INWT Statistics

**INWT** Statistics GmbH

# A Bayesian long-short term event memory state-space model for multi-party elections

Dr. Marcus Groß | Meetup-BerlinBayesians | July 1, 2021

# Motivation

**What are the potential benefits of a model based election forecast compared to polls / poll averages?**

- Improved forecasts
  - Polls ask for voting intention for potential election next sunday, *not* election itself
  - Polls can be biased / of varying quality $\rightarrow$ correction / adjustments beneficial
- More accurate uncertainty measures, more insights e.g. probabilistic statements (what is the probability that a coalition gains majority in parliament?)
- Predictions on government formation via additional data and model (what is the probability that X will be chancellor or that coalition A will be forming the government)

# Motivation

**Approaches for US presidential elections:**

- fivethirtyeight.com
- economist.com
- https://www.nytimes.com/
- ...

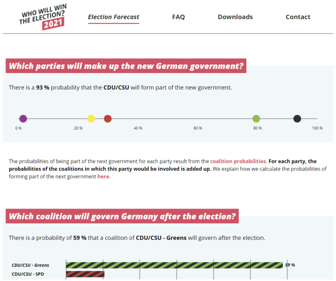
**In contrast: German federal elections (or comparable election systems):**

- More major parties (six currently)
- Popular vote only relevant for seat distribution (overhang mandates w/o influence since 2013)
- Electoral threshold for parliament (5% in Germany)
- Government formation and chancellor elected after election (w/o voter's participation)

# Website for German federal election forecast

`https://wer-gewinnt-die-wahl.de/en/`

- Updated daily, historical forecasts downloadable
- Open-source, full transparency and reproducibility via github repository
- No external funding



# Data

- Polling data since 1994 from 7 pollsters, 4000 polls in total
  - Data is scraped from webpage ( `https://www.wahlrecht.de/`) collecting all available poll data and frequently updated
  - Available for six large parties: CDU/CSU, SPD, Die Grünen, FDP, Die Linke, and AfD
- German election outcomes since 1994
- Government / opposition status of parties
- Expert survey data on coalition ranking

# Model

**Two models:**

1. ***Vote share model:*** Model election vote shares using poll, election and party status data
2. ***Government model:*** Model government formation using vote share model simulations and expert survey data

Both models use Bayesian approaches. Vote share model uses (R)Stan, Government model R only

# I: Vote share model – Overview

Let $y_{pt,t}$ the vote share or vote intentions of party $pt$ if an election was held at time point $t$ (main parameters of interest).

We have **two major objectives**:

1. Given the poll data $poll_{pt,t,pll}$ ($pll$ = pollster), what can we infer about the true vote share $y_{pt,t}$?
2. How can the time series of $y_{pt,t}$ be described? Can our forecast on $y_{pt,t+1}$, $y_{pt,t+2}$ ,.. be improved given the history $y_{pt,t}$, $y_{pt,t-1}$, $y_{pt,t-2}$,.. other than assuming $E(y_{pt,t+1})$, $E(y_{pt,t+2})$ ,.. = $E(y_{pt,t})$?

Related to that one has two major **sources of uncertainty**:

1. Uncertainty in polling
2. Uncertainty about future events, i.e. shocks to vote share

# I: Vote share model – Long-Short term memory

Modeling time series of vote shares $y_{pt,t}$, $t = 1, 2, ..$ (weekly intervals)

Usual approach: (Gaussian) Random walk (e.g. "Pooling the polls over an election campaign", Simon Jackman, 2005, Australian Journal of Political Science):

$$y_{pt,t} = y_{pt,t-1} + \epsilon; \text{ with } \epsilon \sim N(0, \sigma^2_{eps})$$

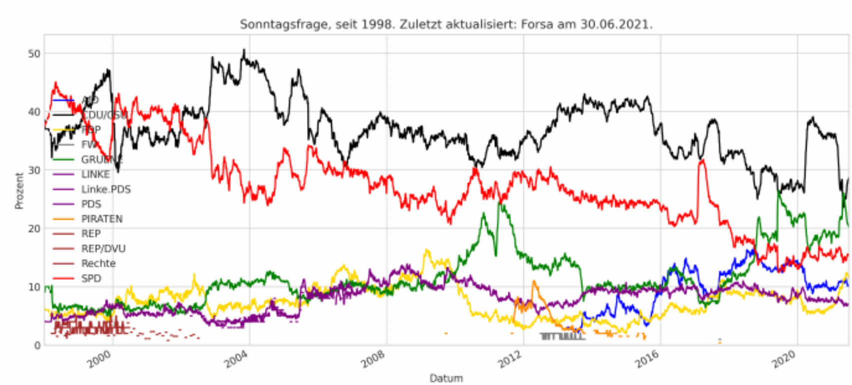
Random walks $\rightarrow$ flat expectation forecast

Do we have a classic random walk? Are the shocks $\epsilon$ normally distributed?

