Cambridge Medium (2023), 1, 1–8

CAMBRIDGE
UNIVERSITY PRESS

ARTICLE TYPE

Supplementary materials for "A Novel Class of Unfolding Models for Binary Preference Data"

Rayleigh Lei* and Abel Rodriguez

Department of Statistics, University of Washington, Seattle, 98195, Washington, United States of America
*Corresponding author. Email: rlei13@uw.edu

1. Details of the Markov chain Monte Carlo algorithm for the static probit unfolding model

As discussed in Section 2.3 of the original manuscript, we consider a data augmentation approach where

$$\begin{pmatrix} y^*_{i,j,1} \\ y^*_{i,j,2} \\ y^*_{i,j,3} \end{pmatrix} \Bigg| \alpha_{j,1}, \alpha_{j,2}, \delta_{j,1}, \delta_{j,2}, \beta_i \sim \mathrm{N}\left(\begin{pmatrix} -\alpha_{j,1}(\beta_i - \delta_{j,1}) \\ 0 \\ -\alpha_{j,2}(\beta_i - \delta_{j,2}) \end{pmatrix} \Bigg| \mathbf{0}_3, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right), \tag{1}$$

for all $i = 1, \ldots, I$ and $j = 1, \ldots, J$, and let $y_{i,j} = 1$ if and only if $y^*_{i,j,2} > \max\left\{y^*_{i,j,1}, y^*_{i,j,3}\right\}$ and $y_{i,j} = 0$ otherwise. For notational convenience, we will let $\widetilde{\boldsymbol{y}}^*_{i,j} = \{y^*_{i,j,1}, y^*_{i,j,3}\}$.

Additionally, because of the mixture structure associated with the prior on $\boldsymbol{\alpha}_j$, $\boldsymbol{\delta}_j$, we consider a second data augmentation scheme and introduce variables $z_1, \ldots, z_J$ such that $z_j = 1$ if and only if $\alpha_{j,1} > 0$ and $\alpha_{j,2} < 0$, and $z_j = -1$ otherwise, where $\Pr(z_i = 1) = 0.5$ a priori.

Sampling $\beta_i, i = 1, 2, \ldots, I$: Conditional on $\boldsymbol{y}^*_{i,j}$'s, $\boldsymbol{\alpha}_j$'s and $\boldsymbol{\delta}_j$'s, the posterior distribution of each β_i is Gaussian with mean μ_{β_i}, and variance $\sigma^2_{\beta_i}$ given by:

$$\sigma^2_{\beta_i} = \frac{1}{1 + \sum_{j=1}^J \boldsymbol{\alpha}_j' \boldsymbol{\alpha}_j} \qquad \mu_{\beta_i} = -\sigma^2_{\beta_i} \left(\sum_{j=1}^J \boldsymbol{\alpha}_j' (\widetilde{\boldsymbol{y}}^*_{i,j} - D_{\boldsymbol{\alpha}_j} \boldsymbol{\delta}_i) \right). \tag{2}$$

Sampling $z_j, j = 1, 2, \ldots, J$: Given $\boldsymbol{y}^*_{i,j}$s, β_is $\boldsymbol{\delta}_j$, and after integrating over $\boldsymbol{\alpha}_j$, the conditional for $z_j = 1$ is proportional to the following:

$$\begin{aligned} &p(z_j = 1 \mid \boldsymbol{y}^*_{i,j} \boldsymbol{\delta}_j, \{\beta_i\}) \\ &\quad \propto \phi\left(\boldsymbol{\delta}_j \mid \boldsymbol{\mu}, \kappa^2 \mathbb{I}_{2\times 2}\right) \int_0^\infty \int_{-\infty}^0 \prod_i p(\boldsymbol{y}^*_{i,j} \mid \boldsymbol{\alpha}_j, z_j = 1, \boldsymbol{\delta}_j, \beta_i) p(\boldsymbol{\alpha}_j \mid z_j = 1) d\alpha_{j,2} d\alpha_{j,1} \\ &\quad = \phi\left(\boldsymbol{\delta}_j \mid \boldsymbol{\mu}, \kappa^2 \mathbb{I}_{2\times 2}\right) \left\{1 - \Phi\left(-(\boldsymbol{\mu}_{\boldsymbol{\alpha}_j})_1 / \left(\Sigma_{\boldsymbol{\alpha}_j}\right)_{1,1}\right)\right\} \Phi\left(-(\boldsymbol{\mu}_{\boldsymbol{\alpha}_j})_2 / \left(\Sigma_{\boldsymbol{\alpha}_j}\right)_{2,2}\right), \end{aligned} \tag{3}$$

