Structural and Trial-Heat Model Combinations to Forecast US Elections

Marco Zanotti

University Milano-Bicocca

Contents

1. Election Forecasting

Problem

To predict **timely and accurately** the election results

Strategical task in the political area since political forces spend millions of dollars in each candidate's campaign and need to know when and where to allocate them

The **US**: presidential election is the result of the voting process in each state

2. Data

Type of Data

Forecasting elections makes use of mainly two different types data:

- **Fundamental indicators**, that is economic or political variables
- **Trial-heat polls**, that is surveys with trail-heat questions issued by official pollsters' agency

Fundamental Indicators

The **economy** strongly affect and anticipate election results. Among the most used economic indicator there are GDP, GNP, unemployment, inflation at national or state level

The **political dimension** of election is also of high relevance and it is usually measured by incumbency, time-for-change, votes of previous elections, presidential home-state advantage, president approval rating

Many models have been developed using only such data and predicted the results within few percentage points

Trial-Heat Polls

Survey responses are related to actual voting process, meaning that polls are connected to observable political behaviours and incorporate the process of updating information of individuals, so that can be used to track the evolution of preferences over time and states

Election polls data suffers of some **well-known problems** such as sampling errors, house effect, question wording, response errors and high variability

Nowadays many pollster agencies exist, producing surveys both at the national and state levels, in particular during the election year

3. Methods

Type of Models

Over the years, three types of election forecasting models evolved:

- **Structural models**, econometric models based on fundamental indicators
- **Trial-heat models**, econometric models relying on polls data
- **Bayesian models**, models that use polls data to update historical forecasts, improving the performance of structural models through the incorporation of voters preferences' evolution

The variable of interest is usually the percentage election outcome of one of the two major parties (π_t)

Structural Models

The **Time-for-change model** is one of the most successful and was proposed by *Abramowitz* in 1988 (and again in 1996 and 2008)

$$\pi_t = \beta_0 + \beta_1 GDP_{t-1} + \beta_2 Approval_t + \beta_3 TC_t$$

It assumes that voters positively evaluate periodic government alternation of the two major parties

It relies only on previous elections' data without incorporating the opinion about the current election

Trial-Heat Models

Using trial-heat polls as **literal forecast** produce very poor results, because the accuracy of election polls depends enormously on when the poll is conducted

Campbell improved the poor trial-heat literal prediction suggesting a simple regression model that used only trial-heat polls at national level and GDP

$$\pi_t = \beta_0 + \beta_1 polls_{t-1} + \beta_2 GDP_{t-1}$$

Information at national level only

Trial-Heat Models

Gelman & King incorporated current polls information within a more complex structural model considering the aggregate trial-heat two months before the election, incumbency, GNP rate, approval rating and state variables

$$\pi_{it} = \beta_0 + \beta_1 polls_{t-1} + \beta_2 GNP_{t-1} + \beta_3 Incumbency_t + \\ + \beta_4 Approval_t + \beta_i State_i$$

It is a state level model but polls data was used as a national and aggregated information

Bayesian Models

Since the late 90s, methods following a Bayesian approach have been introduced also in the context of election prediction

Bayesian models naturally follow the **“voters’ enlightenment”** hypothesis because the weights voters attach to fundamental variables are allowed to change during the campaign, accounting for changes in public opinion

Core idea: to use polls data to update historical forecasts, accounting for current voters’ preferences and improving the performance of structural models

Bayesian Models

Brown & Chappell proposed a model averaging weighting two models by their posterior model evidence. The **hist** equation represents structural model, while the **poll** equation is the polling model

$$\pi_t^{hist} = \beta_0 + \beta_1 GDP_{t-1} + \beta_2 Incumbency_t + \epsilon_t$$
$$\pi_t^{poll} = \alpha_0 + \alpha_1 S_t + u_t$$
$$\pi_t = w^{hist}\pi_t^{hist} + w^{poll}\pi_t^{poll}$$

The prediction is a weighted average. The weights are based on the proportion of the variances of the error terms of the two regressions

