Supplementary Information for Generalized Ideal Point Models

Robert Kubinec

Table of contents

Additional Model Details

In this section I provide a more complete derivation of additional distributions available in the idealstan package.

Ordinal Derivations: Rating Scale and Graded Response

There are two possible ordinal models that can be implemented in the ideal point framework. The first is known as a rating-scale model in the IRT framework and is structurally similar to

the ordered logit model. Imai, Lo, and Olmsted (2016) were the first to implement this model for ideal points using an EM algorithm.

Following the notation I introduced, a rating-scale model for $k \in K$ outcomes can be modeled as follows for observed data:

$$L(Y_{ijtk}|\alpha_{it},\gamma_j,\beta_j) = \prod_{i-1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{T}\begin{cases} 1 - \zeta(\gamma_j\alpha_i - \beta_j - c_1) & \text{if } K = 0 \\ \zeta(\gamma_j\alpha_i - \beta_j - c_{k-1}) - \zeta(\gamma_j\alpha_i - \beta_j - c_k) & \text{if } 0 < k < K, \text{ and} \\ \zeta(\gamma_j\alpha_i - \beta_j - c_{k-1}) - 0 & \text{if } k = K \end{cases}$$

where again $\zeta(\cdot)$ represents the logit function. In this version, each ordinal category k less one is assigned a cutpoint c_k. In ideal point terms, these cutpoints divide the ideal point space across categories. A direct application is to Likert scales such as those commonly found in survey data. Importantly, because `idealstan` allows for mixed outcomes at the observation level, the number of ordinal categories K can likewise vary across items.

A different parameterization is known as the graded response model in IRT. This model allows the cutpoints c_k to vary by item. As a result, instead of including separate cutpoints, we can change the model by indexing each item-specific intercept for each category k:

$$L(Y_{ijtk}|\alpha_{it},\gamma_j,\beta_j) = \prod_{i-1}^{I}\prod_{j=1}^{J}\prod_{t=1}^{T}\begin{cases} 1 - \zeta(\gamma_j\alpha_i - \beta_{jk-1}) & \text{if } K = 0 \\ \zeta(\gamma_j\alpha_i - \beta_{jk-1}) - \zeta(\gamma_j\alpha_i - \beta_{jk-1}) & \text{if } 0 < k < K, \text{ and} \\ \zeta(\gamma_j\alpha_i - \beta_{jk-1}) - 0 & \text{if } k = K \end{cases}$$

Though not employed in political science to any great extent, this alternative formulation to ordered logit permits more flexible inference on the cutpoints. Allowing cutpoints to vary by item will enable much more accurate prediction of categories with fewer numbers of observations. It also increases information about the items/bills, though of course at a cost of including K additional parameters per item. My intention in including it here is not to suggest that it is a superior model to the standard rating-scale approach, but rather as an alternative that may be useful in some situations.

Poisson Distribution

The Poisson distribution has a considerable history as a latent variable model for word counts (Slapin and Proksch 2008), though it can be more broadly applied to any unbounded positive discrete variable. For `idealstan`, the Poisson parameterization is as follows:

$$P(Y_{ijtk} \mid \lambda_{ijtk}) = \frac{\lambda_{ijtk}^{Y_{ijtk}} e^{-\lambda_{ijtk}}}{Y_{ijtk}!} \tag{1}$$

$$\lambda_{ijtk} = e^{\gamma_j \alpha_i - \beta_j} \tag{2}$$

where λ_{ijtk} is the mean count and contains the IRT linear model with the exponential link function.

Normal and Log-Normal Distributions

The next two distributions have not been widely used in the ideal point literature but have been commonly deployed in continuous latent variable methods like principal components and factor analysis: the Normal distribution for continuous and the log-normal distribution for positive-continuous outcomes. While employed Kropko (2013) employed the Normal distribution in a traditional IRT framework, no one has yet used the log-Normal distribution.