where $\phi(\cdot \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the density of a (multivariate) Gaussian distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$, Φ the cumulative distribution function of the (univariate) standard Gaussian distribution, $\boldsymbol{\mu}_{\boldsymbol{\alpha}_j}$

and Σ_{α_j} are defined as

$$\Sigma_{\alpha_j} = \left(\sum_i D_{\beta_i - \delta_j} D_{\beta_i - \delta_j} + \frac{1}{\omega^2} \mathbb{I}_{2,2} \right)^{-1} \qquad \mu_{\alpha_j} = -\Sigma_{\alpha_j} \left(\sum_i D_{\beta_i - \delta} \widetilde{\gamma}^*_{i,j} \right), \tag{4}$$

with $D_{(a,b)}$ indicating a diagonal matrix with diagonal entries a and b for $a, b \in \mathbb{R}$, and $(\mu_{\alpha_j})_k$ and $\left(\Sigma_{\alpha_j}\right)_{k,k'}$ represent the k-th and (k, k')-th entries of μ_{α_j} and Σ_{α_j}, respectively.

Similarly, the conditional for $z_j = -1$ is given by

$$\begin{aligned} & p(z_j = -1 \mid \boldsymbol{\gamma}^*_{i,j} \boldsymbol{\delta}_j, \{\beta_i\}) \\ & \quad \propto \phi\left(\boldsymbol{\delta}_j \mid -\mu, \kappa^2 \mathbb{I}_{2\times 2}\right) \int_{-\infty}^{0} \int_{0}^{\infty} \prod_i p(\boldsymbol{\gamma}^*_{i,j} \mid \boldsymbol{\alpha}_j, z_j = 1, \boldsymbol{\delta}_j, \beta_i) p(\boldsymbol{\alpha}_j \mid z_j = -1) d\alpha_{j,2} d\alpha_{j,1} \\ & \qquad = \phi\left(\boldsymbol{\delta}_j \mid -\mu, \kappa^2 \mathbb{I}_{2\times 2}\right) \Phi\left(-(\mu_{\alpha_j})_1 / \left(\Sigma_{\alpha_j}\right)_{1,1} \right) \left\{ 1 - \Phi\left(-(\mu_{\alpha_j})_2 / \left(\Sigma_{\alpha_j}\right)_{2,2} \right) \right\}. \end{aligned} \tag{5}$$

Sampling $\boldsymbol{\alpha}_j, j = 1, 2, \ldots, J$: The posterior distributions of $\boldsymbol{\alpha}_j$ conditioned on z_j, β_is, $\boldsymbol{\gamma}^*_{i,j}$s and $\boldsymbol{\delta}_j$ is a truncated multivariate normal distributions for $j = 1, 2, \ldots, J$. The posterior mean for $\alpha_{j,1}$ and $\alpha_{j,2}$, $\mu_{\boldsymbol{\alpha}_j}$, and posterior covariance matrix, Σ_{α_j}, are given in (4). Meanwhile, the truncation is determined by z_j. If $z_j = 1$, then the distribution is truncated to the region where $\alpha_{j,1} > 0$ and $\alpha_{j,2} < 0$, and it is truncated to the region $\alpha_{j,1} < 0$ and $\alpha_{j,2} > 0$ if $z_j = -1$.

Note that we sample z_j and $\boldsymbol{\alpha}_j$ in the same Gibbs step because we compute the posterior of $\boldsymbol{\alpha}_j$ when sampling for z_j. Of course, we perform this Gibbs step by sampling z_j conditioned on β_i's and $\boldsymbol{\delta}_j$ and then sampling $\boldsymbol{\alpha}_j$ conditioned on β_i's, z_j, and $\boldsymbol{\delta}_j$.

