Qualitative Analysis of Election Forecasts

1. Introduction

1.1 Background

The field of election forecasting has significantly evolved, becoming a pivotal aspect of political analysis. This research, a joint effort by Aaryan Sharma (UIN: 661307552) and Moath Abdallah (UIN: 662216728) under the guidance of Professor Ian Kash, delves into the intricacies of election forecasting. The accuracy and complexity of these forecasts have seen considerable improvements with advancements in statistical methods and data availability, underscoring their importance in understanding voter behavior and political dynamics.

1.2 Importance of Accurate Forecasting

Accurate election forecasting guides political campaigns, informs voters, and shapes policy-making. It also holds significance in media and public discourse, influencing the narrative around electoral campaigns. However, forecasting elections accurately remains challenging due to the unpredictable nature of human behavior and the multitude of influencing factors.

1.3 The Role of the Brier Score in Forecast Evaluation

In this study, we emphasize using the Brier score, a metric that quantifies the accuracy of probabilistic predictions by measuring the difference between predicted probabilities and actual outcomes. This provides a reliable means to assess and compare the performance of various election forecasting models.

1.4 Research Objectives

Our primary objective is to comprehensively evaluate election forecast models using qualitative and quantitative analyses, with the Brier score as a central metric. Specific goals include:

We are replicating and operationalizing renowned models from platforms like FiveThirtyEight and The Economist as well as a model from an individual user on GitHub.

We are refining R/Python code for enhanced Brier score calculations.

1.5 Structure of the Paper

Structured to provide a holistic view of our research, the paper progresses as follows: Section 2 encompasses a detailed literature review, laying the theoretical and empirical groundwork. Section 3 outlines our methodology, describing the various techniques and processes employed. The subsequent sections discuss the results, insights, and conclusions derived from our study, culminating with recommendations for future research in this domain.

2. Literature Review

2.1 Introduction to Forecasting and the Brier Score

Our literature review begins with a foundational understanding of forecast evaluation, primarily through the lens of the Brier score. The seminal work, as suggested by Professor Kash, is the classic article detailing the problem of forecast evaluation and introducing the Brier Score[1]. This score, as further elaborated on Wikipedia[2], has become a standard metric for assessing the

[1] Department of commerce Charles Sawyer, secretary F. W ... - USC. Accessed November 23, 2023. https://viterbi-web.usc.edu/~shaddin/cs699fa17/docs/Brier50.pdf.
[2] "Brier Score." Wikipedia, August 17, 2023. https://en.wikipedia.org/wiki/Brier_score.

accuracy of probabilistic predictions in various fields, including meteorology and finance, and is now extensively applied in election forecasting.

2.2 Challenges in Election Forecasting

Delving deeper into election forecasting, we explore the specific challenges it presents. An article from the Journal of Decision Making[3] provides insights into the complexities of forecasting elections. This includes the unpredictability of voter behavior, the impact of external socio-political factors, and the dynamic nature of public opinion.

2.3 Review of Established Forecast Models

We then turn our attention to examining established forecast models. FiveThirtyEight, renowned for its data-driven approach, provides a comprehensive forecast model, as seen in their 2022 election forecast[4]. Their interactive tools[5] and a repository of prediction data[6] serve as valuable resources for understanding modern forecasting techniques.

Similarly, The Economist's model[7] and an accompanying academic paper[8] offer insights into their methodology, which combines statistical analysis with journalistic expertise.

On the other hand, Pkremp's model[9] is just that one of a solo passion project and is not funded by a wealthy company.

2.4 Diverse Approaches in Forecast Modeling

---

[3] Information, incentives, and goals in election forecasts. Accessed November 23, 2023. http://www.stat.columbia.edu/~gelman/research/published/jdm200907b.pdf.

[4] NateSilver538. "2022 Fivethirtyeight Election Forecast." FiveThirtyEight, November 8, 2022. https://projects.fivethirtyeight.com/2022-election-forecast/.

[5] Ryanabest. "Explore the Ways Republicans or Democrats Could Win the Midterms." FiveThirtyEight, November 8, 2022. https://projects.fivethirtyeight.com/2022-flip-senate-house/.

[6] "Election-Forecasts-2022." GitHub. Accessed November 23, 2023. https://github.com/fivethirtyeight/data/tree/master/election-forecasts-2022.

[7] "The Economist's 2022 Senate Forecast." The Economist. Accessed November 23, 2023. https://www.economist.com/interactive/us-midterms-2022/forecast/senate/how-this-works.

[8] Heidemanns, Merlin, Andrew Gelman, and G. Elliott Morris. "An Updated Dynamic Bayesian Forecasting Model for the US Presidential Election." Harvard Data Science Review, October 27, 2020. https://hdsr.mitpress.mit.edu/pub/nw1dzd02/release/1.

