2024 US Presidential Election Forecasting*

Trump forcasted to win popular vote by state over Harris by 0.8%

Mariko Lee   Karen Riani   Cristina Su Lam

November 1, 2024

We forecast the 2024 U.S. presidential winner using state-level poll averages for Kamala Harris and Donald Trump, weighted by poll quality and sample size. Applying a Bayesian model, we find a slight 0.5% lead for Trump over Harris, using polls taken since the reelection bid on July 21, 2024. Weighting the polls increases the influence of the most reliable data in shaping the forecast—limiting skewed results and showing a clearer voter sentiment amidst a particularly competitive election. These results show who may be leading the elections via polls, giving campaigns the insights they need and reinforcing public trust in the election process.

1 Introduction

Amid the chaotic, high-stakes 2024 U.S. presidential election, forecasting models offer crucial insights into voter sentiment as Americans face a tight race between top candidates Kamala Harris and Donald Trump. With the attempted assassination of Trump, Joe Biden's endorsement of Harris, and recent crises such as hurricanes disrupting voter access in swing states, public opinion remains volatile (C. News 2024; A. News 2024b). Key events like Trump's policy proposals on social security and public safety, Harris's advocacy for reproductive rights, and other rising debates on gun control have further polarized the electorate, making forecasting models essential in offering clarity amid this competitive landscape (Watch 2022).

This study joins the growing body of forecast models—including media outlets' own polls-of-polls approaches and various Bayesian methods—by refining the aggregation of poll data with Bayesian regression, weighted by sample size and pollster rating (Times 2024; Press 2024). In an environment where neither candidate shows a decisive lead, this model aims to add precision by emphasizing out higher-quality polls, improving forecast reliability for strategic decision-making and public confidence in election projections (Bloomberg 2024).

---

*Code and data are available at: https://github.com/karenrni/Forecasting-the-2024-US-Presidential-Elections

We estimate a slim 0.5% lead for Trump over Harris in the overall popular vote by aggregating state-level polling data, a close margin that highlights the importance of updated, representative data amidst competing forecasts and fluctuating polls. By using polls weighted by reliability and focusing on data collected after Harris' campaign declaration, this model maintains transparency, offering a robust perspective on public sentiment and helping to clarify perceptions in a highly competitive election.

The following structure of this paper is as follows: Section 2 presents the data sources and methodology, Section 3 presents the forecasting model, then Section 4 followed by Section 5, which discusses the results and their implications. Section 6 concludes with a discussion of the limitations of the study and suggestions for future research.

2 Data

TODO: PROPER FORMATTING FOR DATA SEC, ADD FINAL DETAILS AND FIGURES

2.1 Overview

Our study uses state-level polling data from Donald Trump and Kamala Harris to predict the 2024 US presidential election. The polls used were sourced from FiveThirtyEight's 2024 Presidential Election Forecast Database (FiveThirtyEight 2024), which compiles polling data from multiple organizations and rates each poll's quality. These ratings assist in identifying and weighing the most reliable polls. The dataset used for this analysis spans polls starting from July 21, 2024, when Kamala Harris announced her reelection bid.

Our analysis focuses on the following variables: Polling rating (Pollscore): A numerical score reflecting the historical reliability of polling organizations used to weight polls by accuracy. Sample size: The number of respondents per poll, influencing the poll's margin of error and its weight in the forecast. Percentage support (pct): The proportion of respondents who express support for each candidate. State: The US state where the poll was conducted OR labeled "National" for nationwide polls. Pollster: The organization that conducted the poll, influencing the credibility and weight of our analysis.

Table X - Sample of the cleaned dataset, highlighting key variables crucial in the analysis. [Table]

2.1.1 Measurement and Limitations

We applied several measurement and limitations considerations to ensure the validity of our dataset: Poll Quality: Temporal Dynamics: Geographic Coverage: Response Bias:

2.1.2 Outcome Variables

The Percentage of support (pct) for Donald Trump and Kamala Harris is the main outcome variable. We forecast which candidate is leading by aggregating these results across states. Figure X represents the distribution of support percentages for the two candidates, showing clusters ranging from 40% to 60%, indicating the election's competitive nature.