Sampling $\delta_j, j = 1, 2, \ldots, J$: Finally, the posterior distributions of $\boldsymbol{\delta}_j$ conditioned on z_j, β_i's, $\boldsymbol{\gamma}^*_{i,j}$'s, and $\boldsymbol{\alpha}_j$ is a multivariate normal distributions with posterior mean $\boldsymbol{\mu}_{\delta_j}$ and posterior covariance matrix Σ_{δ_j} given by:

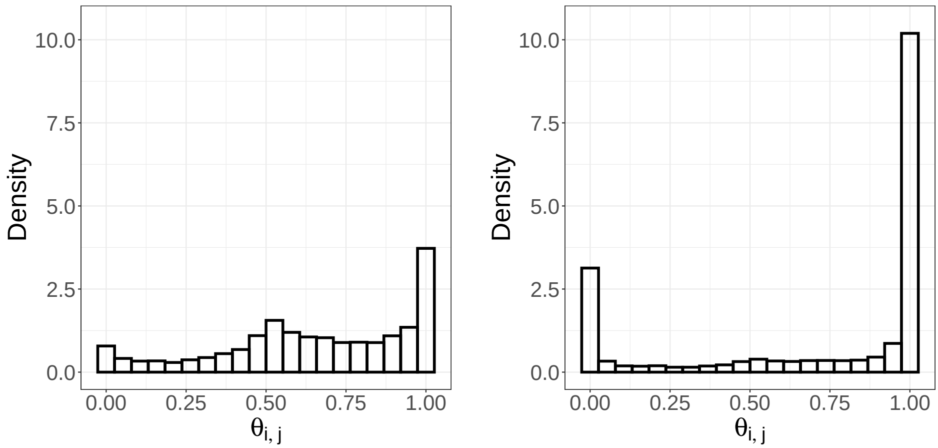
$$\Sigma_{\delta_j} = \left(I D_{\boldsymbol{\alpha}_j} D_{\boldsymbol{\alpha}_j} + \frac{1}{\kappa^2} \mathbb{I}_{2,2} \right)^{-1} \qquad \mu_{\delta_j} = \Sigma_{\delta_j} \left(\sum_{i=1}^{I} D_{\boldsymbol{\alpha}_j} (\widetilde{\boldsymbol{\gamma}}^*_{i,j} - \boldsymbol{\alpha}_j \beta_i) + z_j \frac{\boldsymbol{\mu}}{\kappa^2} \right). \tag{6}$$

Sampling $\boldsymbol{\gamma}^*_{i,j}; i = 1, 2, \ldots, I\ j = 1, 2, \ldots, J$: Conditioned on $\boldsymbol{\alpha}_j$, $\boldsymbol{\delta}_j$, and β_i, we first discuss how to sample $\boldsymbol{\gamma}^*_{i,j}$ if $y_{i,j} = 1$.

1. Draw $y^*_{i,j,1}$ from a normal distribution with mean $-\alpha_{j,1}(\beta_i - \delta_{j,1})$ and variance 1 truncated to the interval, $(-\infty, y^*_{i,j,2})$.
2. Draw $y^*_{i,j,2}$ from a standard normal distribution truncated to the interval, $(\max(y^*_{i,j,1}, y^*_{i,j,3}), \infty)$.
3. Draw $y^*_{i,j,3}$ from a normal distribution with mean $-\alpha_{j,2}(\beta_i - \delta_{j,2})$ and variance 1 truncated to the interval, $(-\infty, y^*_{i,j,2})$.

   Next, we discuss how to sample $\boldsymbol{\gamma}^*_{i,j}$ if $y_{i,j} = 0$.

1. If $y^*_{i,j,3} > y^*_{i,j,2}$, draw $y^*_{i,j,1}$ from a normal distribution with mean $-\alpha_{j,1}(\beta_i - \delta_{j,1})$ and variance 1. Otherwise, draw $y^*_{i,j,1}$ from the same normal distribution, but truncate this normal distribution to the interval $(y^*_{i,j,2}, \infty)$.



