

Appendix A: Descriptive Statistics

	Mean	Std. Deviation	Minimum	Maximum
Age	53.14	9.21	21.78	81.23
Gender	0.33	0.47	0.00	1.00
Left-Right Ideology	-0.09	0.48	-0.81	0.82
Integration Ideology	0.23	0.53	-1.00	1.00
Left-Right Distance	0.05	0.07	0.00	0.85
Integration Distance	0.14	0.22	0.00	1.31
Group Leader	0.07	0.25	0.00	1.00
Natn'l Election Time	2.34	1.30	0.00	5.00
Majoritarian	0.28	0.45	0.00	1.00
Natn'l Party Seat %	29.32	17.36	0.00	63.90
Natn'l Party in Gov't	0.50	0.50	0.00	1.00
Integration Bill	0.23	0.42	0.00	1.00
Absolute Majority	0.11	0.31	0.00	1.00
Group RCV Sponsor	0.24	0.43	0.00	1.00
Group Vote Defection	0.04	0.20	0.00	1.00

Appendix B: Estimation Details

Bayesian CREM estimation requires the analyst to specify prior probability distributions for the model parameters $\boldsymbol{\beta}$, σ_m^2 , and σ_v^2 in equation 1 in the main text. Following Browne & Draper (2000) and Browne, Goldstein & Rabash (2001), we adopt a diffuse multivariate normal prior for the p fixed effects, $\boldsymbol{\beta} \sim N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, where $\boldsymbol{\mu}_0 = 0$ and $\boldsymbol{\Sigma}_0 = 10^6 \mathbf{I}$. Similarly, we select scaled inverse χ^2 priors for the variance terms, $\sigma_m^2 \sim \text{SI}\chi^2(v_m, s_m^2)$ and $\sigma_v^2 \sim \text{SI}\chi^2(v_v, s_v^2)$, where $v_m = v_v = 2 \cdot 10^{-3}$ and $s_m^2 = s_v^2 = 1$. By choosing these prior distributions we indicate that we are uncertain about $\boldsymbol{\beta}$ and assume that the random intercepts are all close to zero, a priori.

Rabash & Browne (In Press) describe a Gibbs sampling algorithm for estimating a CREM with continuous responses. We take advantage of the latent variable interpretation of binary regression and a data augmentation (Tanner & Wong 1987) technique introduced by Albert & Chib (1993) to convert this continuous response estimation algorithm into one that can estimate the BRM in equation 1 in the main text. This algorithm treats the random intercepts $\zeta^{(m)}$ and $\zeta^{(v)}$ as latent variables and introduces a new vector of latent variables \mathbf{z} , such that

$$z_i = \mathbf{x}_i \boldsymbol{\beta} + \zeta_{m(i)}^{(m)} + \zeta_{v(i)}^{(v)} + \epsilon_i \quad (2)$$

where we assume each independent and identically distributed $\epsilon_i \sim N(0, 1)$ and

$$y_i = \begin{cases} 0 & \text{if } z_i \leq 0 \\ 1 & \text{if } z_i > 0. \end{cases} \quad (3)$$

This is the familiar latent variable specification of the probit BRM and implies that

$$\Pr(y_i = 1 | \boldsymbol{\beta}, \zeta_{m(i)}^{(m)}, \zeta_{v(i)}^{(v)}) = \Phi \left[\mathbf{x}_i \boldsymbol{\beta} + \zeta_{m(i)}^{(m)} + \zeta_{v(i)}^{(v)} \right]$$

or, in other words, the latent variable specification in equations 2 and 3 is equivalent to the binary response CREM described by equation 1 in the main text. This parameterization of the model in equation 1 suggests a Gibbs sampling algorithm¹ that incorporates the following steps in each iteration ($n_k^{(c)}$ is

¹Gibbs samplers iteratively sample from the posterior distributions of subsets of the model parameters conditional on current simulated values of the remaining parameters—a process that eventually converges to the model parameters' joint

the number of observations in the k th unit of classification c , n_m is the number of MEPs and n_v is the number of votes):

1. Simulate $\boldsymbol{\beta}$ from $f(\boldsymbol{\beta}|\mathbf{z}, \sigma_m^2, \sigma_v^2, \boldsymbol{\zeta}^{(m)}, \boldsymbol{\zeta}^{(v)}) \sim N_p(\hat{\boldsymbol{\beta}}, \hat{\mathbf{D}})$, where

- $\hat{\mathbf{D}} = [\mathbf{X}'\mathbf{X} + \boldsymbol{\Sigma}_0^{-1}]^{-1}$
- $\hat{\boldsymbol{\beta}} = \hat{\mathbf{D}} \left[\sum_{i=1}^N \mathbf{x}'_i d_i + \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0 \right]$
- $d_i = z_i - \zeta_{m(i)}^{(m)} - \zeta_{v(i)}^{(v)}$

2. Simulate each $\zeta_k^{(m)}$ from $f(\zeta_k^{(m)}|\mathbf{z}, \boldsymbol{\beta}, \sigma_m^2, \sigma_v^2, \boldsymbol{\zeta}^{(v)}) \sim N(\hat{\mu}_k^{(m)}, \hat{D}_k^{(m)})$, where

