the voting dimension, the censored estimates correspond closely with their uncensored counterparts. Of course, the "leadership" dimension is driven mostly by words, and we did not censor those.

While this last exercise may seem a stunt, we note that in heavily whipped parliaments most legislators vote their parties, rather than their preferences (e.g., Kellerman, 2012), yet they still give speeches. In such settings we might use SFA to "bridge" between speeches actually given by members of a parliament to the votes that they would have cast had they not been "whipped," anchoring the exercise by treating the votes of party principals as a genuine reflection of the leaders' preferences.

## 2.2 External Validity

Lastly, we consider assessing SFA's ability to recover preference estimates with external validity. So far, we have used SFA to impute from legislative members to other members during the same session. We next turn from imputing ideal points for legislative members to imputing ideal points for non-legislative actors, namely newspaper editorials.<sup>1</sup>

We apply SFA to word count data from unsigned editorials published during the two years that the  $112^{th}$  Congress was in session in the *New York Times*, the *Wall Street Journal*, and the *Washington Post*, using the same terms we employed in our analysis of the Senate. As above, we combine the word counts of these editorials with the Senate data, treating the editorials as legislators with a missing vote record.

As the term data come from different venues, the Senate floor versus the editorial page, the exercise is one of "out of sample prediction." This leaves us with the question of whether the political meanings of the terms of discourse are the same in both venues. As a first approach to this issue, we treat the ideal points for both groups as coming from a mean-zero distribution. Results appear in Figure 3. We orient the dimension so that the Republicans have a positive value. The densities for the Republican and Democratic Senators are in the background, and the voting dimension legislator preferred outcomes are plotted as hatch marks along the x-axis. The results are largely as expected. If we treat the three sets of editorial boards as legislators who do not vote, we find the Wall Street Journal (**WSJ**) to the right of the *Washington Post* (**Wash Post**) and the New York Times (**NYT**) to its left. The distance between the *Wall Street Journal* and the *Washington Post* is about half the estimated distance between the New York Times and the Post.

#### 2.3 Descriptive Statistics

A summary of our data can be found in Table 1. The first three columns report the session of the Senate, the number of distinct Senators who served, and the total number of votes cast. A "term" is a stemmed unigram or bigram. We restrict our attention to frequent terms, excluding "stop words" see the text for

 $<sup>^{1}</sup>$ We used the Factiva database to download the editorials from the three newspapers published from January 2011 to December 2012 corresponding to the 112th Congress.

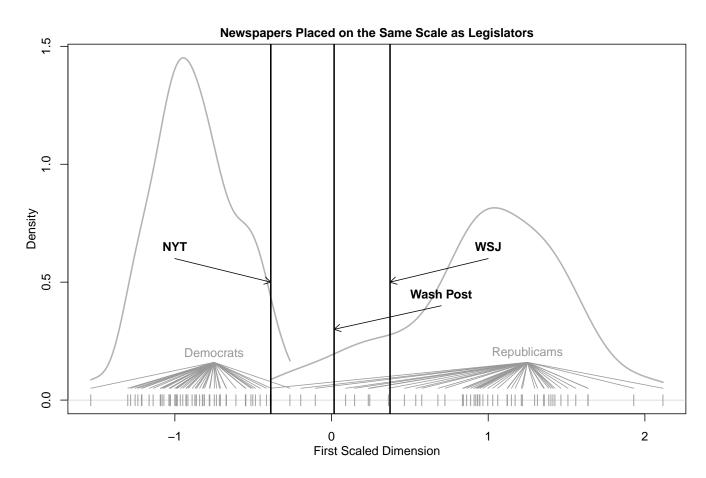


Figure 3: Scaling newspaper editorials given only their text. This figure presents the relative locations and differences between the ideal points for legislators and newspapers.

details.

			Total 7	Terms	Frequent Terms								
Senate	Senators	Votes	Number	% Zero	Number	%Zero	Minimum	Maximum					
105	100	613	2880504	97.8	3168	10.7	0	2838					
106	102	670	940615	94.6	3804	13.9	0	2995					
107	100	633	765696	94.6	3155	13.3	0	4993					
108	100	676	813635	94.6	3384	13.5	0	3702					
109	101	645	788693	94.7	3350	13.7	0	3816					
110	102	658	846720	94.6	3711	15.2	0	5211					
111	110	697	783331	95.0	3473	19.9	0	3494					
112	101	487	572247	94.7	2532	15.1	0	2946					

Table 1: **Data summary by Senate**. The first three columns report the session of the Senate, the number of distinct Senators who served, and the total number of votes cast. A "term" is a stemmed unigram or bigram. We restrict our attention to frequent terms, excluding "stop words" see the text for details.

