# Applying Swing Vote Analysis with Multi-level Regression and Poststratification (MRP) to Improve U.S. Pre-Election Polling Estimates

Kaan Cem Ketenci
ketenci@umich.edu
Michigan Program in Survey and Data Science
UNIVERSITY OF MICHIGAN

**AAPOR Conference 2022**

May 12, 2022

# Motivation

- **"The Mythical Swing Voter"** by Gelman et al., 2016.
- Most U.S. pre-election polling estimates indicate somewhat large fluctuations during the months leading to the elections.
  → Unrealistically high shifts in votes considering the partisan polarization in the U.S. politics, especially in presidential elections.
- Campaign events such as debates and national conventions of parties lead to artificially large shifts in estimated vote shares.
- Analysis by Gelman et al. using the pre-election polls for the U.S. 2012 presidential elections shows that campaign events were especially correlated with survey participation rates rather than actual large changes in vote preferences.
  → **Actual vote swing is small.**

# A Successful Example: UK Exit Polling Methodology

- "Exit polling in a cold climate: the BBC–ITV experience in Britain in 2005" by Curtice & Firth, 2008: Analyzes the improvements made in the U.K. exit polling methodology since 2001 U.K. general elections.
- The ***change*** in each party's vote shares between consecutive elections varies much less than the overall variation of parties' actual vote shares.

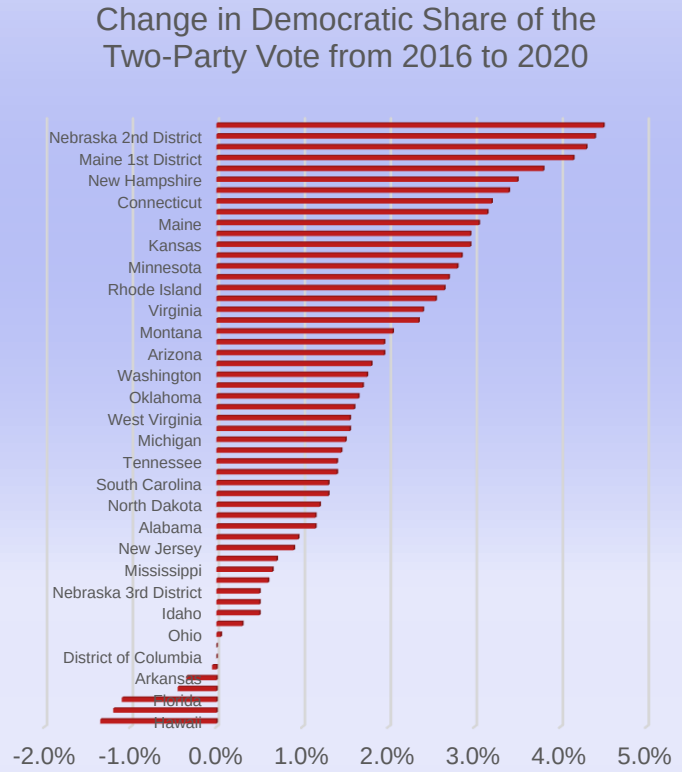
| Constituency-level Std: | Labour | Conservative | Liberal Democrat |
|---|---|---|---|
| Std. of vote share % | 15.1 | 14.0 | 10.4 |
| Std. of absolute change in vote share % (2001 to 2005) | 4.0 | 3.0 | 4.4 |

- Even a non-representative polling station in terms of party vote shares may be representative in terms of vote swings in consecutive elections.

| | 1992 | 1997 | 2001 | 2005 | 2010 | 2015 | 2017 |
|---|---|---|---|---|---|---|---|
| Difference between predicted and actual # seats | 62 | 29 | 6 | 0 | 1 | 15 | 4 |

# Swing Vote Analysis in the U.S. Elections

- Focusing on the ***change*** in party vote shares, i.e., swing vote analysis is especially useful when there is a **polarized electorate** so that there is relatively small swing in voters' behavior in consecutive elections.
- Swing vote analysis can be applied in settings of any number of parties/candidates, however it performs best when the vote shares are dominated by **two parties / candidates** as in the U.S. elections.



# Data and Methods

- Longitudinal panel datasets by the Pew Research Center's American Trends Panel and by ANES (American National Election Studies) covering pre-election and post-election surveys of 2012, 2016, 2018 and 2020 U.S. presidential and midterm elections, supplemented by voter validation data when available.
- To estimate the Democratic candidates' share of the two-party vote:
  - Code respondent as +1 for Republican-to-Democrat swing
  - Code respondent as -1 for Democrat-to-Republican swing
  - Code respondent as 0 for intending to vote to the same party as before
  - Categorize 3rd party voters together with the non-voters.
- Conduct swing vote analysis separately for validated 2016 voters and for validated 2016 non-voters. This reduces the variance in estimates as voter-to-nonvoter and nonvoter-to-voter swing is much larger than the swing between the two parties.

