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":

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

and "Nay" alternatives:

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

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:

$$\begin{aligned}
V_{lp}^* &= U_l^{vote}\left(Aye; \{x_{ld}\}_{d=1}^D, \{z_{pd}^{aye}\}_{d=1}^D\right) - U_l^{vote}\left(Nay; \{x_{ld}\}_{d=1}^D, \{z_{pd}^{nay}\}_{d=1}^D\right) \\
&= -\frac{1}{2}\sum_{d=1}^D a_d(z_{pd}^{aye} - x_{ld})^2 + \tilde{\xi}_{lp}^{aye} - \left(-\frac{1}{2}\sum_{d=1}^D a_d(z_{pd}^{nay} - x_{ld})^2 + \tilde{\xi}_{lp}^{nay}\right) \\
&= \sum_{d=1}^D \frac{a_d}{2}(z_{pd}^{nay^2} - z_{pd}^{aye^2}) + \sum_{d=1}^D \left(\frac{a_d}{2}\cdot 2x_{ld}(z_{pd}^{aye} - z_{pd}^{nay})\right) + \tilde{\xi}_{lp}^{aye} - \tilde{\xi}_{lp}^{nay} \\
&= \left(\sum_{d=1}^D \frac{a_d}{2}(z_{pd}^{nay^2} - z_{pd}^{aye^2}) + E\{\tilde{\xi}_{lp}^{aye}\} - E\{\tilde{\xi}_{lp}^{nay}\}\right) + \sum_{d=1}^D a_d x_{ld}(z_{pd}^{aye} - z_{pd}^{nay}) \\
&\quad - \left(\tilde{\xi}_{lp}^{nay} - \tilde{\xi}_{lp}^{aye} + E\{\tilde{\xi}_{lp}^{aye}\} - E\{\tilde{\xi}_{lp}^{nay}\}\right)
\end{aligned} \tag{3}$$

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),

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

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

$$\begin{aligned}
V_{lp}^{*} &= U_l(\{x_{ld}\}_{d=1}^{D}, \{p_d\}_{d=1}^{D}) - U_l(\{x_{ld}\}_{d=1}^{D}, \{q_d\}_{d=1}^{D}) \\
&= \underbrace{\left(\sum_{d=1}^{D} \frac{a_d}{2}({z_{pd}^{nay}}^2 - {z_{pd}^{aye}}^2) + E\{\tilde{\xi}_{lp}^{aye}\} - E\{\tilde{\xi}_{lp}^{nay}\}\right)}_{c_l^{vote}+b_p^{vote}} + \sum_{d=1}^{D} a_d x_{ld} \underbrace{(z_{pd}^{aye} - z_{pd}^{nay})}_{g_{pd}^{vote}} \\
&\qquad - \underbrace{\left(\tilde{\xi}_{lp}^{nay} - \tilde{\xi}_{lp}^{aye} + E\{\tilde{\xi}_{lp}^{aye}\} - E\{\tilde{\xi}_{lp}^{nay}\}\right)}_{\epsilon_{lp}^{vote}} \\
&= c_l^{vote} + b_p^{vote} + \sum_{d=1}^{D} a_d x_{ld} g_{pd}^{vote} - \epsilon_{lp}^{vote}. \qquad (4)
\end{aligned}$$

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

$$\theta_{lp}^{vote} = c_l^{vote} + b_p^{vote} + \sum_{d=1}^{D} a_d x_{ld} g_{pd}^{vote} \qquad (5)$$

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

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

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

where,

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

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

The prior structure is given by:

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

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

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

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

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

$$x_{ld} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 4) \tag{15}$$

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

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

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

Combining the likelihood and prior gives us the posterior:

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

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:

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

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

$$
\begin{aligned}
d.|\widetilde{\tau}^2_{\cdot} &\sim \mathcal{N}(0_D, \widetilde{D}_{\widetilde{\tau}}) &(21)\\
\widetilde{D}_{\widetilde{\tau}} &= diag(\widetilde{\tau}_1^2, \widetilde{\tau}_2^2, \cdots, \widetilde{\tau}_D^2) &(22)\\
\widetilde{\tau}_1^2, \widetilde{\tau}_2^2, \ldots, \widetilde{\tau}_D^2 &\sim \prod_{1 \leqslant d \leqslant D} \frac{\lambda^2}{2} e^{-\lambda^2 \widetilde{\tau}_d^2/2} d\widetilde{\tau}_d^2 &(23)
\end{aligned}
$$

where, after integrating out $\widetilde{\tau}_l^2$, 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 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 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:

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

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

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

$$
c_l|\cdot \sim \mathcal{N}\left(\frac{\sum_{j=1}^{J} Z^*_{lj}}{J+1}, \frac{1}{J^2+1}\right) \tag{26}
$$

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

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

Now, we update $x_{\cdot\cdot}$, $a_{\cdot}$, $v_{\cdot\cdot}$ 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.

