Dynamic Forecasting of US Elections

Marco Zanotti
University of Milano Bicocca

June 19, 2023

Abstract

The text of your abstract 200 or fewer words.

Keywords: election forecast, bayesian modelling, polls, web data

1 Introduction

Purpose Background Challenges Gaps Directions

point out any controversy in the field

Voters at least base their decisions on relatively known and measurable variables [gelman 1993] These fundamental variables measure their interests and include economic conditions, party identification, proximity of the voter's ideology and issue preferences to those of the candidates, etc. All the serious forecasting methods try to predict the election result using some versions of the same fundamental variables to measure economic well-being, party identification, candidate quality and so forth.

1.1 Elections may be hard to predict

Nonostante la previsione nazionale è considerata essere prevedibile per via del fatto che il risultato è considerato essere basato su variabili fondamentali che sono in place before the election campaign (for instance the economic situation of the US and forte senso di appartenenza dei cittadini americani ad uno dei due partiti)

First, close elections will always be hard to predict since in these cases the best possible forecast will be statistically indistinguishable from 50%.

In primaries, low-visibility elections, and uneven campaigns, or uninformed elections we would not expect forecasting based on fundamental variables meas- ured before the campaign to work. The fast-paced events during a primary campaign (such as verbal slips, gaffes, debates, particularly good photo, opportunities, rethorical victories, specific policy proposals, previous primary results, etc) can make an important difference because the can affect voters' perceptions of the candidates' positions on fundamental issues. Also, primary

election candidates often stand so close on fundamental issues that voters are more likely to base their decision on the minor issues that do separates the candidates.

Moreover, the inherent instability of a multi-candidate race.

Difficulty within some well-known states.

The outcome of elections with uneven campaigns would also be hard to predict based on fundamental variables alone.

However, in the general election campaign for the president (high information, balanced campaigns) these events are ephemeral having little effect on the final outcome. gelman 1993

1.2 "Mental" Process of Voters (gelman 1993)

A well-accepted hypothesis of voters process during election is the so called enlightened preferences of **?**. Essentially, voters based their preferences on fundamental variables and the function of the electoral campaign is to inform individuals about them and their appropriate weights. Hence, individuals are not rational but use increasing amounts of information over the campaign. At the beginning of the campaign voters have low level of information and this is reflected in polls answers, while the day before the election the voters have full information. Essentially the voters information set improves over the course of the campaign.

Based on this assumption a model aiming at forecasting the presidential election correctly has to incorporate the process of "voters' enlightment", implying that, since the values of the fundamental variables do not change, the weigths respondents attach to these variables have to change during the campaign, accounting for changes in public opinion.

2 Body

explain data experimental evidence: describe important results

3 Data

List, differences in the availability Nation variables or state variables

3.1 Economic and Political Indicators

Economy matters! An in-party presidential candidate running in the context of a booming economy would win a greater share of the vote than with a slugghish economy.

Given the relevance of the topic, numerous researchers over many decades discovered and analysed the importance of some economic variables that strongly affect and anticipate the election results.

This data is typically available before the start of the campaign.

growths in GDP, GNP, unemployment, inflation in the last available quarter before the elections growths in gdp (or others) by states

incumbency (usually a dummy), national and states vote of last two elections, presidential home-state advantage, partisanship of a state (proportion of democrats in last legislature), president approval rating (officially estimasted by national agencies), this consider he's foreign affairs, personal style, communication skills, honesty, integrity, domestic agenda, etc distance between state and candidate ideologies, state's religion, time-for-change (if a party has controlled the White House for two or more terms)

Sometimes also regional political variables have been adopted to highlight southern and northern differences mainly (used to remove anomalies in previous elections)

Many models have been developed using only such data and predicted the results within few percentage points. Forecasting models based on economic and political variables measured before the start of the campaign have performed well in the past.

PROBLEMS: - more recent economic changes are difficult to incorporate directly to economic variables since data is usually not available yet

3.2 Individuals Indicators

party, ideology, race, sex, income, education, religion, region

3.3 Trial-Heat Polls

horse-race aspects, interpreting each short-term change in the public opinion polls as a serious change in the likely fortunes of candidates

Data before 1988 are usually from Gallup, then other polling organizations emerged and are used too.

