

Introduction

This report outlines the methodology behind forecasting the outcome of the upcoming Icelandic Parliamentary Elections scheduled for November 30th. The forecast is based on a dynamic linear model implemented in Stan, incorporating polling data over time and adjusting for polling house effects while accounting for overdispersion.

Model Specification

We model the polling percentages for each political party over time using a dynamic linear model with a Dirichlet-Multinomial observation component. The model captures the evolution of party support and accounts for variations between different polling houses.

Notation

Input Data

- P : Number of political parties (*including the Other category*)
- T : Number of time points (dates) at which we have polling data
- H : Number of polling houses
- N : Number of observations (polls)
- $y_{n,p}$: Count of responses for party p in poll n
- Δ_t : The time difference between polls at $t - 1$ and t in days

Parameters

- $\beta_{p,t}$: Latent support for party p at time t (for $p = 2, \dots, P$)
- $\gamma_{p,h}$: Effect of polling house h for party p (for $p = 2, \dots, P$)
- $\mu_{\gamma,p}$: Mean house effect for party p
- $\sigma_{\gamma,p}$: Scale of house effects for party p
- σ_p : Scale parameter for the random walk of party p
- ϕ : Overdispersion parameter

Dynamic Party Effects

The latent support for each party (except the reference category) evolves over time following a random walk with scaled innovations:

$$\beta_{p,1} = \beta_p^{(0)}, \quad \beta_{p,t} = \beta_{p,t-1} + \sigma_p z_{p,t} \sqrt{\Delta_t} \quad \text{for } t = 2, \dots, T, \quad p = 1, \dots, P-1$$

where $z_{p,t} \sim \mathcal{N}(0, 1)$ and $\sqrt{\Delta_t}$ scales the innovations according to the time difference between polls.

Polling House Effects

Polling house effects are modeled hierarchically to account for systematic biases:

$$\gamma_{p,1} = 0, \quad \gamma_{p,h} = \mu_{\gamma,p} + \sigma_{\gamma,p} \tilde{\gamma}_{p,h} \quad \text{for } h = 2, \dots, H,$$

where $\tilde{\gamma}_{p,h} \sim \mathcal{N}(0, 1)$. The parameters $\mu_{\gamma,p}$ and $\sigma_{\gamma,p}$ control the mean and variability of polling house effects for each party.

Elections are set to be the first polling house and therefore $\gamma_{p,1} = 0$.

Overdispersion

To account for overdispersion in the polling data, we introduce an overdispersion parameter ϕ :

$$\phi = \frac{1}{\phi_{\text{inv}}},$$

where $\phi_{\text{inv}} \sim \text{Exponential}(1)$.

Data and Likelihood

The observed counts $\mathbf{y}_n = (y_{n,1}, \dots, y_{n,P})$ are modeled using a Dirichlet-Multinomial distribution:

$$\mathbf{y}_n \sim \text{Dirichlet-Multinomial} \left(\sum_{p=1}^P y_{n,p}, \phi \cdot \pi_n \right),$$

where $\pi_n = \text{softmax}(\eta_n)$ and η_n includes the latent support and polling house effect for each party, with the first party's linear predictor constrained to be the negative sum of other parties' predictors:

$$\eta_{n,p} = \begin{cases} -\sum_{p^*=2}^P (\beta_{p^*,t_n} + \gamma_{p^*,h_n}) & \text{if } p = 1 \\ \beta_{p,t_n} + \gamma_{p,h_n} & \text{if } p > 1 \end{cases}$$

Prior Distributions

The priors are specified as follows:

- **Initial Party Effects:** $\beta_p^{(0)} \sim \mathcal{N}(0, 1)$
- **Random Walk Innovations:** $z_{p,t} \sim \mathcal{N}(0, 1)$
- **House Effect Means:** $\mu_{\gamma,p} \sim \mathcal{N}(0, 1)$ with $\sum_p \mu_{\gamma,p} \sim \mathcal{N}(0, 1)$
- **House Effect Scales:** $\sigma_{\gamma,p} \sim \text{Exponential}(1)$
- **Random Walk Scales:** $\sigma_p \sim \text{Exponential}(1)$
- **Overdispersion Parameter Inverse:** $\phi_{\text{inv}} \sim \text{Exponential}(1)$

Inference

Bayesian inference is performed using Markov Chain Monte Carlo (MCMC) sampling via Stan. Posterior distributions of the latent variables $\beta_{p,t}$ and $\gamma_{p,h}$ are obtained, allowing for probabilistic forecasting of election outcomes. The overdispersion parameter ϕ helps in capturing extra variability in the polling data beyond the multinomial assumption.

Posterior Predictive Checks

To assess the model's fit, posterior predictive simulations are conducted:

$$\mathbf{y}_{\text{rep},d} \sim \text{Dirichlet-Multinomial}(n_{\text{pred}}, \phi \cdot \pi_d),$$

where $\pi_d = \text{softmax}(\eta_d)$ and η_d is constructed such that its first component is the negative sum of the remaining components, which are the latent party support values $\beta_{p,d}$.

Conclusion

The dynamic linear model effectively captures the temporal evolution of party support and adjusts for polling house biases while accounting for overdispersion in the data. By leveraging Bayesian methods, we obtain a comprehensive probabilistic forecast of the election outcomes, accounting for uncertainty in the estimates.