[Figure X: Distribution of Candidate Support]

2.1.3 Predictor Variables

The following predictor variables were used to build the logistic regression model and refine our prediction: Pollscore: Historical reliability score for pollster (range:) Numeric Grade: (scale: 0-4) Transparency Score: (scale: 0-10) Sample Size: Number of respondents in each poll (typically between 500-3000) Methodology: Survey method (eg. phone interviews, online panels) State: State-level indicators that capture regional variations in support.

These selected variables are based on their relevance to polling accuracy and availability across the dataset. The analysis regarding their relationships with the outcome variables (percentage support) is further explored.

2.1.4 Cleaning Process and Analysis

The data cleaning process involved the following steps: Filtering for High-Quality Polls: We included only polls with a numeric grade of X or higher to reduce the impact of unreliable data. Standardizing Dates and Locations: Variable Transformation: Converted percentages into absolute numbers using sample size Created binary indicators for candidates (Trump = 0, Harris = 1)

Some of our data is of marriage license (**?@fig-marriage**), from (**TorontoOpenData?**)

3 Model

TODO: FINALIZE MODEL SECTION - VALIDATION - ALTERNATIVES - JUSTIFICATION - UPDATE MODEL BELOW

$$\text{logit}(\mu_i) = \beta_0 + \beta_1 \times \text{Pollster}_i + \beta_2 \times \text{State}_i + \beta_3 \times \text{Sample Size}_i + \beta_4 \times \text{Pct}_i$$

- y_i is the dependent variable, representing the count of respondents who support Harris in a given poll, modeled as a binomial outcome. - β_0 is the intercept term, indicating the baseline log-odds of Harris support when all predictors are zero. - β_1, β_2, β_3, and β_4 are the coefficients

for the predictor variables **Pollster**, **State**, **Sample Size**, and **Pct** (the percentage of support for Harris in a poll), respectively: - β_1 represents the adjustment in log-odds based on the specific pollster conducting the survey. - β_2 accounts for the impact of the state in which the poll is conducted. - β_3 adjusts for the influence of the poll's sample size. - β_4 represents the effect of the poll's percentage support for Harris on the log-odds. - The function $\text{logit}(\mu_i)$ transforms the linear combination of predictors into probabilities, providing the estimated probability of support for Harris.

4 Results

TODO: FINISH RESULTS INCLUDING GRAPHS, TABLES, SUMMARY STATS

Overall Percentage for Kamala Harris: 49.72895 % Overall Percentage for Donald Trump: 50.27105 % ... state_winner n 1 Harris 19 2 Trump 23

5 Discussion

TO DO:

FINISH DISCUSSION, EDIT CURRENT BITS

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

What are some weaknesses of what was done? + implications

Environmental and Societal Factors in Surveys, contributing to the – bias

Recent envoronmental disasters have disruoted life in major swing states

Didnt account for partisian pollsters and the potential bias ,

Weaknesses in forecasting itself, How trump could lose despite pollster estimates

Lack of data, could introduce bias, not entirely representative, uncertainty due to declaration for Kamala, This can cause

6 Limitations

What is left to learn or how should we proceed in the future?