Constantly updates the historical forecasts

Bayesian Models

Rigdon et al. created a **Dirichlet-Multinomial** model with a conjugate prior based on past election results and a likelihood based on current poll data to estimate each candidate's probability of winning a state

$$p = (p_1, p_2, p_3, p_4) \sim Dirichlet(b_1, b_2, b_3, b_4)$$
$$X = (X_1, X_2, X_3, X_4) \sim Multinomial(n, p_1, p_2, p_3, p_4)$$
$$P(p|X) \sim Dirichlet(x_1 + b_1, x_2 + b_2, x_3 + b_3, x_4 + b_4)$$

where X_i are the sample proportions in a state poll and p_i are the shares in a state of candidate i

Takes into account the proportions of **third-party candidates and undecided**

Bayesian Models

Lock & Gelman used a **Normal-Normal** model assuming normality of both the prior, based on historical election results, and the likelihood, based on the poll data

$$\pi_{it}^{hist} \sim N(\pi_{i,t-1}, \sigma_h^2)$$
$$\pi_{it}^{poll} \sim N(\pi_{i,0}, \sigma_p^2)$$
$$\pi_{it} \sim N\left(\frac{\frac{1}{var(\pi_{it}^{poll})}\pi_{it}^{poll} + \frac{1}{\sigma_h^2}\pi_{i,t-1}}{\frac{1}{var(\pi_{it}^{poll})} + \frac{1}{\sigma_h^2}}, \frac{1}{\frac{1}{var(\pi_{it}^{poll})} + \frac{1}{\sigma_h^2}}\right)$$

States with higher prior precision place more weight on past election results and viceversa (this happened for almost every state)

Bayesian Models

Linzer unified historical forecasts based on **structural variables** with state-level **poll data**

$$\pi_{it} = logit^{-1}(\beta_{it} + \delta_t)$$
$$\beta_{i,t-1} \sim N(\beta_{i,t}, s_\beta^2) \qquad \delta_{t-1} \sim N(\delta_t, s_\delta^2)$$
$$\beta_{iT} \sim N(logit(h_i), \sigma_i^2)$$

where π_{it} is the share of voting preferences for state i and day t, β_{it} represents the historical voting preferences in state i and δ_t is a national effect and h_i is the estimate from the TFC model

The posterior probability the Democratic candidate wins in state i is calculated as the proportion of draws from π_{iT} greater that 0.5

4. Conclusions

Conclusions

- A Bayesian approach produces **continuously revised forecasts** as new poll data is released
- **Comparing performances** of different models is difficult (different elections, different horizons, different purposes)
- In general, forecasts are accurate **within 2 months** before the election day and forecasting using both structural variables and poll data outperform others
- It is **still difficult** to produce timely and accurate forecasts
- **Problems** arise in forecasting accuracy and uncertainty for states that are polled few and in those days with no polls at all

Conclusions

- **Web data** as a source of spontaneous public opinions to solve lack of data issues (as *Rizk et al., 2023*)
- **Correlated vote intentions** across states (for example, if a candidate is performing bad in a state, this might indicate that he will also underperform in other states)

These factors may help to improve early forecast accuracy of the models

Bibliografy

Abramowitz, A. I. (2008), 'Forecasting the 2008 Presidential Election with the Time-for-Change Model.', PS: Political Science and Politics 41(4), 691–695.

Brown, L. B. & Chappell, H. W. J. (1999), 'Forecasting presidential elections using history and polls.', International Journal of Forecasting 15(2), 127–135.

Campbell, J. E. (1996), 'Polls and Votes: The Trial-Heat Presidential Election Forecasting Model, Certainty, and Political Campaigns', American Politics Research 24(4), 408–433.

Gelman, A. & King, G. (1993), 'Why Are American Presidential Election Campaign Polls So Variable When Votes are So Predictable?', British Journal of Political Science 23(1), 409–451.

Bibliografy

Linzer, D. A. (2013), 'Dynamic Bayesian Forecasting of Presidential Elections in the States', Journal of the American Statistical Association 108(501), 124–134.

Lock, K. & Gelman, A. (2010), 'Bayesian Combination of State Polls and Election Forecasts.', Political Analysis 18(3), 337–348.

Rodrigue Rizk, e. a. (2023), '280 Characters to the White House: Predicting 2020 U.S. Presidential Elections from Twitter Data.', Comput Math Organ Theory.

Steven E. Rigdon, e. a. (2009), 'A Bayesian Prediction Model for the U.S. Presidential Election.', American Politics Research 37(4), 700–724.

Thank you!