(a) Alternative prior

(b) Original prior

Figure 1. Histograms for 10,000 draws of the implied prior distribution on $\theta_{i,j}$ for our probit GGUM model under an alternative prior with $\boldsymbol{\mu} = (-2, 10)'$, $\omega^2 = 1$ and $\kappa^2 = 9$ (left panel) and under the original prior with $\boldsymbol{\mu} = (-2, 10)'$, $\omega^2 = 25$ and $\kappa^2 = 10$.

2. Draw $y^*_{i,j,2}$ from a standard normal distribution truncated to the interval, $(-\infty, \max(y^*_{i,j,1}, y^*_{i,j,3}))$.
3. If $y^*_{i,j,1} > y^*_{i,j,2}$, draw $y^*_{i,j,3}$ from a normal distribution with mean $-\alpha_{j,2}(\beta_i - \delta_{j,2})$ and variance 1. Otherwise, draw $y^*_{i,j,3}$ from the same normal distribution, but truncate this normal distribution to the interval $(y^*_{i,j,2}, \infty)$

2. Details of the Markov chain Monte Carlo algorithm for the dynamic probit unfolding model

Similarly to Section 1, we augment the sampler with variables

$$\begin{pmatrix} y^*_{i,j,t,1} \\ y^*_{i,j,t,2} \\ y^*_{i,j,t,3} \end{pmatrix} \Bigg| \alpha_{j,t,1}, \alpha_{j,t,2}, \delta_{j,t,1}, \delta_{j,t,2}, \beta_{i,t} \sim \mathrm{N}\left(\begin{pmatrix} -\alpha_{j,t,1}(\beta_{i,t} - \delta_{j,t,1}) \\ 0 \\ -\alpha_{j,t,2}(\beta_{i,t} - \delta_{j,t,2}) \end{pmatrix} \Bigg| \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right), \tag{7}$$

where $y_{i,j,t} = 1$ if and only if $y^*_{i,j,t,2} > \max\left\{y^*_{i,j,t,1}, y^*_{i,j,t,3}\right\}$ and $y_{i,j,t} = 0$ otherwise, and also define $z_{j,t} = 1$ if $\alpha_{j,t,1} > 0$ and $\alpha_{j,t,2} < 0$, and $z_{j,t} = -1$ otherwise, where $\Pr(z_{i,t} = 1) = 0.5$ a priori.

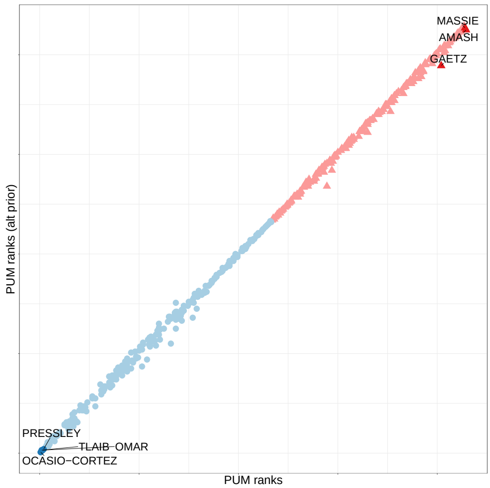
The steps associated with sampling the $\boldsymbol{\delta}_{j,t}$s, $\boldsymbol{\alpha}_{j,t}$s, $z_{j,t}$s and $\boldsymbol{y}^*_{i,t,j}$s are analogous to those described in the previous section. We describe the steps for sampling $\boldsymbol{\beta}_t$ and ρ below:

Sampling $\boldsymbol{\beta}_i$, $i = 1, 2, \ldots, I$: To ensure proper mixing of our Markov chain, we sample the whole trajectory of ideal points for the i-th justice, $\boldsymbol{\beta}_i$, from its joint full conditional distribution. This corresponds to a multivariate normal distribution with variance matrix $\Sigma_{\beta_i} = \left\{B + \boldsymbol{\Omega}(\rho)^{-1}\right\}^{-1}$ and mean $\mu_{\beta_i} = -\Sigma_{\beta_i}\mathbf{m}$, where $\boldsymbol{m}$ is a vector with entries