For a normal distribution, the ideal point model is used to predict the mean of Y_{ijt} given item-specific residual standard deviation σ_j:

$$L(Y_{ijt}|\alpha_{it}, \gamma_j, \beta_j, \sigma_j) = \prod_{i-1}^{I} \prod_{j=1}^{J} \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{(Y_{ijt} - (\gamma_j \alpha_{it} - \beta_j))^2}{2\sigma_j^2}}$$

The log-normal distribution is the same distribution except that the domain of the distribution is on $log(Y_{ijt})$:

$$L(Y_{ijt}|\alpha_{it}, \gamma_j, \beta_j, \sigma_j) = \prod_{i-1}^{I} \prod_{j=1}^{J} \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{(log(Y_{ijt}) - (\gamma_j \alpha_{it} - \beta_j))^2}{2\sigma_j^2}}$$

which naturally imposes the condition that $Y_{ijt} > 0$.

The Log-Normal parameterization is useful for continuous variables that are only defined on positive real numbers. For variables with substantial skew, such as income, this parameterization has a more useful interpretation than the more general Normal distribution. With both of these distributions, most continuous distributions can be modeled with reasonable accuracy.

The residual standard deviation σ_j is given a weakly informative exponential prior:

$$\sigma_j \sim E(1)$$

Practically speaking, unbounded continuous items should be standardized (mean 0 with an SD of 1) so that the range of the outcome remains reasonably close to zero. Because MCMC algorithms are geometric, variables with wide variation can diminish sampling efficiency. Furthermore, if other items are included with different scales, the nature of a shared linear model across items can increase model misfit substantially.

Ordered Beta Distribution

The ordered beta distribution is a robust model for dependent variables that are in the contained $[0, 1]$ interval (Kubinec 2022). Examples of these distributions in social science data include percentages, proportions, and sliders (such as those used in divide-the-pie experiments). The ordered beta distribution is implemented instead of the simpler beta distribution in order to allow for responses at the bounds; i.e., observations of 0 (or 0%) and 1 (or 100%). Such observations are quite common in social science data and the ordered beta distribution only needs two additional parameters (ordered cutpoints) to model the full outcome. Without observations at the bounds of the scale, the ordered beta distribution effectively collapses to the beta distribution on the open interval $(0, 1)$.

The `idealstan` parameterization is as follows:

$$f(Y_{ijt} \mid \Omega, \Delta, \mu_j, \phi_j) = \begin{cases} \Omega & \text{if } Y_{ijt} = 0 \\ (1 - \Omega)\Delta & \text{if } Y_{ijt} = 1 \\ (1 - \Omega)(1 - \Delta)\text{Beta}(\mu_j, \phi_j) & \text{if } Y_{ijt} \in (0, 1) \end{cases}$$

The quantities Ω and Γ are the probabilities of the lower endpoint 0 and the probability of a response in the open interval $(0, 1)$. These probabilities are parameterized by ordered cutpoints $k \in \{1, 2\} : k_1 > k_2$. In other words, in the first stage of the model, either a continuous $(0, 1)$ or a discrete (0 or 1) outcome is selected, and then if the outcome is continuous, then it is also modeled with the Beta distribution.

The Beta distribution is further defined as follows:

$$f(Y_{ijt}, \omega_j, \tau_j) = \frac{\Gamma(\omega_j + \tau_j)}{\Gamma(\omega_j)\Gamma(\tau_j)} y_i^{\omega_j - 1}(1 - Y_{ijt})^{\tau_j - 1}$$

This Beta distribution is further reparameterized to allow for a linear model to predict the expected value (μ) while ϕ is the dispersion of the distribution. To do so, we must substitute for the shape parameters ω and τ:

$$\mu_j = \frac{\omega_j}{\omega_j + \tau_j}$$

4

$$\phi = \omega + \tau$$

Finally, to model $E[Y_{ijt}]$, we use the same IRT linear model as above:

$$\mu_j = g(\gamma_j \alpha_{it} - \beta_j)$$

where $g(\cdot)$ is the inverse logit function.