$$
\begin{aligned}
a.|\cdot &\sim \mathcal{N}\left(A^{-1}\widetilde{X}^\top vec(Z^{**}), A^{-1}\right) \text{ where} \\
&\widetilde{X} = \left[vec\left(x_{\cdot 1}g_{\cdot 1}^\top\right) : vec\left(x_{\cdot 2}g_{\cdot 2}^\top\right) : \ldots : vec\left(x_{\cdot L}g_{\cdot L}^\top\right)\right] \text{ and} \\
&A = \widetilde{X}^\top\widetilde{X} + T^{-1} \text{ with } T = diag(\tau_l^2)
\end{aligned} \tag{29}
$$

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

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

$$
Z^{**}_{lj,-\tilde{d}} = Z^{**}_{lj} - \sum_{d \neq \tilde{d}} x_{l\tilde{d}} g_{jd} a_d \tag{32}
$$

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

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

We fix throughout this and the EM implementation

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

and we sample β_1, β_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:

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

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

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:

   - Calculate $Z^{**}_{lj,-\tilde{d}} := Z^{**}_{lj} - \sum_{d \neq \tilde{d}} x_{l\tilde{d}} g_{jd} a_d$
   - $x_{\tilde{l}\tilde{d}} \leftarrow \frac{\sum_{j=1}^{J} Z^{**}_{\tilde{l}j,-\tilde{d}} a_{\tilde{d}} g_{j\tilde{l}}}{\sqrt{\sum_{j=1}^{J}\left(a^2_{\tilde{d}} g^2_{j\tilde{d}} + \frac{1}{4J}\right)}}$
   - $g_{\tilde{j}\tilde{d}} \leftarrow \frac{\sum_{l=1}^{L} Z^{**}_{l\tilde{j},-\tilde{d}} a_{\tilde{d}} x_{l\tilde{d}}}{\sqrt{\sum_{l=1}^{L}\left(a^2_{\tilde{d}} x^2_{l\tilde{d}} + \frac{1}{4L}\right)}}$

3. Second E step: Updating the rest.

   - $a \leftarrow A^{-1}\widetilde{X}^{\top} vec\,(Z^{**})$; A defined above
   - $\tilde{\tau}^2_d \leftarrow \lambda / a_d$
   - $\lambda \leftarrow \frac{L+1}{\sum_{d=1}^{D}\tilde{\tau}^2_d + 1.78}$
   - Numerically integrate β_1, β_2 to calcualte $\mathbb{E}(\beta_1|\cdot)$; $\mathbb{E}(\beta_2|\cdot)$ using kernel

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

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

1.2.3 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 during the burnin 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:

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

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

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

we instead do updates of the form:

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

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 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 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 magnitudereturns values of 1 or 0, leaving it impossible to evaluate the logarithm.

Extrapolating from the observed values yields the linear approximation:

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

where

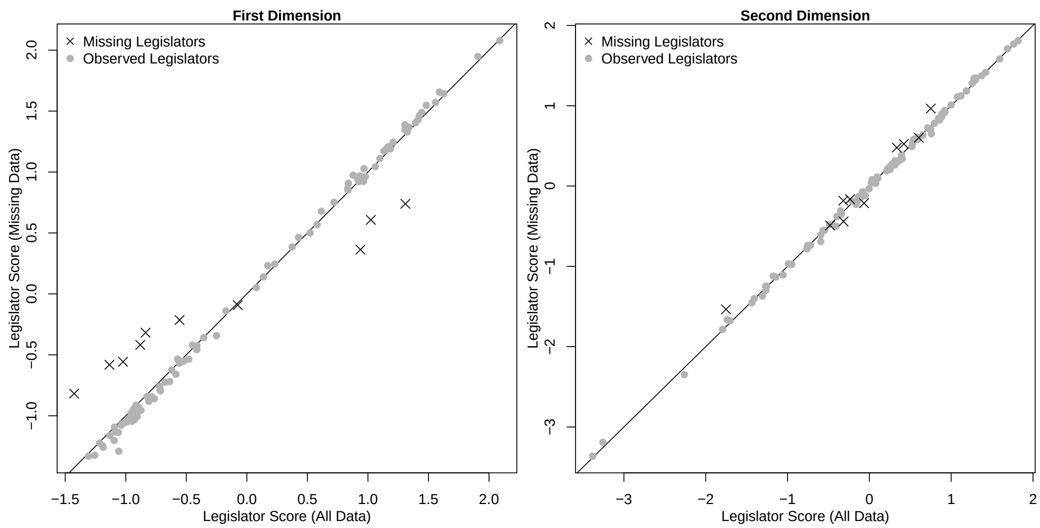


Figure 1: **Estimated Ideal Points for Ten Legislators Missing at Random**. The lefthand panel compares the censored and uncensored estimates (marked by X's) of the preferred outcomes for the ten randomly censored legislators on the first dimension, while the righthand panel makes the analogous comparison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

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

We derived the values for γ from fitting a model over the range $4 \leqslant b < a \leqslant 8$. 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

