# My title*

**My subtitle if needed**

David Qi Haobo Ren Xinrui Xie

November 4, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

## 1 Introduction

## 2 Data

### 2.1 Overview

Josh Pasek have discussed how to the use of Aggregation and Predictive Modeling on polling can increase prediction effectiveness(Pasek 2015). In this article, we will make use of predictive modeling to create a model that aim to predict the support rate of Kamala Harris in the 2024 United States presidential election using a linear regression model trained form past polling results and Dow Jones Industrial Average.

### 2.2 Measurement

### 2.3 Outcome variables

### 2.4 Predictor variables

## 3 Model

We will use a linear model to predict the general polling result of Harris.

*Code and data are available at: https://github.com/HaoboRrrr/USA_Election

### 3.1 Model set-up

For each poll, we assumed they are being sampled similarly.
Let the support rate for Harris at day $t$ be denoted by $R_t$. And we denote the Dow Jones Industrial Average at day $t$ by $D_t$. Then we will use linear regression model:

$$R_t = D_{t-60} + D_{t-90} + \epsilon_t$$

Where $\epsilon_t$ is an independent, normally distributed error term.

### 3.2 Model justification

We expect a positive relationship between the support rate of Harris and Dow Jones Industrial Average 60 and 90 days before. Since people tend to support the party in power when the economy is better, where the condiction of the economy is reflected by $D_t$.

TODO: find paper to support this. finishing reading https://dspacemainprd01.lib.uwaterloo.ca/server/api/core/
72b7-44e5-8fdc-8a5cf0b77b24/content

## 4 Results

## 5 Discussion

### 5.1 First discussion point

### 5.2 Second discussion point

### 5.3 Third discussion point

### 5.4 Weaknesses and next steps

# Appendix

## A Additional data details

## B Model details

### B.1 Posterior predictive check

### B.2 Diagnostics

# C Appendix 1: Polling methodology

## C.1 Overview of Morning Consult

Morning Consult is a globally renowned enterprise technology company that specializes in providing intelligent data to support leaders' decision-making processes. Morning Consult has established itself as a key player in the field of artificial intelligence and decision-making consulting. The company offers a suite of services that include market research, brand analysis, consumer behavior studies, and more. So that the company can optimize decision-making processes and provide clients with a competitive edge(Consult 2024b)

## C.2 Features of the Sample

- Population: The survey aims at all the American citizens who is eligible to vote in the 2024 presidential election in all the parties.
- Sampling Frame: A list of the American citizen's email address and telephone number.
- Sample Size: The polling samples showed a time range from 2011 to 2024. During each election period, the sample size varies a lot. The sample size ranges from 111 to 78247 across the voting period.

## C.3 Sample Recruitment

The company used many sampling methods, such as online panel, live phone call, text, emails and a mix of these approaches. During the recruitment, the data was collected by contacting landlines via Interactive Voice Response. The sampling teams made phone calls and sent text message to randomly selected phone numbers, and collect their responses. The online team collect votes from the active users in the websites and other online panel of voters provided by research marketplace such as CINT. However, some of the polling results come from a survey from long ago, which means that it might be incorrect or biased(Consult 2024a).

## C.4 Sample Approach

Both probability and non-probability sampling methods are involved in the sample collection. For the people surveyed via phone calls and text messages, the main sampling method is Simple Random Sample (SRS). While for those sampled via online panel, the main approach is convenience sample.

The Morning Consult company used so-called "tracking polls" (Consult 2024a), interviews from a certain period are incorporated into future polls, getting reweighted with different samples until they are too old and dropped from the analysis.

## C.5 Strengths and Weakness

Strengths: Since both probability and non-probability sampling methods are used, the pollster can cover a wider range of people surveyed. This means that the sample result can better represent the whole population. Also, the reweighting mechanism can guarantee different methods can have different influence on the final result.

Weakness: According to the data collected, majority of the pollster are done by the convenience sample, which is a non-probability sampling. Thus the actual surveyed population may be biased from the target population, and the result may be less representative. Additionally, the survey showed some overlapping time period, which means that some opinions of the interviewers may be counted more than once.

## C.6 Non-response Handling

Just like all the surveys, non-response bias is the biggest challenge met during the pollster. This is also the greatest factor that may affect the effectiveness of the survey. In all these sampling recruitment and sampling approaches, a portion of the respondents would fail to answer the survey questions as required for various reasons (such as refusing to participate, being unreachable, or being unable to understand the survey questions).

Based on the online research, Morning Consult company usually adopt these measures to minimize the effect of the non-response bias, such as increasing the survey response rate, following up with non-respondents, and using a broader sample for the survey. Additionally, by comparing and analyzing the characteristics of respondents and non-respondents and supplementing the data with other relevant sources, researchers can further assess the influence of non-response bias on the study results.

According to the collected data, we can also see that all the sample recruitment methods request the use of the cell phone. This automatically left out the people who do not use cell phones or who do not use Internet, such as some old people. But this type of non-response bias is still unsolved.

## C.7 Comments of the Questionnaire

The good thing about the questionnaires is that most of the questions are direct and easy to understand. So that this can avoid the bias due to misunderstanding of the questions. Secondly, the questionnaire is short, which makes more people willing to answer without causing waste of time.