For further definition of this model, see Kubinec (2022).
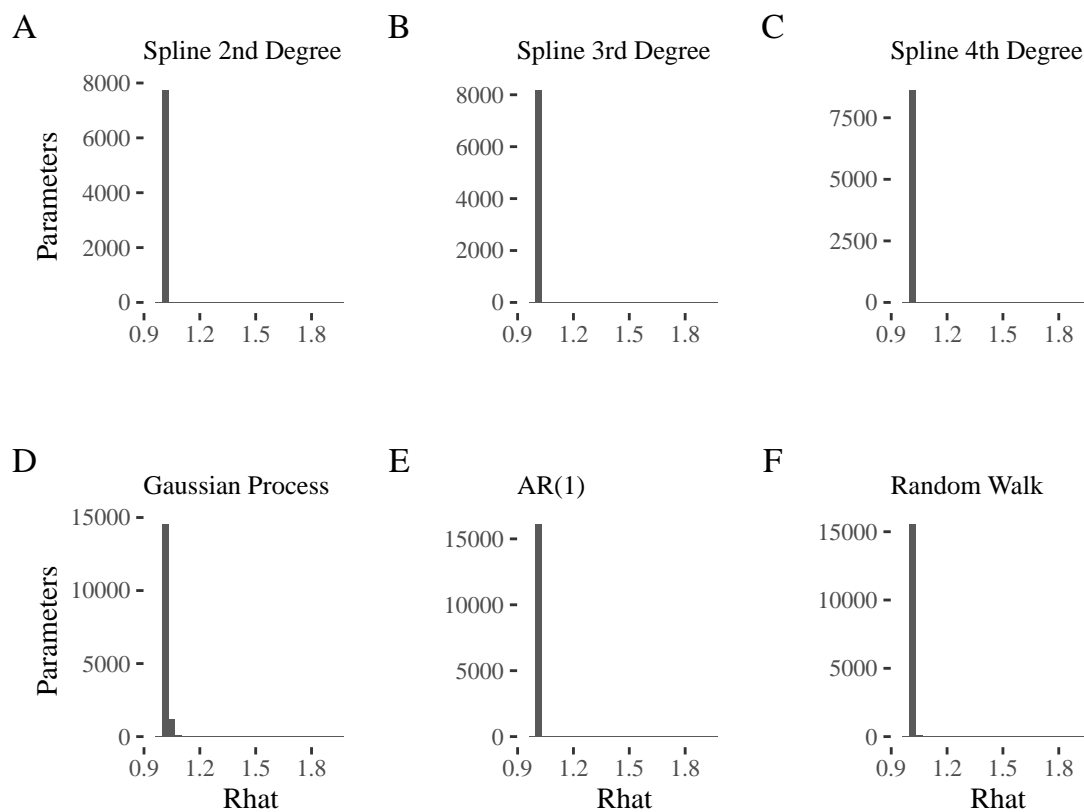
Model Convergence Diagnostics

In Figure Figure 1 I show the convergence diagnostics for all of the time-varying ideal point models of the 115th Congress. These convergence diagnostics employ the split-Rhat metric, which measures the ratio of variance of a Markov chain to itself over time and to other chains run independently. For this estimation, three chains were run with 500 post-warmup draws each, providing a difficult test for convergence. All of the models have almost all of their Rhat values below 1.1 with the exception of the Gaussian process model, which has a few parameters larger than 1.1. This more limited convergence indicates that the Gaussian process' much more demanding time-series function requires more draws than the other models to ensure a fully-explored posterior. However, even with 500 draws per chain, there is clear evidence of convergence for most parameters and inferences should be stable.

Parallelization Speed-up Simulation

In this section I report on simulations that examined the speed improvements with the use of multiple cores for within-chain parallelization of HMC gradient calculations. Figure 2 shows how estimation decreases for a given number of cores used for within-chain parallelization. The results for each core are the average across 2,000 Monte Carlo simulations of a binary/Bernoulli `idealstan` model with no missing data adjustment. For each draw, a number of persons between 10 and 200 and a number of items between 10 and 200 was selected so that the simulation includes substantial heterogeneity. Each time estimate in Figure 2 is an average across this heterogeneity and as such should be a reasonably robust estimate.