**Let's look at the data!**

# I: Vote share model – Long-Short term memory

Polls (as proxies for vote intentions) since 1998 (from `https://www.dkriesel.com/sonntagsfrage`):




Vote shares more stable than expected from random walks over long time ranges

Weekly shocks seem heavy tailed with rare large spikes

# I: Vote share model – Long-Short term memory

Other model approaches and poll averagers also observe some sort of **mean-reversion process**:

- *The basic principle is that polling has some systematic biases, in particular a tendency to overstate changes from the previous election*; http://www.electionforecast.co.uk/
- *Empirically, using more smoothing early in the race and less smoothing late in the race works best. In other words, the trend line starts out being quite conservative and becomes more aggressive as Election Day approaches*; FiveThirtyEight
- ...

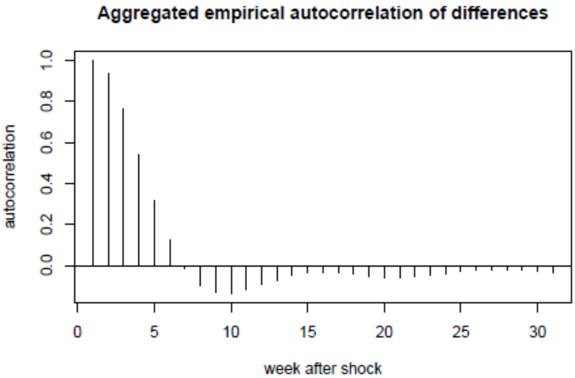
$\rightarrow$ Mixture between non-stationary random walk and a mean-reversing process for vote share time?

**Two alternatives:**

1. Structural model (including a model component using e.g. unemployment data or gdp growth data)
2. Long-short term memory model

# I: Vote share model – Long-Short term memory

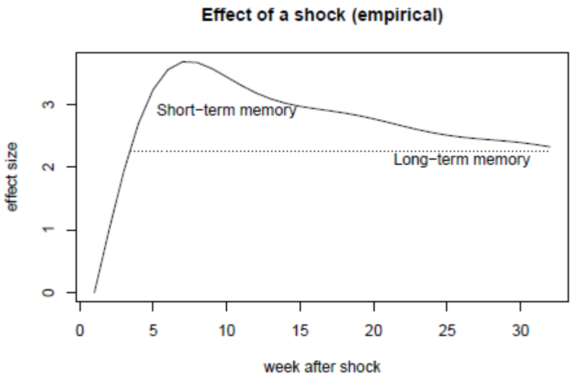
Time Series? Look at autocorrelations*:




*(weekly interpolated polls as proxy for vote intentions, acf's on differences aggregated over 5 major parties)

# I: Vote share model – Long-Short term memory

Cumulated autocorrelations:




Interpretation: News that influence vote share get spread through media, peaking after some weeks. Afterwards events slowly get forgotten and/or attention is focussed on new events → **Long-short event memory model**

# I: Vote share model – Long-Short term memory

How to mimic the LS-memory effect in the model? $\rightarrow$ Multiple AR(1) resembling models:

- weekly shocks $\epsilon$ are multivariate (correlations between shocks) t-distributed with 3.5 degrees of freedom because of heavy tails
- positive short-term memory parameter $\nu_{pt,t}$; $\theta_2$ marks the spreading speed of the weekly event effect

$$\nu_{pt,t} = \theta_2 \cdot (\nu_{pt,t-1} + \epsilon_{pt,t-1})$$

- diminishing short- and mid- term effect parameter $\eta 1_{pt,t}$ and $\eta 2_{pt,t}$; $\alpha$ and $\alpha_2$ govern the amount of "forgetting"; $\theta_0$ and $\theta_1$ mark the short- and mid-term memory decay speed, fixed to 0.7 and 0.95 (half life times of 2 weeks and 3 months):