- $\hat{D}_k^{(m)} = \left[n_k^{(m)} + \frac{1}{\sigma_m^2} \right]^{-1}$
- $\hat{\mu}_k^{(m)} = \hat{D}_k^{(m)} \left[\sum_{i \text{ s.t. } m(i)=k} \left(z_i - \mathbf{x}_i \boldsymbol{\beta} - \zeta_{v(i)}^{(v)} \right) \right]$

3. Simulate each $\zeta_k^{(v)}$ from $f(\zeta_k^{(v)}|\mathbf{z}, \boldsymbol{\beta}, \sigma_m^2, \sigma_v^2, \boldsymbol{\zeta}^{(m)}) \sim N(\hat{\mu}_k^{(v)}, \hat{D}_k^{(v)})$, where

- $\hat{D}_k^{(v)} = \left[n_k^{(v)} + \frac{1}{\sigma_v^2} \right]^{-1}$
- $\hat{\mu}_k^{(v)} = \hat{D}_k^{(v)} \left[\sum_{i \text{ s.t. } v(i)=k} \left(z_i - \mathbf{x}_i \boldsymbol{\beta} - \zeta_{m(i)}^{(m)} \right) \right]$

4. Simulate σ_m^2 from $f\left(\frac{1}{\sigma_m^2}|\mathbf{z}, \boldsymbol{\beta}, \sigma_v^2, \boldsymbol{\zeta}^{(m)}, \boldsymbol{\zeta}^{(v)}\right) \sim \text{Gamma}\left[\frac{n_m+v_m}{2}, \frac{1}{2} \sum_{j=1}^{n_m} (\zeta_j^{(m)})^2 + v_m s_m^2\right]$

5. Simulate σ_v^2 from $f\left(\frac{1}{\sigma_v^2}|\mathbf{z}, \boldsymbol{\beta}, \sigma_m^2, \boldsymbol{\zeta}^{(m)}, \boldsymbol{\zeta}^{(v)}\right) \sim \text{Gamma}\left[\frac{n_v+v_v}{2}, \frac{1}{2} \sum_{j=1}^{n_v} (\zeta_j^{(v)})^2 + v_v s_v^2\right]$

6. Simulate each z_i from the truncated normal distributions

$$f(z_i|y_i, \boldsymbol{\beta}, \sigma_m^2, \sigma_v^2, \zeta_{m(i)}^{(m)}, \zeta_{v(i)}^{(v)}) \sim \begin{cases} N_{(0,\infty)}(\mathbf{x}_i \boldsymbol{\beta} + \zeta_{m(i)}^{(m)} + \zeta_{v(i)}^{(v)}) & \text{if } y_i = 1 \\ N_{(-\infty,0)}(\mathbf{x}_i \boldsymbol{\beta} + \zeta_{m(i)}^{(m)} + \zeta_{v(i)}^{(v)}) & \text{if } y_i = 0 \end{cases}$$

This algorithm directly simulates from the posterior distributions of not only the fixed coefficients and random effects variances but also from the posteriors of all three sets of latent variables, allowing the analyst to work with the estimated posterior distributions of the random intercepts and to perform residual analysis based on \mathbf{z} using familiar techniques from linear regression modeling.

In our analysis of MEP voting behavior, we estimated each model by running the Gibbs sampler for 60,000 iterations, discarding the first 10,000 iterations and retaining every 50th iteration for a final posterior distribution. For an introduction to Gibbs samplers, and MCMC in general, see Gelman, Carlin, Stern & Rubin (2004) or Gill (2002).

posterior sample of 1,000 observations. Standard diagnostic tests generated results consistent with chain convergence for all three models and results are robust to variation in chain starting values and prior specification. In addition, penalized quasi-likelihood estimates of logistic versions of these models produce substantively similar results to the MCMC probit approach. We performed all MCMC computation in C++ using the Scythe Statistical Library (Pemstein, Quinn & Martin 2007).

References

- Albert, James H. & Siddhartha Chib. 1993. “Bayesian Analysis of Binary and Polychotomous Response Data.” *Journal of the American Statistical Association* 88(422):669–679.
- Browne, William J. & David Draper. 2000. “Implementation and Performance Issues in the Bayesian and Likelihood Fitting of Multilevel Models.” *Computational Statistics* 15:391–420.
- Browne, William J., Harvey Goldstein & Jon Rabash. 2001. “Multiple Membership Multiple Classification (MMMC) Models.” *Statistical Modelling* 1:103–124.
- Gelman, Andrew, John B. Carlin, Hal S. Stern & Donald B. Rubin. 2004. *Bayesian Data Analysis*. 2nd ed. Boca Raton: Chapman & Hall.
- Gill, Jeff. 2002. *Bayesian Methods: A Social and Behavioral Sciences Approach*. Boca Raton: Chapman & Hall.
- Pemstein, Daniel, Kevin M. Quinn & Andrew D. Martin. 2007. “Scythe Statistical Library v1.0.1.” <http://scythe.wustl.edu>.
- Rabash, Jon & William J. Browne. In Press. Non-Hierarchical Multilevel Models. In *Handbook of Quantitative Multilevel Analysis*, ed. Jan De Leeuw & I.G.G Kreft.
- Tanner, T.A. & W.H. Wong. 1987. “The Calculation of Posterior Distributions by Data Augmentation.” *Journal of the American Statistical Association* 82:528–549.