## 2.4 Selection of $\alpha$ based on Empirical Criteria

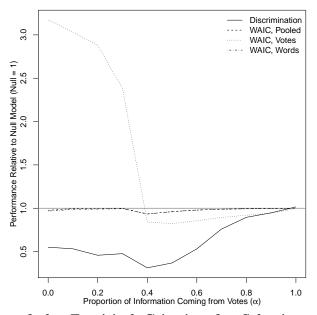


Figure 4: **Performance of the Empirical Criterion for Selecting**  $\alpha$ , **averaged over.** We consider several model fit criteria for choosing  $\alpha$ : our discrimination statistic, the WAIC for word data, WAIC for vote data, and WAIC for all data. All results are scaled relative to those from a null model, denoted by the horizontal line at 1. The criteria were evaluated at  $\alpha \in \{0.00, 0.09, 0.15, 0.24, 0.36, 0.50, 0.64, 0.76, 0.85, 0.91, 1.00\}$ . The WAIC for words and for the whole data are difficult to differentiate as they agree quite closely, given the preponderence of words in the data. There is a stable maximum around 0.36 across years, and the criteria is reasonably concave with a well-defined extremum.

Words with Negative Level	Words with Positive Level
ordered rise today hud fort today introduc thank chairman credit union without objection objection order <b>ef</b>	Dimension 1 can time bill say year one amend want want wark
ordered appropri rise today ask hud act fort author today introduc section thank chairman consent credit union ask unanim without objection unanim objection orderethanim consent nasa committe	Dimension 2 famili peopl children tobacco health school percent care cut tax
billion balanc budget fund social secur spend feder balanc social budget tax	nato think children know school said leader week hope say
said attorney judiciari debt law unit judg constitut	Dimension 4 health educ care unanim consent ask unanim consent health care fda school
school student provid medic protect women public health children	Dimension 5 say get thing tax think said billion want now
drug think tobacco balanc budget nomine judciari law court judg	Dimension 6 militon nuclear unit state trade year north dakota nation billion

# Figure 5: Word Dimensions from 105th Senate

Words with Negative Level										_\	Vo	rds	wi	th F	Pos	itiv	e L	eve	el →		
can	time	want	year	one	bill	say	work	amend	peopl		thank chairman	clean water	fish wildlif	harvest	museum	taiwan	nativ american	hawaii	habitat	risk manag	Dimension 1
health	school	care	children	prescript	famili	medicar	cut	drug	prescript drug		chief	motion	lewinski	ask	committe	proceed	ask unanim	unanim	unanim consent	consent	2 Dimension
partialbirth	foc	1aa	ergonom	stockpil	irrig	pharmaci	sheriff	nuclear	basebal		consum protect	900	troops	motion lay	nasa	credit card	secur medicare	foster care	bankruptci bill	creditor	Umension 3
tax	budget	social	secur	social secur	surplus	billion	spend	energi	bill		victim	women	violenc	juvenil	children	justic	judg	court	gun	crime	Dimension 4
act	provid	program	servic	includ	legisl	section	effect	health	provis		know	tax	peopl	thing	said	get	talk	think	want	say	Cilinension o
farmer	bankruptci	tax	farm	enforc	milk	bank	attorney	law	crime		billion	friend	student	arm	defens	million	school	nation	nuclear	militari	

# Figure 6: Word Dimensions from 106th Senate

Words with Negative Level	Words with Positive Level
legisl introduc without objection scout colonel objection ordered soldier madam rise pay tribut rise today girl scout	Dimension 1 bill can time say peopl vote one year want
thing farm talk peopl want think get say tax	Dimension 2 act court judg author section includ nomine general law confirm
judiciari suprem court drug confirm nomine generic generic judg andsan bahkruptici	emiss secur librari bulget shall fiscal oreningi visa
weapon distinguish homeland iraq missil say said defens constitut war	Dimension 4 health program school educ care m children insur provid busi patient
law tax bill year percent drug provid state	4 Dimension 5 D motion proceed motion reconsid reconsid laid table nevada quorum laid
men thank live student day educ women teacher school	farmer farmer bill agricultur d committe tax farm bill nomine circult export

# Figure 7: Word Dimensions from 107th Senate

Words with Negative Level	Words with Positive Level
zone motion fort vote pleasur proceed rise today leader infantri morn thank chairman ask iraqi freedom unanim ordered consent without objection ask unanim objection orderethanim consent	Dimension 1 bill time one year can amend peopl work say vote
motion vote leader morn ask unanim consent ask unanim anim consen	Dimension 2 feder administr program nation act provid shall requir includ support
benefit job busi prescript senior heath care care medicar heath	Intellig nuclear iraq weapon secur defens war money amend depart
circuit judici suprem court case law justic nomine judg court	Dimension 4 amend secur intellig budget billion bill appropri program report nation
tax cut talk job pay cut say tax	Dimension 5 nomin committe judg act confirm circuit proceed court consider judici
section gas feder author meet market committe trade tax bill energi	iraq war administr famili bush american women day school men