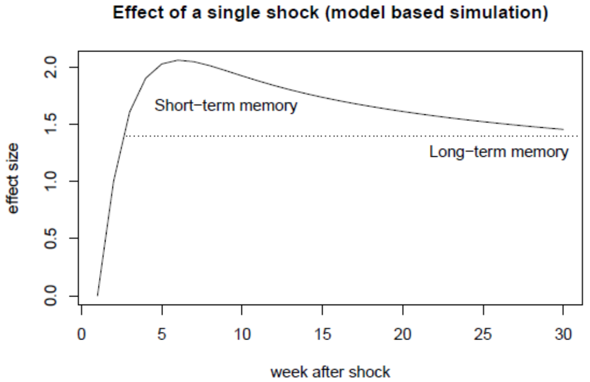
$$\eta 1_{pt,t} = \theta_0 \cdot \eta 1_{pt,t-1} + (1 - \theta_0) \cdot \alpha \cdot \nu_{pt,t-1}$$
$$\eta 2_{pt,t} = \theta_1 \cdot \eta 2_{pt,t-1} + (1 - \theta_1) \cdot \alpha_2 \cdot \nu_{pt,t-1}$$

- Thus, vote share: random walk with 3 additional LS-term components:

$$y_{pt,t} = y_{pt,t-1} + \epsilon_{pt,t} + \nu_{pt,t} - \eta_{1,pt,t} - \eta_{2,pt,t}$$

# I: Vote share model – Long-Short term memory

Result from model fit:



# I: Vote share model – Long-Short term memory

Advantages compared to structural model:

- less data needed
- simpler model, self-containing
- economic or other indicators often lagging
- more direct approach (voters know best themselves whom they are going to vote)
- comparable to market efficiency hypothesis in equity markets
- less danger of overfitting and variable selection bias

Overall: LS-memory model and structural model have similar forecasting tendencies (partial mean-regression)

# I: Vote share model – party status effect

Party status: Government or Opposition in German federal elections between 1994 and 2017:

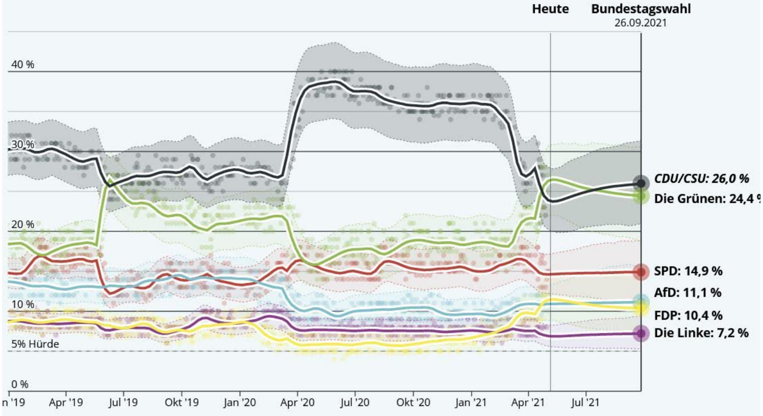
- Government parties lost vote share in 10 of 12 cases in the following election
- Opposition parties gained vote share in 15 of 20 cases
- Same effect in neighboring countries with similar multi-party election systems like Netherlands or Sweden

$\rightarrow$ include non-zero expectation parameters for weekly shocks, positive for opposition parties, negative for government

Increases long-term stability of party vote share, can be used to improve predictive power

# I: Vote share model – Long-Short term memory

Example of LS-memory effects, early May 2021: Green and CDU/CSU expect to cross again in the future after large shifts in polls / vote shares:



# I: Vote share model – poll errors

Going back to first major objective: *Given the poll data $poll_{pt,t,pll}$ (pll = pollster), what can we infer about the true vote share $y_{pt,t}$?*

$y_{pt,t}$ can only be observed when elections take place

Polling error, i.e. difference between $poll_{pt,t,pll}$ and $y_{pt,t}$, can be decomposed:

1. polling uncertainty due to limited sample size
2. house bias of pollster for specific parties (long term and short term)
3. common bias of all pollsters (changes over time, reset after election, correlated between parties)

$$logit(poll_{pt,t,pll}) \sim N(y_{pt,t} + bias_{pt,pll} + bias_{pt,pll,el} + \epsilon_{\text{pollError}_{pt,el}}, \sigma^2_{pt,pll} + 0.0002^2)$$

- bias/errors are normally distributed, common distribution over parties, pollsters, election cycles
- logit transformation for variance stabilization
- deep hierarchical structure $\rightarrow$ very hard to fit
- elections treated as poll with all biases / errors set to 0 (except 'counting error' of 0.0002 sd for fitting)

# I: Vote share model – poll errors

What is *common bias*? Pollsters do:

- Weighting (to achieve representativity)
- Adjustments and corrections (non-response, socially desirable response behavior,..) lead to bias shared among all pollsters.