As can be seen in Figure 2, the most dramatic speedups are for moving from a single core to 2 to 4 cores. Above that number of cores, estimation time continues to decrease but a decreasing rate (i.e. the second derivative is less than 0). This is a known behavior of parallel processing, especially when there is computational overhead from running multiple processes. For Markov Chain Monte Carlo, it is not possible to run embarrassingly parallel processes in which there is a low fixed cost to using multiple cores. Rather, the process is partially parallelized via

A   Spline 2nd Degree

B   Spline 3rd Degree

C   Spline 4th Degree

D   Gaussian Process

E   AR(1)

F   Random Walk

Plots show split–Rhat values from three independent
Hamiltonian Markov Chain Monte Carlo chains with 500 post–warmup iterations per chain.

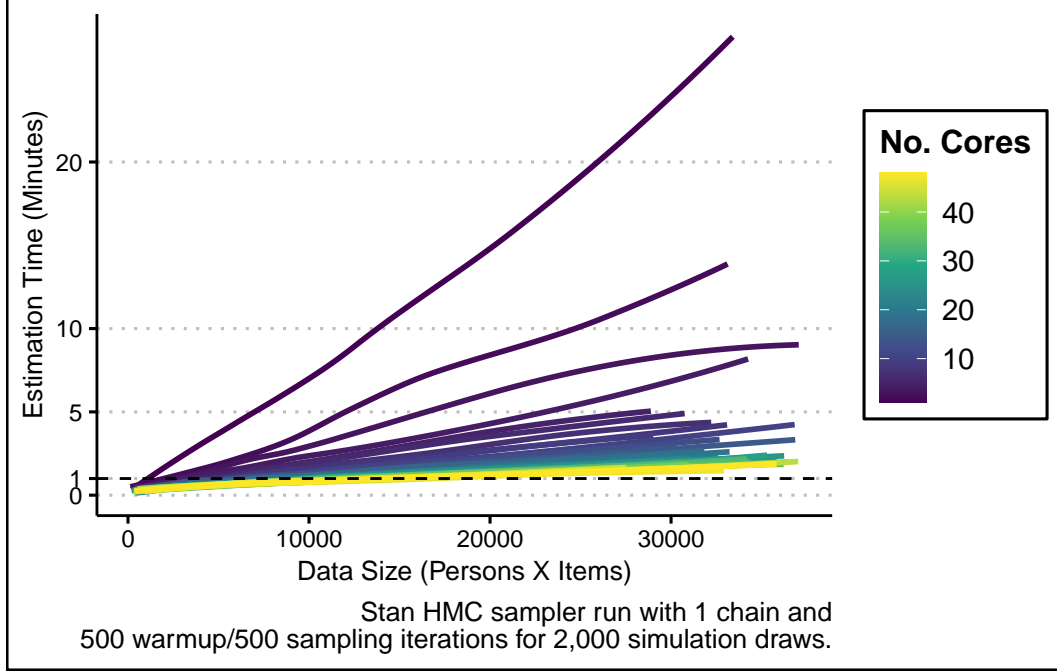Figure 1: Comparison of Convergence Diagnostics Across Specifications

Figure 2: Estimation Time for Idealstan Models by No. of Cores

the computations required to calculate the gradients necessary for Hamiltonian Markov Chain Monte Carlo, i.e., where the algorithm should "jump" next.

The `idealstan` package attempts to minimize this overhead by using Stan's map-reduce functionality to estimate the likelihood over subsets of data defined either by persons or items. For time-varying models, map-reduce must be performed over persons because the value of the person ideal point in time t depends on the prior value $t-1$, and as a consequence, attempting to estimate the likelihood for each item separately would yield biased results. However, using map-reduce over time-varying trajectories is often optimal in any case as the time-varying parameters for each person are calculated separately in each process, permitting less computational overhead.

For static models, the user can choose to parallelize over either items or persons. There is no clear guideline for which may be faster, and in general the most important criterion may be the number of items or persons relative to cores. For example, if there are five persons and 100 items in the data but only five cores, it would make the most sense to use persons as that would yield an efficient parallelization scheme.