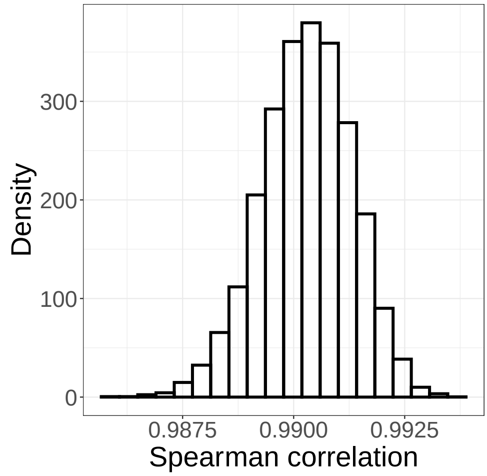
$$m_t = -\sum_{j: y_{i,j,t} \in \{0,1\}} \left(\alpha_{j,t,1}(y^*_{i,j,1} - \alpha_{j,t,1}\delta_{j,1}) + \alpha_{j,t,2}(y^*_{i,j,3} - \alpha_{j,t,2}\delta_{j,t,2}) \right) \tag{8}$$

and $\mathbf{B}$ is a diagonal matrix with entries

$$B_{t,t} = \sum_{j: y_{i,j,t} \in \{0,1\}} \boldsymbol{\alpha}'_{j,t}\boldsymbol{\alpha}_{j,t}. \tag{9}$$



(a) Posterior median ranks



(b) Spearman rank correlation

Figure 2. Comparison of the posterior distribution over legislator's ranks for the 116th House under two alternative priors. The left panel shows a comparison of the median ranks for each legislator. Democrats are shown with blue triangles, Republicans are shown with red triangles, and independents are shown with a rhombus and the color of the party that they caucus with. The right panel shows the posterior distribution of the Spearman correlation between both sets of ranks.

Sampling ρ: To sample ρ, we use a random walk Metropolis-Hasting step on the logit scale. More specically Let ρ' denote the proposed value for ρ which is obtained as

$$\rho' = \frac{1}{1 + \exp\left\{-\log \frac{\rho}{1-\rho} + \nu\right\}},$$

where ν follows a normal distribution with mean 0 and variance τ^2. Then ρ' is accepted with probability:

$$\min\left\{1, \frac{\left[\prod_{i=1}^{I} \phi(\boldsymbol{\beta}_i \mid \mathbf{0}, \boldsymbol{\Omega}(\rho'))\right] \phi_{(0,1)}(\rho' \mid \eta, \lambda^2)\rho'(1-\rho')}{\left[\prod_{i=1}^{I} \phi(\boldsymbol{\beta}_i \mid \mathbf{0}, \boldsymbol{\Omega}(\rho))\right] \phi_{(0,1)}(\rho \mid \eta, \lambda^2)\rho(1-\rho)}\right\}.$$

where $\phi(\cdot \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the density of a (multivariate) normal distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$, and $\phi_{(0,1)}(\cdot \mid \eta, \lambda^2)$ denotes the density of a (univariate) normal distribution with mean η and variance λ^2, truncated to the interval $[0, 1]$.

3. Evolution of inter-party spread in revealed preferences

Figure 3 presents the ratio of the spread of Democrats' revealed preferences to the spread of Republican's preferences based on three alternative metrics of spread: range, standard deviation and interquartile range (IQR). The first plot just repeats Figure 5 of the main manuscript. We can see that the overall picture is very similar no matter what metric is used, but that there some minor differences. For example, the differences between IDEAL and the unfolding models between the 110th and 114th House are less pronounced under the interquartile range and the standard deviation.

