

## 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, adjusting for polling house effects, accounting for overdispersion, and modeling the impact of significant political events.

## 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, accounts for variations between different polling houses, incorporates overdispersion, and includes the effect of notable political events.

## Notation

### Input Data

- $P$ : Number of political parties.
- $T$ : 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  $n$ .
- $s_t$ : Indicator variable for the occurrence of the political event at time  $t$ .

### Parameters

- $\beta_{p,t}$ : Latent support for party  $p$  at time  $t$ .
- $\gamma_{p,h}$ : Effect of polling house  $h$  for party  $p$ .
- $\sigma_p$ : Scale parameter for the random walk of party  $p$ .
- $\phi$ : Overdispersion parameter.
- $\beta_p^{\text{event}}$ : Effect of a significant political event for party  $p$ .
- $\Delta_t$ : The time difference between polls at  $t - 1$  and  $t$  in days.

### Dynamic Party Effects

The latent support for each party evolves over time following a random walk with Student's t-distributed innovations:

$$\beta_{p,1} = \beta_{0,p}, \quad \beta_{p,t} = \beta_{p,t-1} + \epsilon_{p,t} \quad \text{for } t = 2, \dots, T+1,$$

where  $\epsilon_{p,t} \sim t_3(0, \sigma_p \sqrt{\Delta_t})$ , a Student's t-distribution with 3 degrees of freedom, scale  $\sigma_p \sqrt{\Delta_t}$ , and  $\Delta_t$  is the time difference between consecutive polls in days.

$\beta_{p,T+1}$  is the predicted support for party  $p$  at election day where  $\Delta_{T+1}$  is the number of days between the most recent poll and the day of the elections.

### Polling House Effects

Polling house effects are modeled to account for systematic biases:

$$\gamma_{p,1} = 0, \quad \gamma_{p,h} \sim \mathcal{N}(0, \sigma_{\text{house}}^2) \quad \text{for } h = 2, \dots, H,$$

where outcomes of elections are coded as the first polling house and thus the first polling house effect is set to zero. The hyperparameter  $\sigma_{\text{house}}$  controls the variability of polling house effects.

### Political Event Effect

We include an effect for a significant political event (e.g., government collapse):

$$\beta_p^{\text{event}} \sim \mathcal{N}(0, \sigma_{\text{event}}^2),$$

and incorporate it into the model as:

$$\eta_{n,p} = \beta_{p,t_n} + \gamma_{p,h_n} + \beta_p^{\text{event}} \times s_{t_n},$$

where  $s_{t_n}$  is an indicator (typically 0 or 1) denoting the occurrence of the event at time  $t_n$ .

## Overdispersion

To account for overdispersion in the polling data, we introduce an overdispersion parameter  $\phi$ :

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

where  $\phi_{\text{inv}} \sim \text{Exponential}(1)$ . This parameter adjusts the concentration of the Dirichlet-Multinomial distribution where a larger value of  $\phi_{\text{inv}}$  implies greater dispersion (more observational variance) and a smaller value of  $\phi_{\text{inv}}$  implies less overdispersion.

## 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  $\eta_n = (\eta_{n,1}, \dots, \eta_{n,P})$  includes the latent support, polling house effect, and event effect for each party:

$$\eta_{n,p} = \beta_{p,t_n} + \gamma_{p,h_n} + \beta_p^{\text{event}} \times s_{t_n}.$$

Here,  $t_n$  is the date of poll  $n$ , and  $h_n$  is the polling house of poll  $n$ .

## Predictions

We then predict the true latent party support as

$$\mathbf{y}_t^* \sim \text{Dirichlet-Multinomial} (n_{\text{pred}}, \phi \cdot \pi_t^*),$$

where  $\pi_t^* = \text{softmax}(\eta_t^*)$  and  $\eta_t^* = (\eta_{t,1}^*, \dots, \eta_{t,P}^*)$  includes the latent support and event effect for each party, but not the house effects:

$$\eta_{t,p}^* = \beta_{p,t} + \beta_p^{\text{event}} \times s_t.$$

## Prior Distributions

The priors are specified as follows:

- **Initial Party Effects:**  $\beta_{0,p} \sim \mathcal{N}(0, 1)$ .
- **Random Walk Innovations:**  $\epsilon_{p,t} \sim t_3(0, \sigma_p \sqrt{\Delta_t})$ .
- **Polling House Effects:**  $\gamma_{p,h} \sim \mathcal{N}(0, \sigma_{\text{house}}^2)$  for  $h = 2, \dots, H$ .
- **Scale Parameters:**  $\sigma_p \sim \text{Exponential}(1)$ .
- **Overdispersion Parameter Inverse:**  $\phi_{\text{inv}} \sim \text{Exponential}(1)$ .
- **Event Effects:**  $\beta_p^{\text{event}} \sim \mathcal{N}(0, \sigma_{\text{event}}^2)$ .

## Inference

Bayesian inference is performed using Markov Chain Monte Carlo (MCMC) sampling via Stan. Posterior distributions of the latent variables  $\beta_{p,t}$ ,  $\gamma_{p,h}$ , and  $\beta_p^{\text{event}}$  are obtained, allowing for probabilistic forecasting of election outcomes. The overdispersion parameter  $\phi$  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:

$$\mathbf{y}_{\text{rep},d} \sim \text{Dirichlet-Multinomial}(n_{\text{pred}}, \phi \cdot \pi_d), \quad d = 1, \dots, D+1,$$

where  $\pi_d = \text{softmax}(\beta_d)$  and  $\beta_d = (\beta_{1,d}, \dots, \beta_{P,d})$ .

These simulations generate replicated data under the model to compare with the observed data, aiding in the evaluation of model adequacy.

## Conclusion

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