# Swing Vote Analysis Reduces Sampling Variance

- To improve accuracy in estimates and reduce measurement error due to false recall of respondents' votes in the past election, use post-election survey responses obtained shortly after the past election.
  i.e., 2016 post-election survey data is used jointly with 2020 pre-election survey data.
- Voter validation data for the past election cycle is especially necessary as false reports of voting in the past election is very common.

| Weighted Sample | 2012 to 2016 | 2016 to 2018 | 2016 to 2020 |
|---|---|---|---|
| Dem-to-Rep swing among past Democratic voters | 10.5% | 3.6% | 3.7% |
| Rep-to-Dem swing among past Republican voters | 5.9% | 4.5% | 7.5% |

- Swing vote analysis reduces sampling variance among past voters by a factor of 3.1 in 2016, by 6.1 in 2018, and by 4.5 in 2020 elections.
  Larger variance remains among likely voters who have not voted in the past elections.

# Demographic Shift in Consecutive Elections

- Demographic changes during the years between the consecutive elections must also be accounted for in addition to the estimated vote swing. This is especially necessary in recent U.S. elections since young voters who become eligible to vote for the first time have been Democratic voters at much higher rates than the general electorate.
- - Eligible voters who turned 18 for the first time before an election
  - Estimated deaths by age & party ID between the two consecutive elections (Life table statictics are used to get estimates of deaths by age.)
  - Immigrants who became eligible to vote for the first time
- e.g. 18-21-year-old voters in 2020 elections were ineligible to vote in 2016. This group is not part of the swing vote analysis. Additional vote margins for Democrats are calculated using estimated turnout among 18-21-year-olds.

# Using MRP Methods

- MRP (Multilevel Regression and Poststratification) is used together with the swing vote analysis.
- This allows swing vote estimates to be obtained for each of the multi-dimensional demographic category. The conclusions can be informative similar to the exit poll results investigating voting outcome by demographic groups.
- Poststratification cells include age group, gender, race/ethnicity, education, income group, and vote in the past elections. Benchmark proportions of the electorate population are obtained using exit poll data from the past elections corrected using Pew Research Center's estimates by voter validation, and applying raking procedures to get estimates for each cell.

# Caution with Exit Poll Estimates

- Some of the most widely cited exit polls have substantial biases in their demographic estimates of actual voters.
- e.g. 2016 CNN Exit Poll estimated college graduates as 50% of actual voters while the more accurate estimate by Pew Research Center using voter validation data was only 37% college graduates among all voters.
- Similarly voters aged 18-29 was estimated to be 19% of voters in the 2016 CNN Exit Poll while the voter-validated estimate was only 13% by Pew Research Center.
- Exit poll reweighting process generates final weights using actual known election results, and are therefore guaranteed to give correct total share of the votes by candidates for the whole nation. However, exit polls still suffer from large biases due to differential rates of non-response within voters of the same demographic group.

# Findings for 2016, 2018 and 2020 Elections

- Final estimates of vote shares are obtained by final weights generated through swing vote analysis, MRP, demographic shift estimates between consecutive elections, and likely voter models.
- Estimated two-party vote shares vs actual results:

| Two-party Vote Shares | Predicted Democratic Vote Share | Actual Democratic Vote Share |
|---|---|---|
| 2016 Presidential Elections | 51.5% | 51.1% |
| 2018 Midterm House Elections | 54.1% | 54.4% |
| 2020 Presidential Elections | 53.7% | 52.2% |

- Predicted votes in 2016 and 2018 elections are close to the actual election outcome. 2020 estimates are somewhat off (7.4% national popular vote margin predicted for Biden vs the true outcome of 4.5% vote margin). However, it is an improvement over alternative predictions of close to 9% vote margin for Biden.

# Further Research for Improvement

- 2020 pre-election polling estimates likely suffered substantially by differential rates of non-response both between voters of different parties and also within voters of the same party. Swing vote analysis eliminates most of the bias due to differential rates of survey participation among different parties. However, there are still large unaccounted factors influencing survey participation within voters of the same party even after controlling for main demographic variables.
- Need key variables to reduce biases due to within-party differential rates of non-response in surveys. Some candidate variables:
  - Volunteering character
  - Actual voters of primary elections
  - First-time voters
  - Trust in authority including trust in government, health authorities, etc.
  - Trust & anger towards congress: e.g. Responses of "Never" to the question "How often can you trust the federal government in Washington to do what is right?"
- One particular aim is to detect the differences in opinion among more traditional voters vs ‘anti-establishment’ voters.