Example: German 2005 election, CDU/CSU polling error hardly explainable w/o common bias:

| Institut | Datum | CDU/CSU | SPD | GRÜNE | FDP | LINKE | AfD | Befragte | Sonstige |
|---|---|---|---|---|---|---|---|---|---|
| Allensbach | 2005-09-16 | 0.415 | 0.325 | 0.070 | 0.080 | 0.085 | NA | 1682 | 0.025 |
| Allensbach | 2005-09-13 | 0.417 | 0.329 | 0.072 | 0.070 | 0.085 | NA | 2000 | 0.027 |
| Kantar(Emnid) | 2005-09-13 | 0.420 | 0.335 | 0.070 | 0.065 | 0.080 | NA | 2000 | 0.030 |
| Forsa | 2005-09-12 | 0.420 | 0.350 | 0.070 | 0.060 | 0.070 | NA | 2504 | 0.030 |
| GMS | 2005-09-12 | 0.420 | 0.330 | 0.080 | 0.070 | 0.070 | NA | 1008 | 0.030 |

| CDU.CSU | SPD | GRÜNE | FDP | LINKE | AfD | Sonstige | Year | Datum | Institut |
|---|---|---|---|---|---|---|---|---|---|
| 0.352 | 0.342 | 0.081 | 0.098 | 0.087 | NA | 4 | 2005 | 2005-09-18 | Election |

Assumption: Common bias is reset after election (pollsters adjust their weighting / correction methods), slowly builds up until next election

# I: Vote share model – model fitting

- fitting with RStan, version 2.2.1
- max_treedepth needs to be increased to 13 or better 14 due to multiple hierarchical level structure
- for daily fitting: 8 chains, 275 burnin-iterations, 400 iterations total $\rightarrow$ 1000 posterior simulations
- runtime: ca. 15 hours
- weak-medium informative priors and parameter bounds

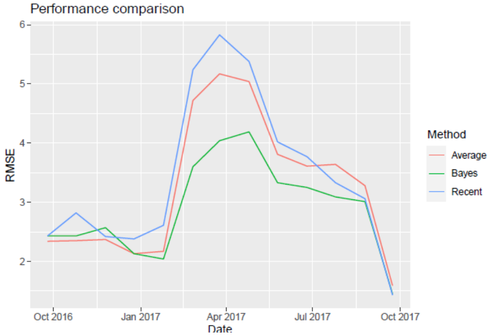
1000 simulations on vote share distribution $\rightarrow$ apply 5% electoral threshold to get parliament ("Bundestag") composition

Uncertainty intervals, probabilities of events straightforward to derive

# I: Vote share model – predictive performance

Backtesting for election 2017. Compute forecasts for election monthly starting from October 2016 to September 2017. Competitors:

- Most recent poll ('Recent')
- Average of most recent polls from each pollster ('Average'), only polls less than 2 weeks old.




- On average model has 20% lower RMSE than 'Recent' and 15% lower RMSE than 'Average'
- Similar result for 2013
- Advantage lower, the closer the election

# II: Government model – Motivation

Vote share model gives valuable insights by simulating election outcomes + seat distribution of parliament.

**However:**

- How will the government be constituted? Who will be chancellor? Which party will be part of government?
- Multiple coalitions conceivable for given election outcome
- Government coalition or chancellor decision is decided by parties/parliament not voter

# II: Government model – Expert survey

**Idea:** Incorporate expert knowledge on coalition forming and combine results with vote share model simulations

*Think of a situation, where representatives of all parties are locked into a room with the goal to form coalitions and name the parties which would get together first. If that coalition is ruled out, who would find together next and so on.*

- Expert: person active in politics / party, working for a political institution or having a degree in political science
- Coalition ranking for selected number of coalitions that are considered remotely conceivable (first 12, now 18 coalitions)
- Ranking should be done independently to the election result, i.e. 'orthogonal'
- Interviews were assisted to prevent errors and misunderstandings
- Last interviews: 21 experts asked
- Coalition preferences between parties don't change much over time, but regular interviews desirable

# II: Government model – Sketch

Example with parties A, B and C. Vote share model: 100 simulations of election – 4 scenarios:

1. In 10 simulations only coalitions with parties A+B or A+B+C had a majority
2. In 15 simulations A+B or A+C or A+B+C had a majority
3. In 60 simulations A or A+B or A+C or A+B+C had a majority
4. In 25 simulations A+C or A+B+C had a majority

What do **experts** say?