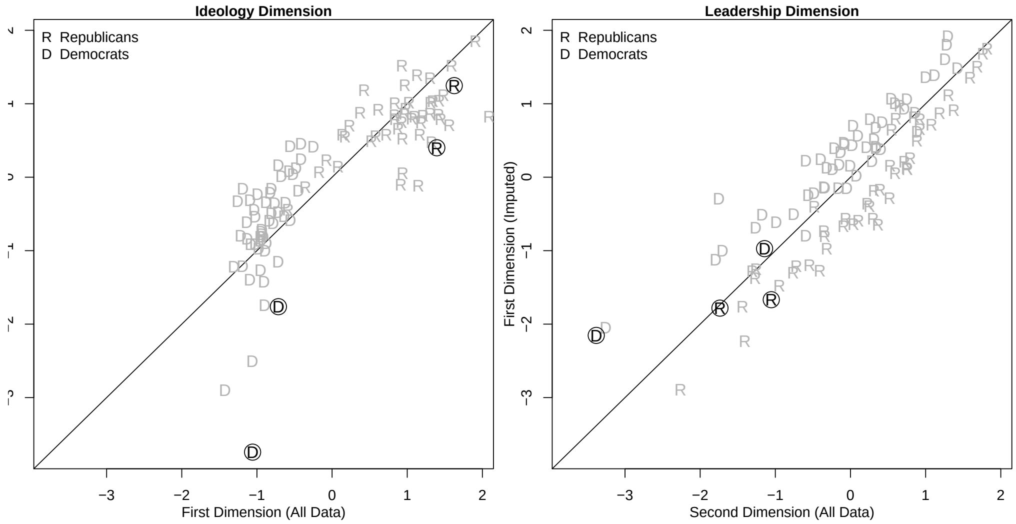


Figure 2: **Estimated Ideology when Only Leaders Votes are Informative.** The voting dimension estimates appear in the left panel, with the censored estimates measured on the vertical (y-axis) while the uncensored ones appear on the horizontal (x-axis). In the censored data the salience of the voting 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.[1]

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

[1] 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.