# Figure 8: Word Dimensions from 108th Senate

Words with Negative Level	Words with Positive Level
heritag section coastal author natur resourc committe ordered consider rise today motion fort unanim consent objection ordered ask unanim call roll unanim clerk call proceed	Dimension 1 bill time peopl year can amend one work american vote
heritag section coastal author natur resourc committe ordered consider rise today motion fort unanim consent without objection consent bijection ordered ask unanim call roll unanim celerk call proceed	Dimension 2 tax pay social money get cut say said country want
rriend love busi today help care ilfe life famili	Dimension 3 court law amend boder judg think vode limmigr illeg bill
ask unanim spend unanim consent deficit get money a budget una billion tax	3 Dimension 4 court iraq judg justic republican women protect constitut judiciari
- floor come said said vote unanim ask unanim unanim consent consent ask	Dimension 5 program act provid health state shall section includ fund requir
• health care wage insur justic republican health cut budget tax	State state unit iraq said nation bill secretari oil shall rememb

# Figure 9: Word Dimensions from 109th Senate

Words with Negative Level ←	Words with Positive Level
get work american time year bill	Dimension 1 pay tribut brigad qualiti life rise today idaho wyom furthermore fort aviat foster
support state ensur legisl law nation court provid	Dimension 2 get thing say lot talk want tax oil
live iraqi worker militari alqaida countri american peopl war iraq	tax tax committe medicar busi amt budget financ revenu clotur altern
veteran educ food help school famili care children health	Dimension 4 court general tax judg intellig congress terrorist law amend confirm re
care pay billion budget program increas drug percent tax health	n 4 Dimension 5 Dimensi iraq school war children motion peopl proceed get clotur student s friend drug t leader iowa republican leader republican 
barrel tax gas market indian trade price energj	Dimension 6 school children get student drug lowa fda right right

# Figure 10: Word Dimensions from 110th Senate

Words with Negative Level											\	Noi	rds	wi	th F	Pos	itiv	e L	eve	el 		
, call roll	endur	madam rise	pay tribut	public servic	fort	rise today	sergeant	nativ	tribut			make	american	health	can	time	get	one	year	bill	peopl	Dimension 1
spend	tax	debt	trillion	say	money	medicar	billion	think	get		•	support	help	also	justic	nation	serv	nomin	court	provid	act	Dimension 2
republican	health	health care	care	judg	democrat	medicar	court	insur	nomine		•	fund	indian	nuclear	issu	oi	trade	credit	bank	energi	financi	Dimension 3
vote	thing	nomin	think	judg	said	clotur	proceed	republican	tri.		•	insur	bill	provid	tax	busi	health care	small busi	care	small	health	Dimension 4
feder	govern	congress	court	law	unit	billion	constitut	spend	judg			morn	motion reconsid	famili	motion	unemploy	work	ask unanim	get	insur	job	Dimension 5
amend	motion	nevada	budget	minut	tax	leader	follow	bill	control			bank	money	judg	bush	compani	justic	right	peopl	countri	court	Dimension 6

# Figure 11: Word Dimensions from 111th Senate

Words with Negative Level	Words with Positive Level	
american time can bill	today wish madam rise rise today army colleagu support deploy air forc resources recognit legaci	Dimension 1
money peopl get tax think budget trillion debt spend	nomin district protect women judici court confirm nation nation support	Dimension 2
economi legisl agreement econom trade small busi job tax busi	say court money want peopl judg judg iudici tell circuit	Dimension 3
motion nomine leader consid judg obama clotur confirm r nomin	health student care school work compani famili job children million	Dimension 4
on table ne laid er ask unanim id action debate laid upon na proceed in motion reconsid in motion	feder law govern spend congress report administr state	Dimension 5
famili school loan colleg dream fee bank card debt	amend say say state	Dimension 6

# Figure 12: Word Dimensions from 112th Senate

# 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.
- Kellerman, Michael. 2012. "Estimating Ideal Points in the British House of Commons Using Early Day Motions." American Journal of Political Science 56(3):757–771.
- 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. Vol. 2 of CRC Handbooks of Modern Statistical Method Chapman and Hall pp. 113–162.
- Park, Trevor and George Casella. 2008. "The bayesian lasso." Journal of the American Statistical Association 103(482):681–686.