Dynamic Linear Election Model for Icelandic Parliamentary Elections Forecast

Brynjólfur Gauti Guðrúnar Jónsson Rafael Daniel Vias

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.

Model Specification

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

Notation

- P: Number of political parties.
- D: Number of time points (dates).
- H: Number of polling houses.
- N: Number of observations (polls).
- $y_{n,p}$: Count of responses for party p in poll n.
- $\beta_{p,t}$: Latent support for party p at time t.
- $\gamma_{p,h}$: Effect of polling house h for party p.
- σ_p: Scale parameter for the random walk of party p.

Dynamic Party Effects

The latent support for each party evolves over time following a random walk:

$$\beta_{p,1} = \mu_p, \quad \beta_{p,t} = \beta_{p,t-1} + \epsilon_{p,t} \quad \text{for } t = 2, \dots, D+1,$$

where $\epsilon_{p,t} \sim \mathcal{N}(0, \sigma_p^2 \times \Delta_t)$, and Δ_t is the time difference between polls at $t-1$ and t.

Polling House Effects

Polling house effects are modeled to account for biases:

$$\gamma_{p,1} = 0, \quad \sum_{h=1}^{H} \gamma_{p,h} \approx 0,$$

where election results are assigned to the the first polling house and therefore the first polling house's effect is set to zero. A soft sum-to-zero constraint is applied to the remaining effects to allow for small amounts of industry-level bias.

Data and Likelihood

The observed counts $y_n = (y_{n,1}, \dots, y_{n,P})$ are modeled using a multinomial distribution with a logit link:

$$y_n \sim \text{Multinomial}\left(\sum_{p=1}^{P} y_{n,p}, \text{softmax}\left(\eta_n\right)\right),$$

where $\eta_n = (\beta_{1,t_n} + \gamma_{1,h_n}, \dots, \beta_{P,t_n} + \gamma_{P,h_n})$, t_n is the date of poll n, and h_n is the polling house of poll n.

Prior Distributions

The priors are specified as follows:

- Initial party effects: $\beta_{0,p} \sim \mathcal{N}(0, 1)$.
- Random walk innovations: $\epsilon_{p,t} \sim \mathcal{N}(0, \sigma_p^2 \times \Delta_t^2)$.
- Polling house effects: $\gamma_{p,h} \sim \mathcal{N}(0, 1)$, with $\sum_h \gamma_{p,h} \sim \mathcal{N}(0, \sigma_{\text{house}}\sqrt{H-1})$ as a soft constraint.
- Scale parameters: $\sigma_p \sim \text{Exponential}(1)$.

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.

Posterior Predictive Checks

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

$$y_{\text{rep},d} \sim \text{Multinomial}\left(n_{\text{pred}}, \text{softmax}\left(\beta_{:,d}\right)\right), \quad d = 1, \dots, D+1.$$

These simulations generate replicated data under the model to compare with the observed data.

Conclusion

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