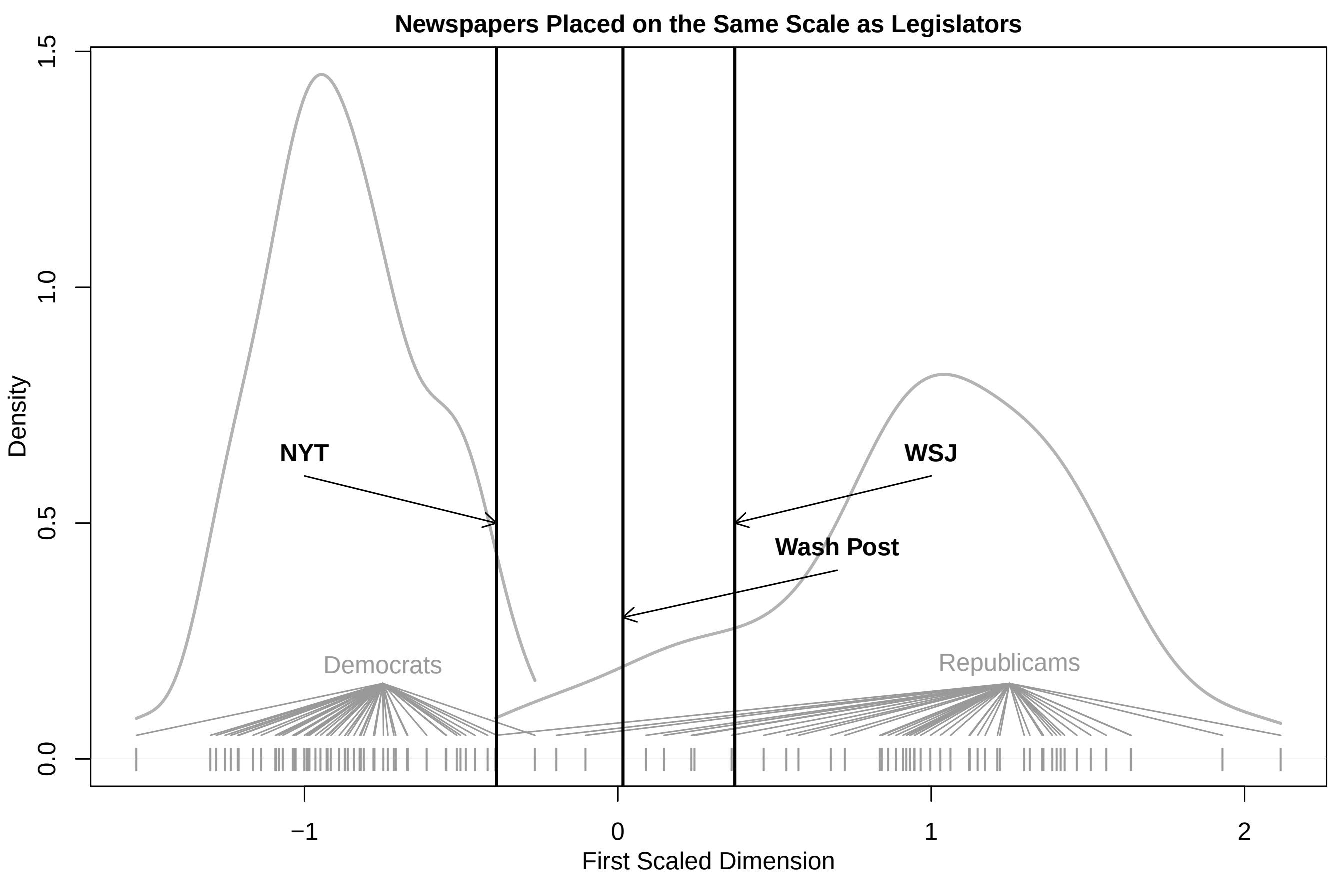


Figure 3: **Scaling newspaper editorials given only their text.** This figure presents the relative locations and differences between the ideal points for legislators and newspapers.

details.

			Total 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 α based on Empirical Criteria

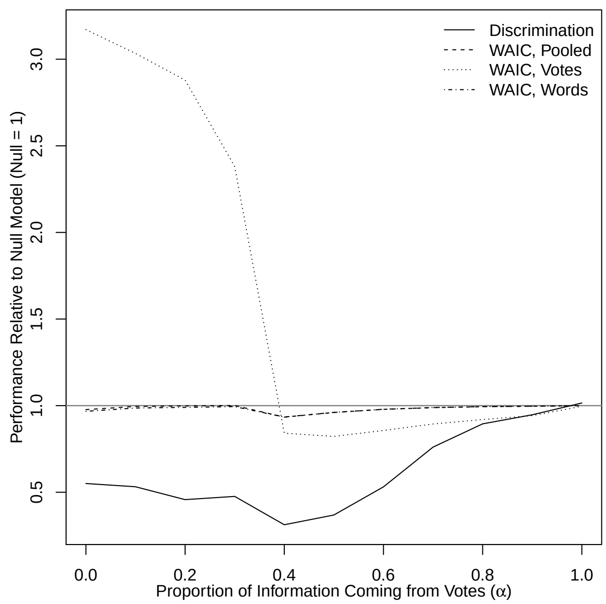


Figure 4: **Performance of the Empirical Criterion for Selecting α, averaged over.** We consider several model fit criteria for choosing α: our discrimination statistic, the WAIC for word data, WAIC for vote data, and WAIC for all data. All results are scaled relative to those from a null model, denoted 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.