[9] "Pkremp/polls". Accessed November 23, 2023. https://github.com/pkremp/polls.

Our review also encompasses models from other sources to highlight the diversity in forecasting approaches. We examine the model developed by researchers Enns and Lagodny, which provides a methodological contrast to the models by FiveThirtyEight, The Economist and Pkremp. Additionally, models from UIUC (Result | Election Analytics @ Illinois), YouGov (Biden vs. Trump | YouGov's 2020 Presidential Election Model), and JHK Forecasts (Forecast Methodology | jhkforecasts.com) are reviewed to understand further the spectrum of methodologies employed in contemporary election forecasting.

2.5 Collaborative Approach to Research

Reflecting on the methodology proposed by Professor Kash, we emphasize the collaborative nature of our research. Using a shared set of Google Slides for organizing and comparing models aligns with modern research practices, promoting transparency and collective learning.

2.6 Conclusion and Path Forward

In summarizing our literature review, we underscore the importance of a multifaceted approach to election forecasting, considering both the technical aspects of model construction and the unpredictable nature of political landscapes. Our review sets the stage for a detailed exploration of forecasting models, their evaluation using the Brier score, and the pursuit of innovative approaches in predicting election outcomes.

3. Methodology

3.1 Data Collection and Preprocessing

Our study commences with the meticulous collection of historical election data, including polling figures, voter demographics, and economic indicators that have demonstrated relevance to election outcomes. This data is primarily sourced from authoritative platforms such as FiveThirtyEight, The Economist, and other established political analysis entities. The integrity of our research is contingent upon the accuracy and consistency of this data; hence, a rigorous preprocessing protocol is employed. This includes standardizing data formats, handling missing values, and verifying the accuracy of data points.

3.2 Model Operationalization

A critical component of our methodology involves replicating and operationalizing existing forecast models, primarily those developed by FiveThirtyEight, The Economist and Pkremp. This process uses R and Python, emphasizing fidelity to the original model specifications. Additionally, based on preliminary data analyses and insights from our literature review, we may integrate novel variables or modify these models to enhance their predictive capacity.

3.3 Brier Score Calculation

The Brier score, a metric for assessing the accuracy of probabilistic predictions, is central to our evaluation process. We adapt existing R/Python scripts to compute the Brier score for each forecast model. This computation facilitates a quantitative assessment of each model's predictive accuracy and allows for a comparative analysis among different models. The Brier score's utility in our study lies in its ability to quantify the magnitude of prediction errors, providing a clear benchmark for model performance.

3.4 Forecast Generation and Evaluation

Single-State Brier Score Calculation

The initial approach in our forecast evaluation employs the Brier score calculation based on single states. This method involves averaging the squared differences between the actual election outcomes and the predicted probabilities for each state. Formally, it is expressed as *avg((reality−prob)^2)*. This granular approach allows us to assess the accuracy of our models at the state level, providing insights into their performance in predicting state-specific outcomes.

Two-State Computation

In addition to the single-state approach, our methodology extends to a two-state computation. This involves extracting the combined probabilities for pairs of states from our models and subsequently calculating the Brier scores for these pairings. This is done by using the formula *avg((conditional*prob−reality)^2)*. To elaborate, the conditional probability is created by taking two states, state1 and state2, and setting a candidate to win state2. The updated score of a candidate to win state1 is the conditional probability. This score is multiplied by the original probability to win state2. This equation is run four times per pair of states depending what candidate wins what state. It is to be noted that not all 50 states are used in this method as the models prevent certain states from being won by a candidate if it deems it impossible. This method offers a nuanced understanding of the models' ability to capture interdependencies and correlations between the electoral outcomes of different states.

Integrated State Probability Method

Building upon the first two methods, we utilize both single-state and two-state probabilities to devise a more comprehensive approach. This integrated method aims to leverage the strengths of both approaches, providing a more holistic view of the model's predictive accuracy. The formula for this method is *avg(brier score of state1 + brier score of state2 + two-state brier score).* We also explored having the two-state brier score weighted more than it currently is in the formula above, 50% instead of the current 33%. Each model has run this method four times, one for each combination of winning candidates. By combining these methods, we can capture the individual and interactive effects of state outcomes on the overall forecast accuracy.

Weighted based on Electoral College

This method is a twist on the Single-State Brier Score Calculation. In the original method, each state was equally weighted when creating the average score. However another approach we took on this method was to give weight to each state based on the number of electoral votes they have. When we did this weighted average on the brier scores, both models noticed a significantly better score than their unweighted counterparts. This is likely due to the fact that many big states such as NY, IL, CA are worth a lot of electoral votes and are very predictable leading to much better brier scores.