Absences in the U.S. Congress

As described in the main text, Figure 3 and Figure 4 show rates of absences over time and within each session in the U.S. House.
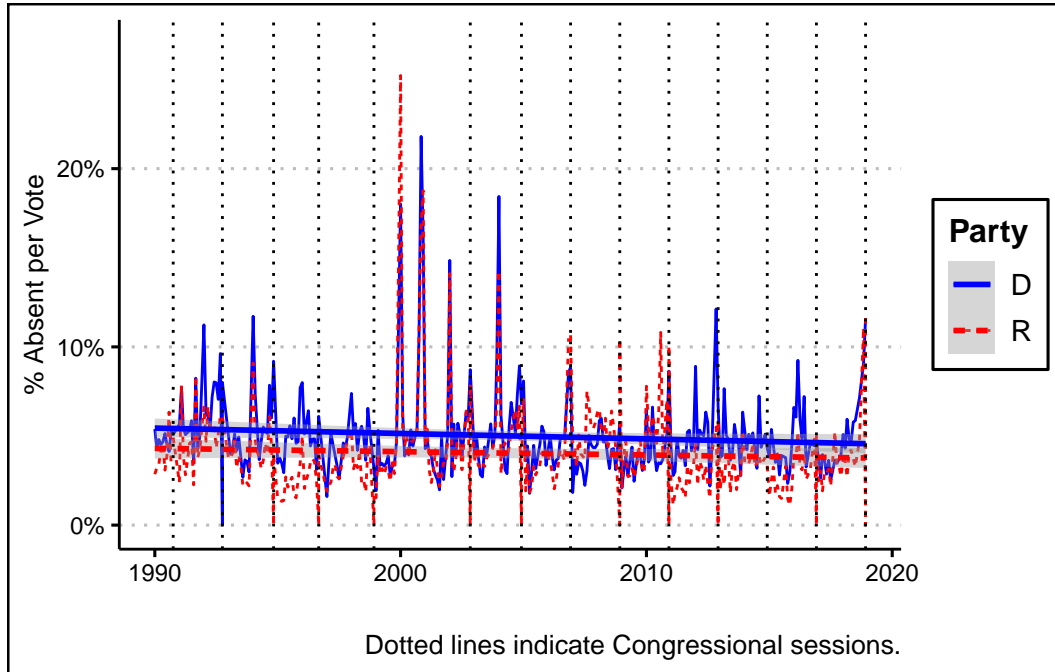

Figure 3: Time-varying Absence Rates in the U.S. House

Coefficients for Hierarchical Covariates

The coefficient values shown in Figure 12 in the main text are shown here in (**tab-coef?**).

Table 1: Table of Coefficients of GOP and District Unemployment Interaction by Model Type (ϕ)

Model	Variable	Missingness	Posterior Median	5% Quantile	95% Quantile
AR(1)	Unemployment	Unmodeled	0.296	0.1607	0.4285
AR(1)	Unemployment	Modeled	0.216	0.1178	0.3116
Gaussian Process	Unemployment	Unmodeled	0.245	0.1788	0.313
Gaussian Process	Unemployment	Modeled	0.185	0.1294	0.246
Random Walk	Unemployment	Unmodeled	0.175	0.0703	0.2872
Random Walk	Unemployment	Modeled	0.159	0.0666	0.2556
Spline 2nd Degree	Unemployment	Unmodeled	0.613	0.4857	0.7346
Spline 2nd Degree	Unemployment	Modeled	0.588	0.4711	0.6985
Spline 3rd Degree	Unemployment	Unmodeled	0.7	0.5754	0.8292
Spline 3rd Degree	Unemployment	Modeled	0.714	0.5998	0.834
Spline 4th Degree	Unemployment	Unmodeled	0.754	0.6187	0.8775
Spline 4th Degree	Unemployment	Modeled	0.736	0.6224	0.8445
AR(1)	UnemploymentXGOP	Unmodeled	-0.293	-0.471	-0.125
AR(1)	UnemploymentXGOP	Modeled	-0.206	-0.344	-0.0739
Gaussian Process	UnemploymentXGOP	Unmodeled	-0.223	-0.3257	-0.1303
Gaussian Process	UnemploymentXGOP	Modeled	-0.146	-0.2265	-0.0643
Random Walk	UnemploymentXGOP	Unmodeled	-0.188	-0.3416	-0.0257
Random Walk	UnemploymentXGOP	Modeled	-0.162	-0.2981	-0.0295
Spline 2nd Degree	UnemploymentXGOP	Unmodeled	-0.608	-0.7863	-0.4225
Spline 2nd Degree	UnemploymentXGOP	Modeled	-0.707	-0.8713	-0.5445
Spline 3rd Degree	UnemploymentXGOP	Unmodeled	-0.684	-0.892	-0.4925
Spline 3rd Degree	UnemploymentXGOP	Modeled	-0.967	-1.1435	-0.7955
Spline 4th Degree	UnemploymentXGOP	Unmodeled	-0.756	-0.9504	-0.5604
Spline 4th Degree	UnemploymentXGOP	Modeled	-0.963	-1.1375	-0.7907