1. 20 / 20 experts ranked A+B above A+B+C, thus A+B gets 100% probability
2. 3 / 20 experts ranked A+B above A+B+C and A+C, 17/20 A+C above A+B and A+B+C, thus A+B gets 15% probability and A+C 85%
3. 20 / 20 experts ranked A above A+B, A+C and A+B+C , thus A gets 100% probability
4. 20 / 20 experts ranked A+C above A+B+C, thus A+C gets 100% probability

**Simple model:** Coalition A+B total probability of (10 * 100% + 15 * 15%) / 100 = 2.25%
Coalition A+C = (15 * 85% + 25 * 100%) / 100 = 37.75%
'Coalition' A (60 * 100%) / 100 = 60%
Coalition A+B+C 0%

# II: Government model – Model

Bayesian model extension: Dirichlet prior on multinomial expert model with all $\alpha_k$ set to 0.5 ('non-informative' Jeffrey's prior)

Further assumptions:

- Senior and junior partner: A+B and B+A are different coalitions – senior partner (higher vote share) will provide chancellor
- No minority governments (very unusual in Germany)
- If coalition with less parties is possible, coalition with more partners is ruled out, e.g. if A+B+C has majority and A+B too, A+B+C is discarded (but not A+C+D)

# II: Government model – Model extension

**Extension:** Higher-order Bayesian model, incorporation of expert's uncertainty.

- Ranking A+B before A+C doesn't mean that expert is sure that A+B is chosen, but rather with some (unknown) uncertainty e.g. 90/10% or 70/30%
- Restricted dirichlet distribution with P(A+B) > P(A+C)
- Set $\alpha_k$ to 0.5 or estimate optimal $\alpha_k$ with 2017 election data (ca. 0.3 via ML-estimation)
- For $\alpha_k = 0.3$ and 2 potential coalitions with P(A+B) > P(A+C): 86.7% for P(A+B); 13.3% for P(A+C)

**Strategy:**

For each coalition scenario:

$\rightarrow$ Sample e.g. n = 1000 times from dirichlet with number of categories set to number of potential coalitions

$\rightarrow$ order returned samples by size

$\rightarrow$ draw n times from multinomial with sampled probabilities for each expert

$\rightarrow$ draw n coalition probabilities from original multinomial-dirichlet model

# II: Government model – 2017 predictions

**Recall:** In 2017 failed attempt of CDU/CSU-FDP-Grüne coalition forming after 2 months after which CDU/CSU-SPD coalition constituted government of Germany

**Predictions on day before 2017 elections**:

| Coalition_parties | simple model | extended model |
|---|---|---|
| CDU/CSU-SPD | 0.599 | 0.572 |
| CDU/CSU-FDP-Grüne | 0.302 | 0.345 |
| CDU/CSU-FDP-~~Grüne~~ | 0.081 | 0.064 |
| CDU/CSU-Grüne | 0.015 | 0.015 |

- Predictions reasonable
- Simple and extended model give similar results
- Chancellor probability for Angela Merkel >99%

# Summary

Goal: Forecast German federal election results

Secondary Goal: Automatize modeling and publish regularly on website

Features:
- Forecast vote shares and uncertainties
- Assign probabilities to events of interest
- Coalition probabilities for final government
- Chancellor probabilities

Data: > 20 years of poll data:
- Historic election results
- Historic government and opposition parties
- Expert survey on coalition ranking

Model Approach:
- Bayesian state-space model with Stan
- Conjugate Bayes model for government formation

INWT Statistics

*Thank you for your time!*


**INWT** Statistics GmbH
Hauptstraße 8
Meisenbach Höfe, Aufgang 3a
10827 Berlin

+49 30 12082310

info@inwt-statistics.de

www.inwt-statistics.de