On the other hand, the questionnaire requests the interviewers to identify their party, which may cause people not willing to vote for the candidates from the other party, even though they actually want to. This will also cause a slight bias in the response.

# D Appendix 2: Ideal methodology and survey

## D.1 Objective

The survey aims to capture voter intention and mood across important demographic and geographic divisions in order to predict the result of the US presidential election. This survey's methodology employs stratify sampling, effective respondent recruitment strategy and data validation techniques to ensure the accuracy of prediction.

## D.2 Sampling Approach

This survey employs Stratified Random Sampling, which involves divide the target population(In this case, the people who are eligible to vote in US) into subgroups(strata) based on their demographics characteristic(Stantcheva, n.d.). To simulate the national electorate, stratify the sample according to age, gender, race/ethnicity, education level, urban/rural domicile, and region. Also refer to the US Census data to determine the proportion of each strata, for example: if the census data shows there are 20% US citizen are between the age of 20 to 30, then there should be 20% of respondent in this age range. We aim for a starting sample of five thousand people since there were budget constraint.

## D.3 Respondent Recruitment

Partner with an online panel provider to find respondents that match each stratum's specifications(Stantcheva, n.d.). Specify quotas to ensure that, for instance, you wish to have 18% of respondents from the Midwest or 25% aging between 30 - 40. Then randomly select respondent within each strata. To promote greater response rates and retention, provide a small monetary reward, such as $5 for each completed survey. To improve inclusivity and take into consideration differences in internet availability, use both phone and online surveys.

## D.4 Data Validation

To keep distinct replies between responses, make sure respondents don't participate more than once each wave, and highlight straight-lining, inconsistent replies, and other low-effort answers for data quality purposes(Horn et al. 1997).

## D.5 Budget Expenditure

- $50k - $60k for panel provider
- $25k for incentives
- $10k for Telephone survey services
- $5k for data analysis

## D.6 Survey Structure

### D.6.1 Introduction:

First of all, thank you for taking the time to participate in this survey.

This survey aims to precisely capture voter intention and mood across important demographic and geographic divisions in order to predict the result of the US presidential election.

Please note:

- Your responses will remain confidential.
- Please answer the question honestly.
- Complete the survey will receive a reward for $5 - $10.

If you have concerns or questions, reach out to haobo.ren@mail.utoronto.ca (Haobo Ren)

### D.6.2 Screener Section: Screening out people who are eligible to vote in each state.

- Are you currently registered to vote?
  - Yes
  - No
- Which state are you currently registered to vote in?
  - (Drop down box)

#### D.6.3 Presidential Ballot Section: Ask participants whether they would vote, who they would vote for, and how confident they are about that.

- Do you plan to vote in the upcoming Presidential election?
  - Yes, I plan to vote in person on election day
  - Yes, I plan to vote in person early
  - Yes, I plan to vote by mail
  - No, I do not plan to vote
- If the election were held today, who would you vote for?
  - Democrat Kamala Harris
  - Republican Donald Trump
  - A third party / Independent candidate
  - Unsure
- How certain are you about your choice
  - Scale 1 to 5

#### D.6.4 Demographics Section: This section collects detailed demographic characteristics of the participants.

- Your gender identification
  - Female
  - Male
  - Other
- Age Range
  - 18 to 24
  - 25 to 29
  - 30 to 39
  - 40 to 49
  - 50 to 59
  - 60 or older
- Which of the following best describes your race or ethnicity
  - White or Caucasian
  - Black or African American
  - Hispanic
  - Asian or Pacific Islander
  - Native American
  - Other

- How would you describe your educational level
  - High School or less
  - College Graduate
  - Post-graduate degree
  - Unsure
- Employment Status
  - Full time
  - Part time
  - Unemployed
- What's your estimated household income?
  - Under $30,000
  - $30k - $49,999
  - $50k - $99,999
  - $100k - $199,999
  - $200k - $249,999
  - $250k or more
  - Unsure
- Do you consider yourself to be lesbian, gay, bisexual, transgender, queer or questioning?
  - Yes
  - No
  - Unsure
  - Prefer not to answer
- Would you best describe the neighborhood or area you live in as
  - Urban
  - Suburban
  - Rural
  - Unsure
- What is your religious background
  - Religious
  - Spiritual
  - Not religious or spiritual

Link to the survey: https://forms.gle/spJjASENRQ6nKK1E6

## References


Consult, Morning. 2024a. "2024 Election State-Level Polls." https://pro.morningconsult.com/trackers/2024-election-state-polls.

———. 2024b. "Imagine Having Foresight to Back Every Decision. Welcome to Decision Intelligence." https://morningconsult.com/?gclid=CjwKCAiAgbiQBh.

Horn, Werner, Silvia Miksch, Gerhilde Egghart, Christian Popow, and Franz Paky. 1997. "Effective Data Validation of High-Frequency Data: Time-Point-, Time-Interval-, and Trend-Based Methods." *Computers in Biology and Medicine* 27 (5): 389–409.

Pasek, Josh. 2015. "Predicting Elections:considering Tools to Pool the Polls." *Public Opinion Quarterly* 79 (2): 594–619. https://doi.org/10.1093/poq/nfu060.

Stantcheva, Stefanie. n.d. "Online Appendix for 'How to Run Surveys: A Guide to Creating Your Own Identifying Variation and Revealing the Invisible'."