Future models could benefit from more complex models which account for other potentially influential factors such as poll methodology (i.e. online vs in-person surveying) and partisanship or respondents' political affiliations (**cambridge_polling_factors_2024?**). Models using polls-of-polls approaches which use multi-level regression and post stratification tend to do particularly well in forecasting elections (REF tab 2), and increasing complexity to account for more influential factors like those mentioned above. They historically perform /…. Better estimation . and accounts for uncertainty propagation, which is particularly relevant in this quarter's election. (Reference in tab) Furthermore, using training data based on historical polls regarding democratic and republican election results and polling data would allow for a more robust model. This is especially because of Biden's late declaration for Harris, thus substituting fot prior state poll behaviour will help account for a lack of polling data. Although many forecasting models aim to estimate the national popular vote, Our model similarly could benefit from estimations for electoral vote

Appendix

A   Methodology Analysis of The Washington Post Polling

With an evaluation of sampling methodology, recruitment, handling non-response, and questionnaire design, this appendix offers an analysis of the polling methodology used by The Washington Post in collaboration with ABC News. The objective is to analyze these approaches' strengths and weaknesses and determine how they affect polling accuracy.

A.1   Population, Frame, and Sample

The Washington Post, in partnership with ABC News, employs a combination of text-to-web polls and random digit dialing (RDD) for landlines and mobile phones to reach a large and representative sample of American adults and registered voters (The Washington Post 2024).

The Washington Post's polling averages use only national and state-level polls that comply with strict quality and transparency criteria. These surveys were chosen because they employ suitable stratification and weighting strategies in addition to random sample approaches (Post 2024). To represent critical demographics such as age, race, gender, and education, the samples are meticulously weighted (The Washington Post 2024) (Post 2024).

A.2   Sample Recruitment

Live phone interviews and text-to-web surveys collect samples for The Washington Post polls, focusing on ensuring comprehensive demographic coverage. In a typical 2024 poll, text-to-web invites reached 21% of respondents, landlines reached 15%, and mobile phones reached 64% of respondents (The Washington Post 2024). Younger and minority voters, who might not be well represented in conventional landline-based surveys, can be efficiently reached by pollsters using this technique.

By using address-based sampling from the Delivery Sequence File of the US Postal Service, ABC News also leverages probability-based recruiting through the Ipos KnowledgePanel. Since internet connections and equipment are offered at no cost, this guarantees that even households without internet connections or digital devices are involved (A. News 2024a).

A.3   Sampling Approaches and Trade-offs

Using stratified random sampling, The Washington Post ensures that important demographic groups are represented proportionately to their voter base. By using stratified sampling, the

polls are more likely to represent the diversity of the voting population accurately. To account for over- or under-representation of particular groups, samples are further weighted (The Washington Post 2024) (Post 2024).

Particularly in situations where state-level polling data is scarce, The Washington Post's polling averages consider the state's voting record in the last two presidential elections (Post 2024). This adjustment offers a more accurate representation of voters' preferences in states with fewer high-quality polls. However, there may be a trade-off since, depending solely on historical data, we may miss recent shifts in voter sentiment (The Washington Post 2024).

A.4 Non-response Handling

The Washington Post uses response weighting, which modifies the results according to demographic variables such as age, race, and education, in order to address non-response bias. In spite of variations in response rates among demographic groupings, this ensures that the final sample more accurately represents the population (Post 2023).

ABC News also addressed non-response bias by applying post-stratification adjustments and sending email reminders to non-respondents. In addition, The Washington Post and ABC News both ensure that their samples are weighted to account for any anomalies in non-response (A. News 2024a) (Post 2023).

Despite these initiatives, non-response bias is still a concern, especially for populations that are less inclined to take part in surveys, including younger or less politically active people (Post 2023).

A.5 Questionnaire Design

To prevent respondents from being guided toward predetermined responses, The Washington Post creates its surveys with neutrality and clarity in mind. The questions are randomized, and respondents are given multiple choices, including "No Opinion," to avoid pressuring answers (The Washington Post 2024). Question order bias can affect how respondents understand and respond to follow-up questions. Therefore, rotation helps mitigate this effect (Post 2023).

Similar ideas are utilized by ABC News, which offers surveys in both Spanish and English to reach a more representative sample of the general public. Leading questions are purposefully omitted from the questionnaires to ensure that the information gathered accurately reflects public opinion (A. News 2024a).