Plot shows hierarchical legislator-levle covariates for different ideal point models, including time functions and whether or not the model accounted for missingness.
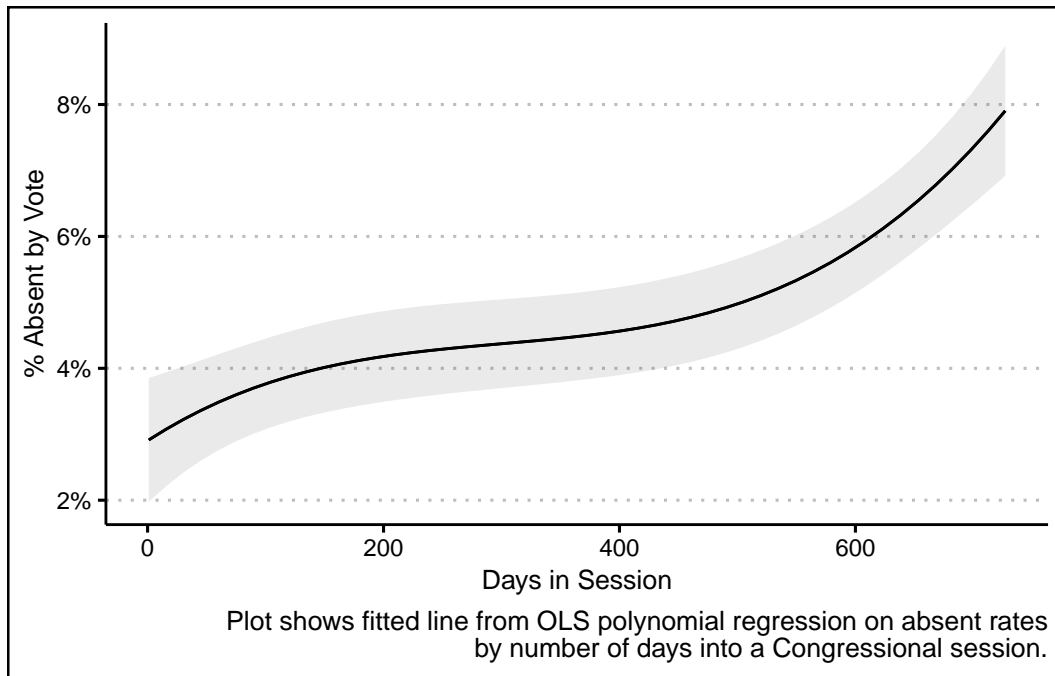
Figure 4: Ordinary Least Squares Polynomial Regression of Days in Session on Absence Rates in the U.S. House

References

Imai, Kosuke, James Lo, and Jonathan Olmsted. 2016. "Fast Estimation of Ideal Points with Massive Data." *American Political Science Review* 110 (4): 631–56. https://doi.org/10.1017/S000305541600037X.

Kropko, Jonathan. 2013. "Measuring the Dynamics of Political Power: A Time-Series IRT Model."

Kubinec, Robert. 2022. "Ordered Beta Regression: A Parsimonious, Well-Fitting Model for Continuous Data with Lower and Upper Bounds." *Political Analysis*, July, 1–18. https://doi.org/10.1017/pan.2022.20.

Slapin, Jonathan B., and Sven-Oliver Proksch. 2008. "A Scaling Model for Estimating Time-Series Party Positions from Texts." *American Journal of Political Science* 52 (3): 705–22.