Exploring Candidate-Specific Polling Questions to Improve Polling Accuracy on US Elections

Jake Snyder, Jie Lu, Kathy Lin, Tin Oreskovic
David Rothschild, Tobias Konitzer
Andrew Gelman

Topic Importance and Data Quality

Nationwide surveys have been widely used to estimate public opinion, evaluating public policies and, ultimately, predicting election outcomes. However, there are several risks associated with traditional surveying methodologies that limit the interpretation of results from election forecasts. If these risks aren’t properly mitigated, the survey results can mislead the public and elevate the existing levels of public mistrust in surveys.

Problem Statement:
1. Is there an efficient and effective way of fully utilizing online poll data to predict two-party voter share in these elections?
2. Is poll data with specific candidate names listed worth investing time/money to collect?

Model Evaluation and Findings

Besides the below metrics, there are noteworthy specific findings about the two surveys: for instance, the behavior prediction for independent, married, white males, >54 y.o, without a bachelor’s degree is the most different across the surveys.

Conclusion:

The estimates using generic ballot polling questions perform marginally better than the estimates using specific candidate names. Since it is more expensive and time consuming to use specific candidate names in polling, we recommend continuing to use generic polling questions, as the investment does not seem to be worthwhile. However, further analysis is needed to determine whether this result holds within specific sub-demographics, such as age or education.

Acknowledgments

We would like to thank Dr. Tobias Konitzer for modeling advice and Dr. David Rothschild and Prof. Andrew Gelman for mentoring us on this project.

References


Tobias Konitzer, Sam Corbett-Davies and David Rothschild. “Non-Representative Surveys: Modes, Dynamics, Party, and Likely Voter Space”