A.6 Strengths and Weaknesses of the Methodology

Strength:
Comprehensive Sampling Method: The Washington Post can reach a broad demographic, including younger and more difficult-to-reach voters, by combining RDD, text-to-web polls, and live phone interviews (The Washington Post 2024) (Post 2023).
Post-stratification Weighting: To account for demographic imbalances and increase the accuracy of their polls, The Washington Post and ABC News both use strong post-stratification weighting (The Washington Post 2024) (A. News 2024a).
Transparent Approach: The Washington Post's polling data is more credible since they only employ high-quality polls in their averages and is transparent about their methodology (Post 2023) (Post 2024).

Weaknesses:
Non-response Bias: Even if both organizations use weighting adjustments, non-response bias still remains a challenge, particularly when it comes to groups that are less likely to respond to surveys (Post 2023) (The Washington Post 2024).
Dependency on Historical Data: In states with fewer polls, The Washington Post relies on historical data (the last two presidential elections), which raises the possibility that the polling averages might not accurately reflect current changes in voters preferences (Post 2024).

A.7 Conclusion

The polling methodologies used by The Washington Post and ABC News offer a strict framework for gauging popular sentiment in the 2024 US presidential election. Their surveys often represent the electorate since they employ various sampling strategies, stratification, and weighting methodologies. However, obstacles such as non-response bias and the use of historical data in some states must be addressed appropriately to protect the accuracy and reliability of their polling averages.

B Idealized Survey & Methodology - $100K Budget

B.1 Overview

Using a $100K budget, this appendix outlines a carefully designed survey methodology for predicting the 2024 US Presidential Election. The objective is to collect representative, high-quality data using recruiting, poll aggregation, and selective sample methods. Through rigorous validation, this approach ensures data accuracy and reduces common survey research errors.

B.2 Sampling Approach

We will implement stratified random sampling to ensure that key demographic and geographic subgroups are fairly represented. This approach reduces bias and offers more reliable insights into voter preference.

Stratification Criteria:
- Age Groups
- Gender
- Education Levels
- Geographic Representation
- Political Affiliation

Sample Size Goal: 10,000 respondents across states and demographics to achieve **high statistical power** with a margin of error below $\pm 1\%$

Trade-offs:
- Although stratified sampling increases representativeness, it necessitates accurate demographic information and may raise operating expenses.
- **Missing Data:** It's possible that some demographics (e.g. men) may be less likely to respond. Post-stratification weighting and data imputation will be used to address this issue.

B.3 Recruitment Strategy

Outline outreach and telephone surveys will be combined in our recruitment strategy to ensure widespread participation from various demographic groups.

Online Recruitment:
- Target ads on Google, Facebook, and Twitter to engage younger voters and urban populations.
- Budget Allocation: $25,000

Random-Digit Dialing (RDD):
- Phone outreach to reach older, rural voters with poor internet connection. - Budget Allocation: $30,000

Incentives:
- Participants are offered $5 gift cards to increase response rates. - Budget Allocation: $20,000

Non-Response Handling:
- Increase recruitment incentives for underrepresented groups and use numerous follow-up reminders.

B.4 Data Validation

To ensure the accuracy and reliability of responses, we will implement several data validation techniques:

Survey Logic Check:
- Recognize and flag responses contradicting one another (e.g., reporting under 18 but registered to vote).
Attention Check:
- Utilize questions to confirm respondents are actively engaged (e.g., "Select 'Confirm' to start questionnaire")
Post-Stratification Weighting:
- Adjusting for over- and under-enumeration and weighting the sample to reflect the demographic of the US population.

Mode and Measurement Errors:
- We mitigate the impact of using mixed modes (online and telephones) by training enumerators and reducing enumerator bias. Misreporting will be reduced through straightforward questions.

B.5 Poll Aggregation Methodology

We will employ a poll-of-polls aggregation method to reduce bias and smooth fluctuation across individual polls.