Log Based

To test a method of grading forecasts not using brier scores, we used a log based method with the equation of $log(|prob - reality|)$. Since log(1) = 0, the prediction is completely off, and log(0) = -INF, the prediction is perfect, the closer the prediction is to -INF the better the model is. By using this method we were able to see if that pattern we were noticing with our brier scores would also carry over in different methods of grading forecasts.

Vote Share Variation Exploration

This final method is one we have spent the most time on aside from the main brier score methods used. We reverse engineered the given probabilities we had been given from the models and converted them into vote shares. With these vote shares we were able to come up with multiple ways of grading the forecasts. The first method used the equation *avg((predicted vote share−actual vote share)^2)* which is practically the original brier score method but with vote shares. The second method we tried out is the same as the first one but only on the states we deemed to be competitive. The third and final method uses pairs of states and the equation used is $((p1-a1)^2+(p2-a2)^2+((p1)^2-(a1)^2)^2+((p2)^2-(a2)^2)^2+((p1)*(p2)-(a1)*(a2))^2)/5$. This method is in parallel to our Integrated State Probability Method as it combines both single state evaluations and two state evaluations.

3.5 Qualitative Analysis

In addition to quantitative metrics like the Brier score, our study also incorporates a qualitative analysis of the forecast models. This involves critically examining the models' components, such as how they account for various socioeconomic and political factors and their overall structure. We aim to identify the strengths and limitations of each model and understand the underlying reasons for any significant discrepancies between the model predictions and the actual election outcomes.

3.6 Visualization and Communication

An essential aspect of our research is the effective communication of our findings. This is achieved by developing comprehensive visualizations that succinctly convey our forecasts' results and the Brier scores' comparative analyses. These visualizations are designed to be intuitive, allowing both academic and non-academic audiences to grasp the nuances of our evaluations.

3.7 Continuous Improvement and Feedback Loop

A structured feedback mechanism, essential for the dynamic and evolving nature of election forecasting, enriches our methodology. Central to this process are the weekly meetings organized by our research team. These sessions serve multiple purposes:

4. Results

GitHub with textfiles for results from 4.1 and 4.2

https://github.com/mabda4/Text-files-CS398

4.1 FiveThirtyEight - 2020 Election

GitHub containing all changes made to FiveThirtyEight's model:

https://github.com/mabda4/2020-FiveThirtyEight-model-modified.

The average state wide brier score is 0.08203393

The average state wide (with constraints) brier score is 0.11845

The average of the Two-State Computation is 0.08390763

The average of the Integrated State Probability Method is:

(BB) 0.0953193

(TB) 0.1243869

(BT) 0.2533886

(TT) 0.4327351

The average of the four Integrated State Probability Methods is 0.226457475

The average of the Integrated State Probability Method with more weight towards the two state comparison:

(BB) 0.127747920541126

(TB) 0.135336491969697

(BT) 0.223190773300866

(TT) 0.334229675898268

The average state wide brier score when weighted by the electoral votes is 0.0012434698558322

The average of the log grading method is -0.303790432997695

The vote share brier score is 0.00134894117647059

The vote share (with constraints) brier score is 0.00086975

The vote share third method is 0.00271949100002399

4.2 Pkremp - 2016 Election

GitHub containing all changes made to Pkremp's model:

https://github.com/mabda4/2016-modified-forecast-model.

The average state wide brier score is 0.0795007

The average state wide (with constraints) brier score is 0.2017719

The average of the Two-State Computation is 0.1673443

The average of the Integrated State Probability Method is:

(CC) 0.18284

(TC) 0.3168405

(CT) 0.3169946

(TT) 0.4450556

The average of the four Integrated State Probability Methods is 0.315432675

The average of the Integrated State Probability Method with more weight towards the two state comparison:

(CC) 0.181053459194776

(TC) 0.273854611090346

(CT) 0.273886805785613

(TT) 0.383899836304335

The average state wide brier score when weighted by the electoral votes is 0.0021392785642857

The average of the log grading method is -0.358404176414859

The vote share brier score is 0.00330192584515034

The vote share (with constraints) brier score is 0.00218779257480633

The vote share third method is 0.00837492332862397

4.3 The Economist - 2016 Election

GitHub containing all changes made to Economist model:

https://github.com/sharmaaaryan4012/Research-Project---Election-Forecast

The average state wide brier score is 0.05268627451

The average state wide (with constraints) brier score is 0.13295

The average of the Two-State Computation is:

(CC) 0.09858197895

(CT) 0.1123547842

(TC) 0.1129008466

(TT) 0.1496829211

The average of the four Two-State Computation is 0.1182315789

The average of the Integrated State Probability Method is:

(CC) 0.120593993

(CT) 0.1251849281

(TC) 0.1251688772

(TT) 0.1376276404

The average of the four Integrated State Probability Methods is 0.1271438596

5. Conclusion

5.1 Comparison of 2016 models

The two models we have been analyzing that cover the 2016 presidential election have been from Pkremp, a user on GitHub, and The Economist. An assumption we had before comparing any analytics is that The Economist would have more reliable scores due to the fact that it is a well established company and that their model for the 2016 election came out after the election had already occurred. It turns out that our assumption would be spot on. On our first method of testing, the average state wide brier score, The Economist's model does decently better. When this average is run with the constraint of using competitive states, both scores got worse with The Economist's change being less than Pkremp's. The same relationship is seen in the Two-State Computation. However an interesting relationship is found in the Integrated State Probability Method. The Economist's model's score stays roughly the same for all scenarios but Pkremp's scores sky rocket when a condition that Trump wins a state or both is introduced. We believe this to be due to the fact that Pkremp's model severely underrated Trump as it gave him only a 10% chance of winning the election. So this led to when being put in scenarios of Trump winning a certain state, it would fall short. On top of that, The Economist model came out after the election was decided, so they were able to properly rate Trump. So overall, in any test The Economist holds up better.

5.2 Comparison of 2016 and 2020 models

Generally speaking, almost all the scores produced by the FiveThirtyEight model were better than either model for the 2016 presidential election. The conclusion we came up with is a mixture of two reasons. First, since the FiveThirtyEight model came out for the 2020 election, naturally it would be better since it is simply newer and can see what old models got right or wrong and grow based on that. Secondly, we looked at real world events that would cause such a difference in the model's results. What we came up with is the fact that the 2016 presidential election was an upset, most forecasts gave the victory to Clinton yet in reality Trump won;

however in the 2020 election, most forecasts were correct in predicting Biden to be the winner. Therefore simply by having a more predictable outcome, FiveThirtyEight's model has an edge over Pkremp or The Economist's models by default.

Error Identification and Resolution: The weekly meetings provide a platform for identifying and addressing errors or inconsistencies in our models or data. This collaborative problem-solving approach ensures swift rectification of issues, thereby maintaining the integrity and accuracy of our forecasts.

Strategic Planning: These gatherings are also pivotal for planning future research activities, discussing methodological improvements, and strategizing how to adapt our models to emerging political and social trends. This forward-looking perspective is crucial for keeping our research relevant and impactful.

Collaboration and Knowledge Sharing: To foster a collaborative research environment and ensure seamless communication among team members, we utilize Google Slides as a central repository for our research data. This platform allows us to:

1. Compile and organize our findings, methodologies, and data analyses in an accessible format.
2. Facilitate easy sharing of insights and developments with other researchers and stakeholders involved in the project.

3. Enable real-time updates and collaborative editing, enhancing the efficiency and productivity of our team.

Implementing this feedback loop, punctuated by regular meetings and supported by collaborative tools like Google Slides, ensures continuous improvement in our forecasting models. This approach enhances our predictions' accuracy and encourages a culture of transparency, collaboration, and innovation within our research team.

6. Bibliography

"Brier Score." Wikipedia, August 17, 2023. https://en.wikipedia.org/wiki/Brier_score.

Department of commerce Charles Sawyer, secretary F. W ... - USC. Accessed November 23, 2023. https://viterbi-web.usc.edu/~shaddin/cs699fa17/docs/Brier50.pdf.

"Election-Forecasts-2022." GitHub. Accessed November 23, 2023. https://github.com/fivethirtyeight/data/tree/master/election-forecasts-2022.

Heidemanns, Merlin, Andrew Gelman, and G. Elliott Morris. "An Updated Dynamic Bayesian Forecasting Model for the US Presidential Election." Harvard Data Science Review, October 27, 2020. https://hdsr.mitpress.mit.edu/pub/nw1dzd02/release/1.

Information, incentives, and goals in election forecasts. Accessed November 23, 2023. http://www.stat.columbia.edu/~gelman/research/published/jdm200907b.pdf.

"Pkremp/polls". Accessed November 23, 2023. https://github.com/pkremp/polls.

NateSilver538. "2022 Fivethirtyeight Election Forecast." FiveThirtyEight, November 8, 2022. https://projects.fivethirtyeight.com/2022-election-forecast/.

Ryanabest. "Explore the Ways Republicans or Democrats Could Win the Midterms." FiveThirtyEight, November 8, 2022. https://projects.fivethirtyeight.com/2022-flip-senate-house/.

"The Economist's 2022 Senate Forecast." The Economist. Accessed November 23, 2023. https://www.economist.com/interactive/us-midterms-2022/forecast/senate/how-this-works.