Figure 5: **Word Dimensions from 105th Senate**

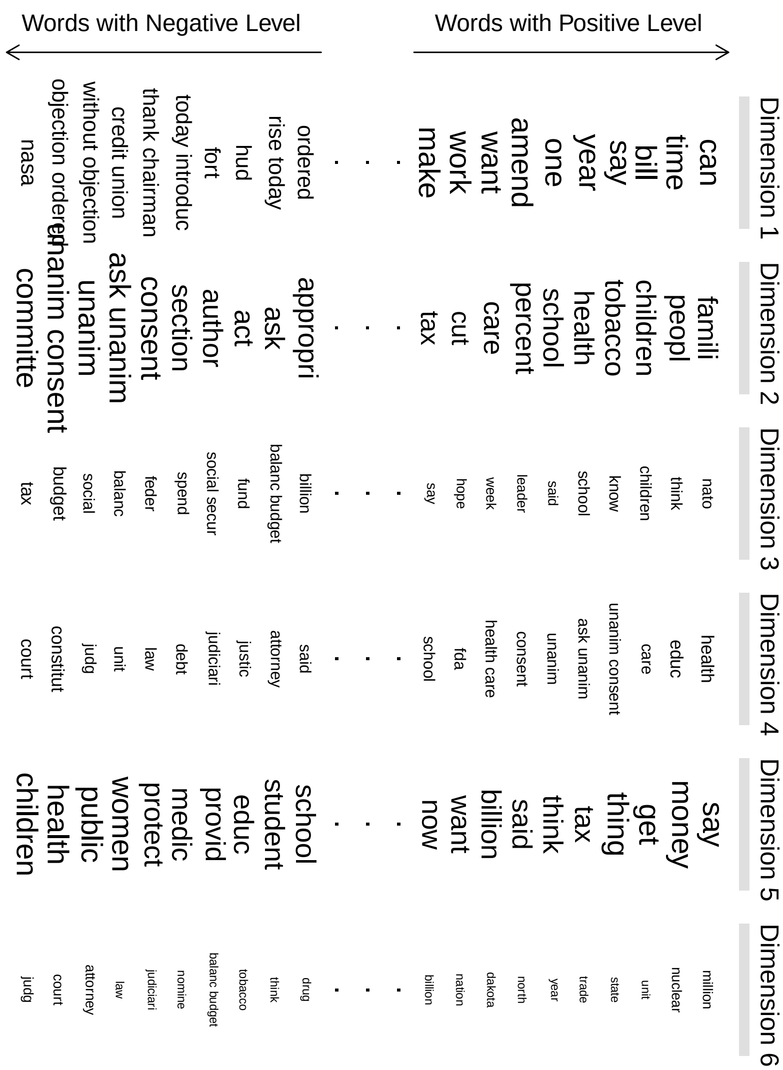


Figure 6: **Word Dimensions from 106th Senate**

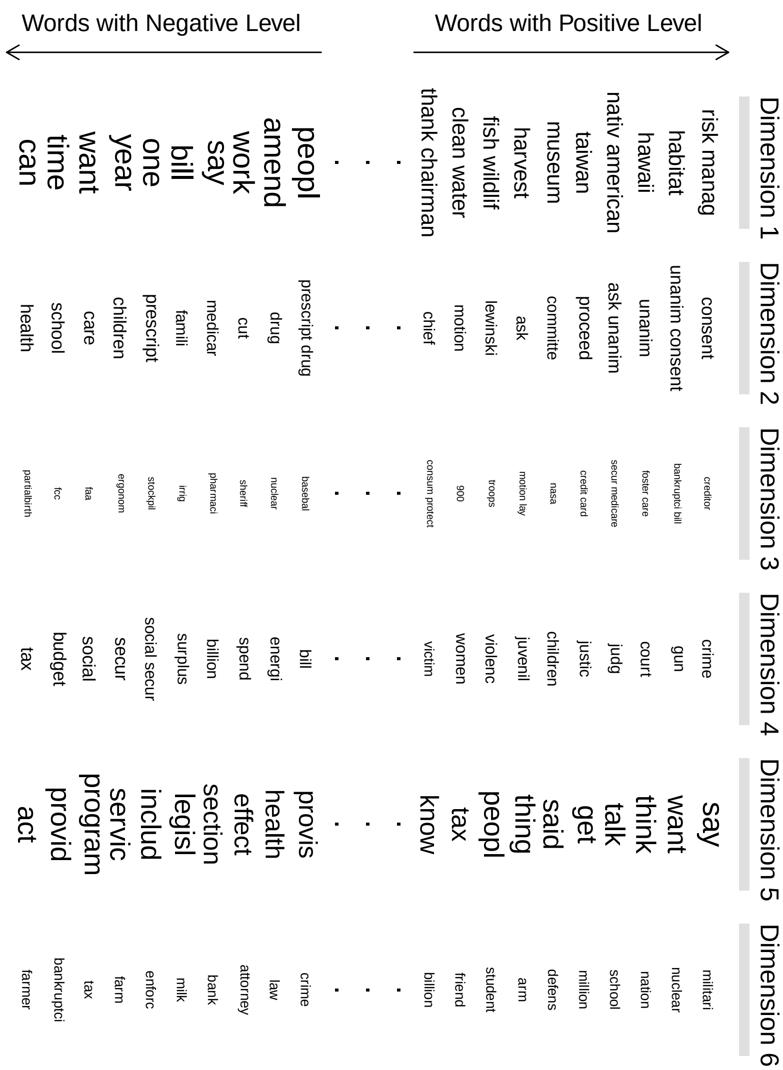


Figure 7: **Word Dimensions from 107th Senate**