Initially only national polls then also state level polls

One can safely merge data from the different polling organizations in order to study trends in candidate support but not the percentage of undecided or not responding. Gelman 1993

The polls converge to a point near the actual election outcome shortly before the election day

Even if early polls in most election years appear to have very little to do with the eventual

outcome of the general election, much evidence exists to conclude that survey responses are related to actual voting process, notably the predictive accuracy of polls immediately after the election. Hence polls are connected to observable political behaviours and incorporates the process of updating information of individuals.

Moreover, can be used to track the evolution of preferences over time and states.

PROBLEMS: - random sampling errors (representativeness) - response errors - question wording - different organization produce systematically different results (organization bias) - high variability in the support for the Democratic and Republican candidates - non-response bias, when the candidate is going bad selectively decide not to answer or saying they do not vote - are affected in the events of the campaign - data limitation (availability) for states but no more a big issue

ONE PRO: - indirectly incorporate more recent economic changes

3.4 Web Conversations

4 Methods & Models

As Gelman & King (1993) pointed out, one of the problems of models based solely on economic and political indicators was that they were based on a single regression specification based only on previous elections.

Using trial-heat polls as literal forecast produce very poor results too, because of all the limitations of the polls. Indeed, the accuracy of trial-heat polls in predicting presidential elections depends to a substantial degree on when during the election year the poll is conducted. It is commonplace now to dismiss early polls as meaningless (same as flipping

a coin) and late polls as obvious. [campbel 1996]

Undecided or non-major party vote are usually discarded or evenly divided between the two major parties.

4.1 Abramowitz 1988 - 1996 - 2008

In 1988 Abram proposed the well-known Time-for-Change model The assumption is that a presidential election is a referendum on the performance of the incumbent president hence how voters cast their ballots should be strongly influenced by their evaluation of the incumbent president's performance. Captures the strength of public sentiment for changes in the government party (based on the hypothesis that voters attach positive value to periodic alternation between the two major parties).

gdp growth second quarter + approval rating of incumbet president + length of time the incumbent president's party has controlled the White House (time for change factor)

4.2 Gelman King 1993

trial-heat 2 months before + incumbency + GNP changes + state variables (state votes last 2 elections + home advantage + partisanship of state + ideology) + regional variables + approval rating + distance between state and candidate ideology

4.3 Campbell 1996

improved trial-heat literal with trial-heat + gdp changes drawback only national

4.4 Brown Chappel 1999

Bayesian model that uses both polls and historical data and allows poll data to be assimilated in an optimal and timely manner to update an earlier forecast Uses time series data over elections

only simple model and generalized with description of regressors

Found gains in using polls to augment historical regression data. Weighted average forecasts have lower MSE than historial or trial-heat only.

5 Conclusion

summarize major points point out significance of results questions that still remain to address

web data + ensembling

By treating forecasting as a Bayesian updating problem, we are able to produce continuously revised forecasts as new poll data are released in the course of the campaign. Allowing to account for the process of voters and incorporating the changing weights assigned to the fundamental variables.

Forecasting using both historical fundamental variables and poll data outperform those based on fundamentals or polls alone (even at the state level)

Forecasts are usually consistently accurate in the 2 months before the election.

6 Notes

Abramowitz (2008) offer some guidance about key ideas about statistical ideas. (Abramowitz 2008) (Brown & Chappell 1999) (Campbell 1996) (Gelman & King 1993) (Lock & Gelman 2010) (Steven E. Rigdon 2009) (Linzer 2013) (Rodrigue Rizk 2023)

[Consistency comparison in fitting surrogate model in the tidal power example.]{#fig-first width=3in}

Table 1: D-optimality values for design X under five different scenarios.

one	two	three	four	five
1.23	3.45	5.00	1.21	3.41
1.23	3.45	5.00	1.21	3.42
1.23	3.45	5.00	1.21	3.43

References

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.

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.