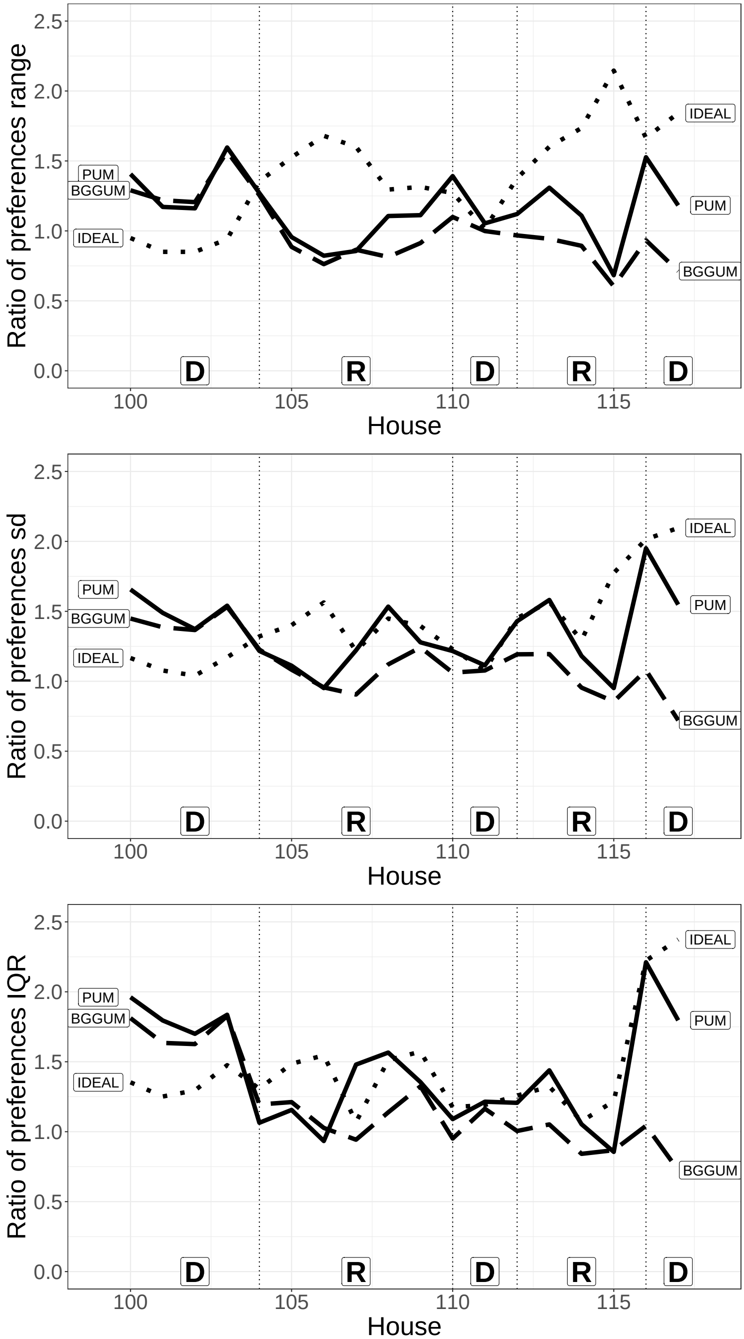
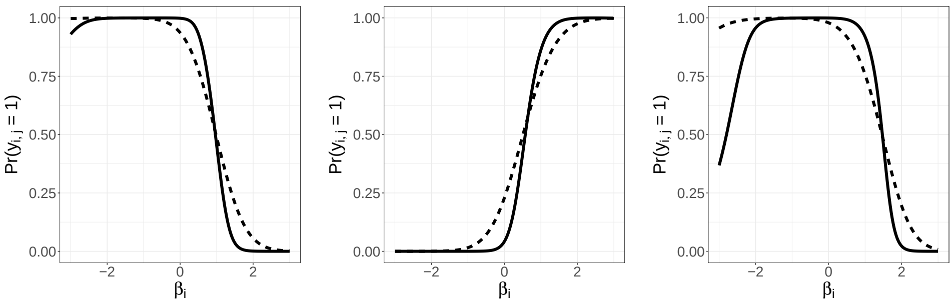


Figure 3. Posterior mean of the ratio of Democrats' range, standard deviation and interquartile range (IQR) over the same metric for Republicans across the various Houses. The solid line corresponds to the probit unfolding model, the dotted line to IDEAL, and the dashed line to BGGUM.