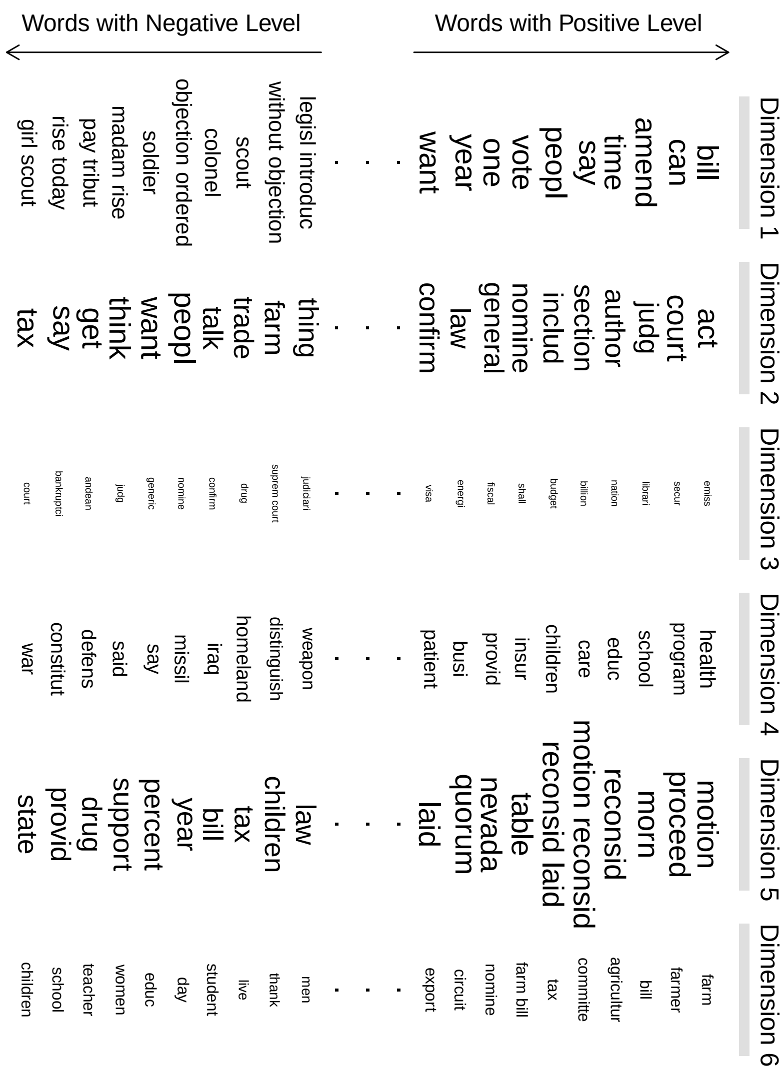


Figure 8: **Word Dimensions from 108th Senate**

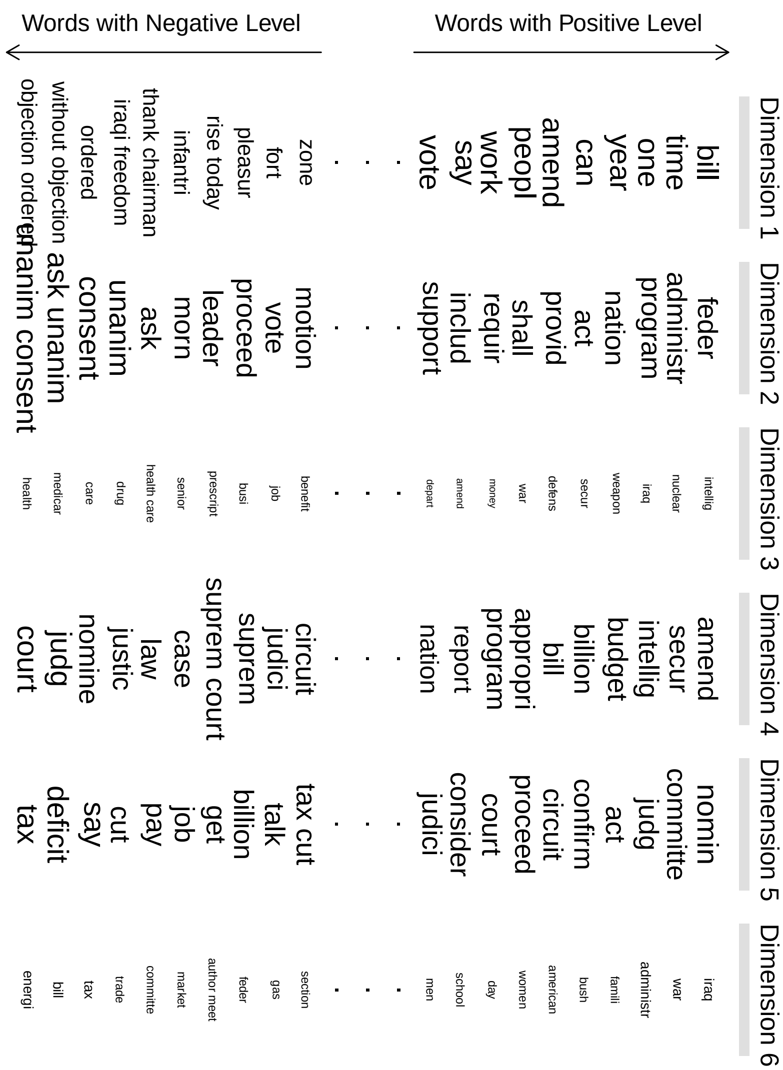


Figure 9: **Word Dimensions from 109th Senate**