Weighting Criteria:
Sample Size: larger samples receive more weight to mirror greater reliability.

Recency: More recent polls are given higher weight to capture modern voter sentiment.

Pollster Rating: Polls from highly rated pollsters receive higher weights to reduce the impact of bias.

B.6 Survey Implementation

Google Forms was used to create and implement the survey, allowing for efficient data collection and safe storage. The main section and sample questions are listed below.

Access the survey: [Google Form](#)

Survey Overview
Title: 2024 US Presidential Election Poll
Purpose: To gather public sentiment and predict election outcomes.
Estimated Time: Less than 5 minutes
Confidentiality: All responses are anonymous and used only for research purposes.

1. What is your age?

- Under 18
- 18-24
- 25-34
- 35-44
- 45-54
- 55-64
- 65+

2. What is your gender?

- Male
- Female
- Non-binary / Prefer not to say

3. What is the highest level of education you have completed?

- Less than high school
- High school diploma or GED
- Some college
- Bachelor's degree
- Master's degree or higher

4. Are you registered to vote?

- Yes
- No

5. Who do you intend to vote for in the upcoming presidential election?

- Donald Trump (Republican)
- Kamala Harris (Democrat)
- Other
- Undecided

6. How likely are you to vote in the upcoming election?

- Very Likely
- Somewhat Likely
- Not Likely

B.7 Budget Breakdown

- **Online Recruitment** : $25,000
- **RDD Recruitment**: $30,000
- **Incentives for Participants**: $20,000
- **Data Processing & Validation**: $15,000
- **Miscellaneous Expenses**: $10,000

Total: $100,000

B.8 Conclusion

This survey methodology uses stratified sampling, multi-channel recruitment, and rigorous data validation procedures to ensure accurate forecasting of the 2024 US Presidential Election. We provide a more stable and reliable prediction through poll-polls aggregation, smoothing out fluctuations across polls. This design balances accuracy, inclusivity, and efficiency with a carefully considered $100K budget, ensuring the poll gathers meaningful insights into voter sentiment and behavior.

References

Bloomberg. 2024. "Trump and Harris Tied in Key States." https://www.bloomberg.com/news/features/2024-10-23/new-poll-has-trump-harris-tied-in-key-states-just-12-days-to-election.

News, ABC. 2024a. "ABC News Polling Methodology Standards." https://abcnews.go.com/US/PollVault/abc-news-polling-methodology-standards/story?id=145373.

———. 2024b. "Hurricanes Helene and Milton Impact on 2024 Election." https://abcnews.go.com/538/hurricanes-helene-milton-affect-2024-election/story?id=114783517.

News, CBS. 2024. "Trump-Harris Campaign Promises in the 2024 Election." https://www.cbsnews.com/news/trump-harris-campaign-promises-2024-election/.

Post, The Washington. 2023. "Polling Methods and Standards (2023)." https://www.washingtonpost.com/documents/f409a864-d58c-430a-8adc-ee6a8a344854.pdf?itid=lk_inline_manual_6.

———. 2024. "Presidential Polling Averages Methodology." https://www.washingtonpost.com/elections/2024/06/26/presidential-polling-averages-methodology/.

Press, Associated. 2024. "Election Polls and Forecasts." https://apnews.com/article/election-polls-kamala-harris-donald-trump-b497bd1015d35ed563e085c8db075802.

The Washington Post. 2024. "ABC News/the Washington Post Polling Methodology (2024)." https://www.washingtonpost.com/documents/b4c28200-18a7-412e-a4da-4e8cf9b91fd5.pdf.

Times, The New York. 2024. "Trump-Harris Polls and the Tight 2024 Election." https://www.nytimes.com/2024/10/21/upshot/trump-harris-polls-election.html.

Watch, Human Rights. 2022. "U.s. Supreme Court Overturns Roe v. Wade." https://www.hrw.org/news/2022/06/24/us-supreme-court-topples-roe-v-wade-blow-rights.