4. Sensitivity analysis for the static probit unfolding model

In this section, we conduct a sensitivity analysis for the results associated with the 116$^{\text{th}}$ U.S. House of Representatives (see Section 3 of the main manuscript). To accomplish this, we refit the model



(a) HRES5

(b) HR21

(c) HRES6

Figure 4. Plots displaying various posterior mean response curves based on the posterior means of $\alpha_{j,1}$, $\alpha_{j,2}$, $\delta_{j,1}$, and $\delta_{j,2}$ from the probit unfolding model for the 116th House under the two priors. The response curves for the prior discussed in the paper is shown with a solid line whereas the response curves for the alternative prior is shown with dashed lines.

setting $\omega^2 = 1$ and $\kappa^2 = 9$. The implied prior on $\theta_{i,j} = \Pr\left(y_{i,j} = 1 \mid \beta_i, \boldsymbol{\alpha}_j, \boldsymbol{\delta}_j\right)$ can be seen on the left panel of Figure 1. To facilitate comparison, the right panel shows again the prior on $\theta_{i,j}$ implied by the original prior (in which $\omega^2 = 25$ and $\kappa^2 = 10$, recall Figure 2a in the main manuscript). Note that the alternative prior favors values of $\theta_{i,j}$ close to 0.5 much more strongly than the original prior.

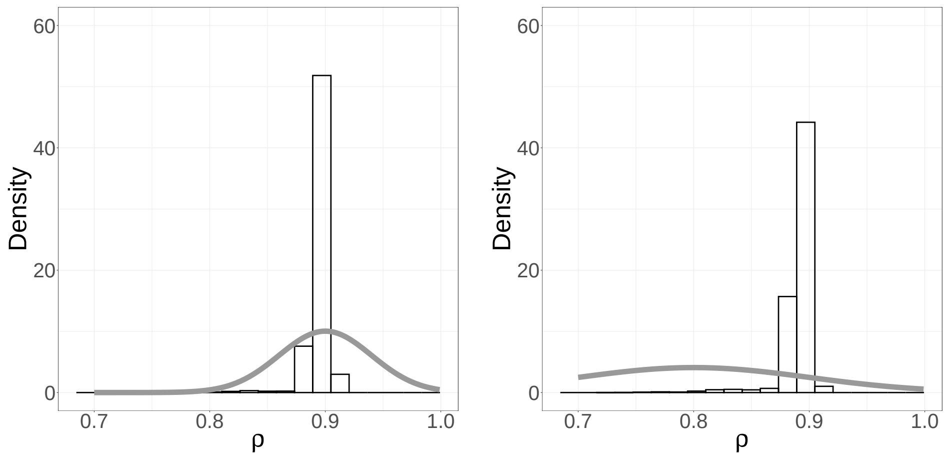
The left panel of Figure 2 compares the posterior median ranks of legislators under both priors, while its right panel shows the posterior distribution of their Spearman correlation. These graphs suggest that the ranks are fairly robust to the change in prior parameters. On the other hand, Figure 4 compares the posterior mean response curves associated with the same three votes discussed in Figure 7 of the main manuscript. Here we see a more pronounced effect of the prior, specially for HRES6. These differences, which mainly seem to be associated with the slopes of the response function, are not too surprising. Indeed, the alternative prior has a much lower variance for $\alpha_{j,1}$ and $\alpha_{j,2}$ than the original, which discourages large values for the slopes.

Finally, we also explored the impact of the alternative prior distribution on the WAIC score of our model (recall Equation (5) in the main manuscript). Interestingly, the WAIC under the alternative prior is worse than the WAIC under the original one, and also worse than the WAIC for BGGUM. However, the WAIC score for our model under the alternative prior is still better than the WAIC for IDEAL. These results suggest that our observation that unfolding models generally seem to better explain voting patterns in the modern U.S. House of Representatives is fairly robust to prior choices. However, it also reminds us that priors can have a substantial effect on model comparison criteria, and that matching priors (as we did in Section 2.1 of the main manuscript) is an important prerequisite to any meaningful application of model selection criteria.

5. Sensitivity analysis for dynamic probit unfolding model

We now study the sentivity to prior choices in the context of the analysis of U.S. Supreme Court vote data discussed in Section 5 of the main manuscript. For this sensitivity analysis, we consider an alternative prior on the autocorrelation parameter ρ that corresponds to a normal distribution with mean 0.8 and standard deviation 0.1, truncated to the $[0, 1]$ interval.