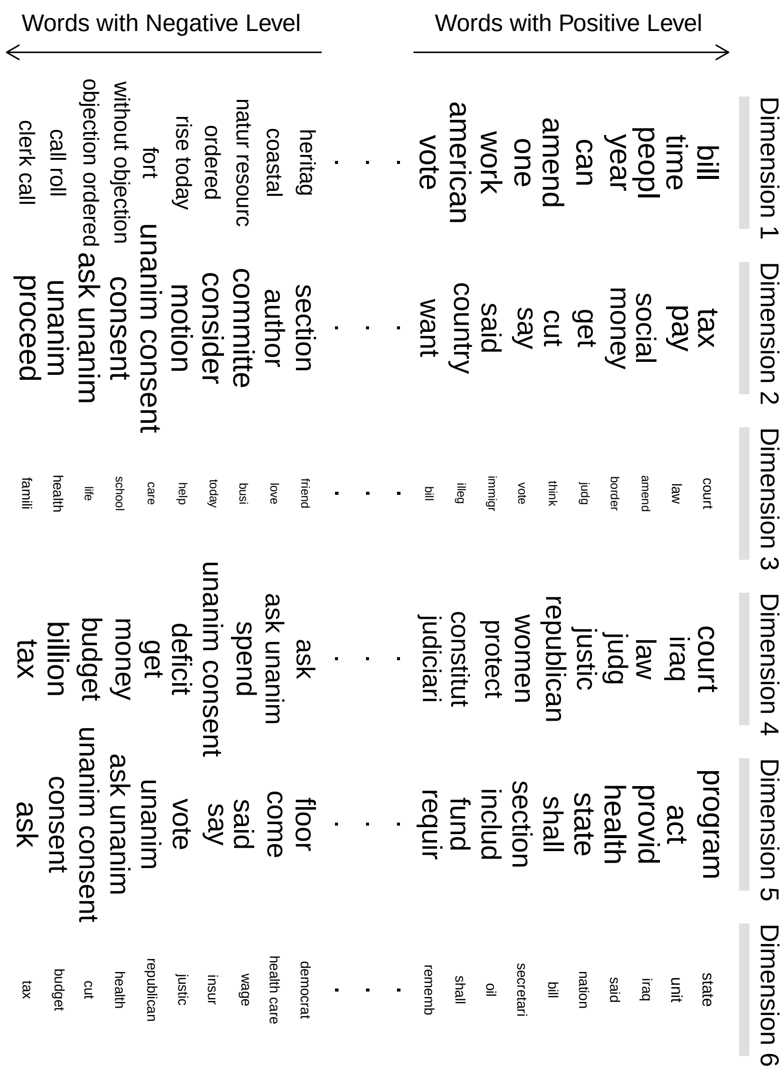


Figure 10: **Word Dimensions from 110th Senate**

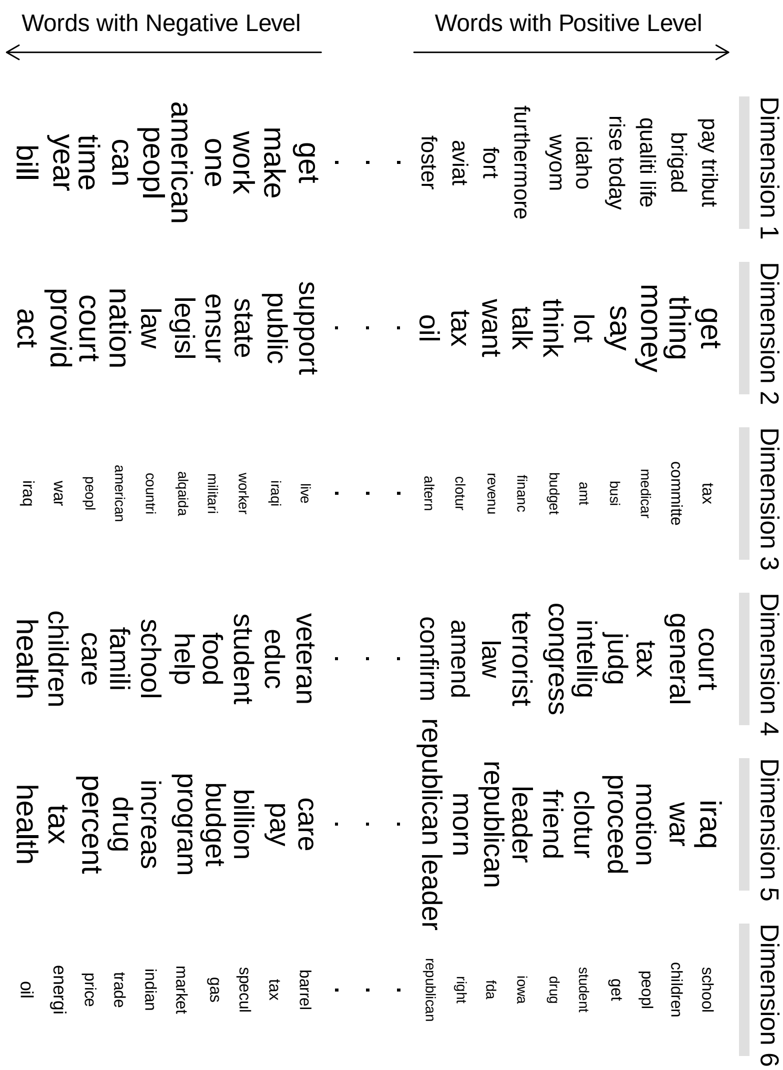


Figure 11: **Word Dimensions from 111th Senate**

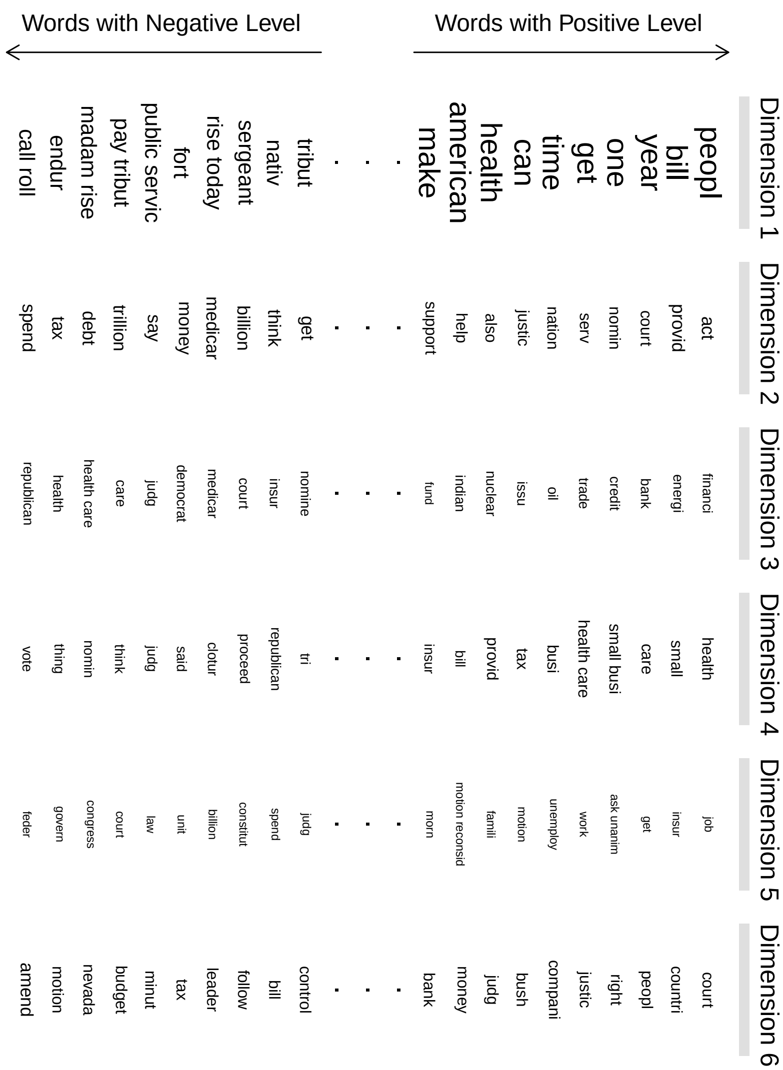
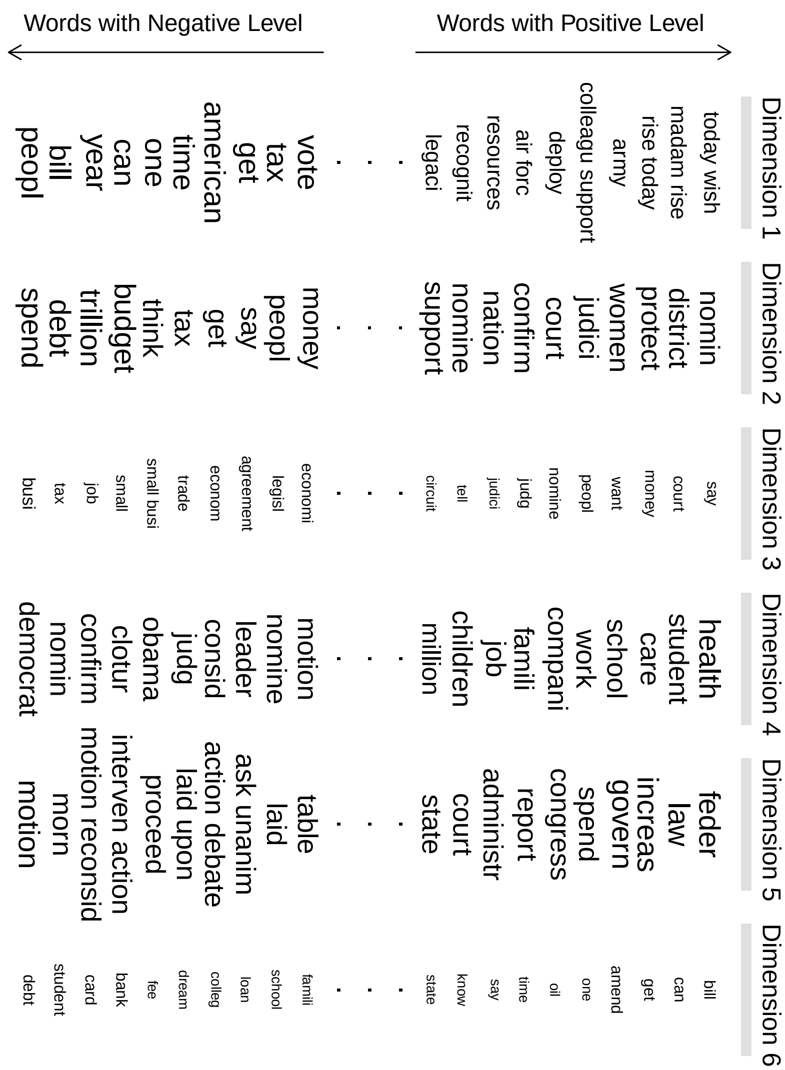


Figure 12: **Word Dimensions from 112th Senate**



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 Commoms 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.