Figure 5 displays histograms of the posterior draws of ρ against the plots of priors' probability density function in each of the two scenarios. As seen in this plot, the alternative prior is centered at a much lower value for ρ and is much more diffuse. Indeed, under the original prior, $P(\rho < 0.85)$ is roughly 0.1, while under the alternative prior, $P(\rho < 0.85)$ is roughly 0.7. However, despite the fact the new prior places less probability on large values of ρ, the posterior is still quite concentrated around 0.9. This suggests that the posterior distribution of ρ is fairly robust to the prior.

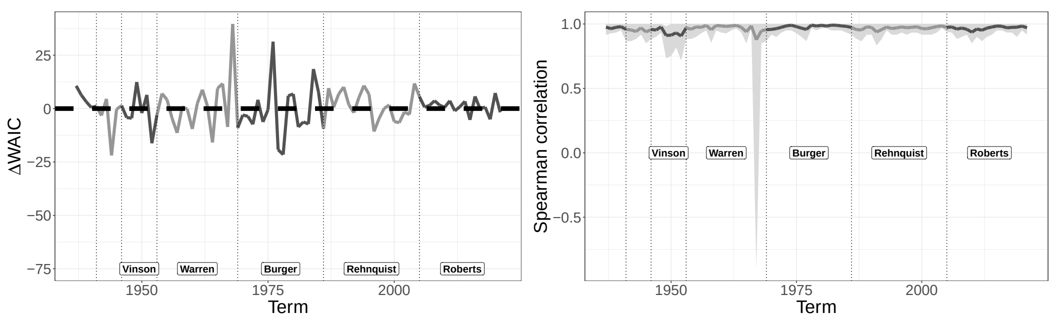


(a) Original prior

(b) Alternative prior

Figure 5. Histograms for the posterior of ρ under various priors with its respective prior probability density function plotted in grey.

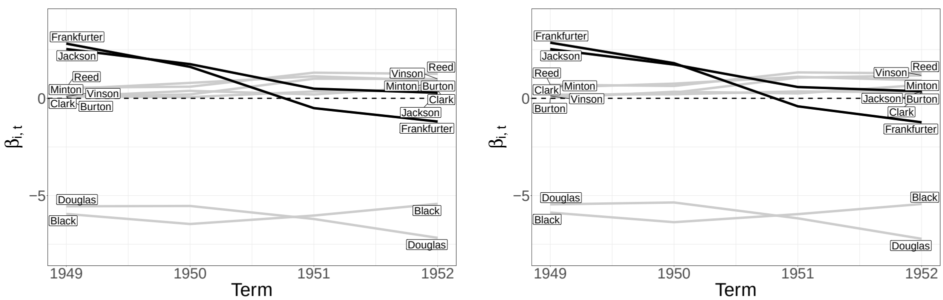
We now examine the downstream effect of this prior choice. Figure 6b shows the posterior mean (solid line) and corresponding 95% credible intervals (shaded region) for the Spearman correlation between the justices' rankings generated under the two priors we consider. Overall, we can see that the ranks are quite similar, although there seems to be some sensitivity during the 1967 term. This is likely related to Justice Black's ideal point for that term. As an example of the overall similarity, Figure 7 displays the posterior mean of the ideal points of SCOTUS justices active during 1949 to 1952 under both the original and the alternative prior. In both cases, the ideal points of Justices Frankfurter and Jackson drop in a similar manner during this time period. On the other hand, Figure 6a shows the difference in WAIC scores between the original and alternative prior on ρ. This graph does not show a clear pattern in favor of either prior.



(a) WAIC

(b) Spearman correlation

Figure 6. Left panel: Difference in WAIC scores for dynamic unfolding model under the original and alternative prior on ρ, ($WAIC$(DPUM alternative prior) - $WAIC$(DPUM original prior)). Note that the way the difference is being computed here is the opposite to the way in which it was computed in Figure 3a in the main text. Right panel: Posterior mean (solid line) and corresponding 95% credible intervals (shaded region) for the Spearman correlation between the justices' rankings generated under the two priors for ρ.



(a) Original prior

(b) Alternative prior

Figure 7. Posterior mean of the ideal points for SCOTUS Justices active during 1949 to 1952 terms under the dynamic unfolding model for various priors for ρ.