Part Two

Oscar Garcia, Shuo Wang

6/28/2022

Contents

1 Importance and Context

Americans voted in record numbers in the 2020 election [1]. Despite this, the US still lags behind other developed nations in electoral participation: 24th out of 35 OECD members, 30th in 2016 [2] At the same time, in the case of the presidential election, the winner has been decided by close margins in a few key states:in 2020, 37 electoral votes were decided by margins of victory of less than 1% (GA, AZ, WI); all of these went to Biden. This highlights the importance of mobilizing voters to ensure turnout is maximized so that a candidate has the best chance to prevail.

We are a team of political consultants and have been mandated by the Democratic Party to advise them on relative difficulty to vote in their potential voters, as compared to Republican voters. The goal is not to be at a disadvantage with respect to the Republican candidates from the start. As a first step, this analysis aims to address the following research question:

Did Democratic voters or Republican voters experience more difficulty voting in the 2020 election?

If we conclude that potential Democratic voters experience more difficulty to vote, the party and its elected representatives should push for further understanding of the causes and for legislative reforms, political campaigns to ease those difficulties or both.

2 Data and Methodology

We use data from the 2020 American National Election Studies (ANES). It includes re-interviews with 2016 respondents as well as fresh survey to addresses randomly drawn from the USPS, and post-election surveys with respondents from the General Social Survey (GSS). This dataset includes 8280 individuals.

Who is a "voter"? since greater difficulty would entail not being to vote in the most extreme cases, we want to include respondents who *intended* to vote but could not do so. The data includes a self-reported level of voting difficulty, but the universe of that question included only those who did vote. To have a broader picture of who is a potential voter, we define "voter" as follows:

- Respondents who stated they intended to register to vote (variable V201019 in ANES), intended to to vote for President (V201032), intended to vote for US House (V201041), intended to vote for US Senate (V201051), intended to vote for Governor (V201069).

We remove respondents who are not "voters according to the definitions above. This leaves 7338 observations.

To operationalize the concept of who is a "Republican" or a "Democrat", we leverage V201231x in ANES, a summary variable that includes 7 valid responses in an ordinal scale from 1. Strong Democrat to 7. Strong Republican, with 4. being Independent. Research shows that self-described "Lean Republican" or "Lean Democrats" are in fact partisans [3], so we map the responses as follows:

- V201231x responses 1 to 3 are mapped to "Democrat".
- V201231x responses 5 to 7 are mapped to "Republican".

We discard the remaining "Independent" respondents for the purposes of this study, resulting in 7316 observations.

To operationalize the concept of "difficulty voting", we leverage two groups of variables: one for respondents who managed to vote, and one for those who did not finally vote. The new Difficulty rating we build has the following elements:

- Answers to Variable 202119, the question "How difficult was it to vote": ranging from 1 (not difficult at all) to 5 (Extremely difficult). We take these valid responses directly, as they apply to respondents who voted in 2020.

[1] Pew Research Center. "Turnout soared in 2020 as nearly two-thirds of eligible U.S. voters cast ballots for president." (2021).
[2] Pew Research Center. "In past elections, U.S. trailed most developed countries in voter turnout." (2020).
[3] "Measuring party support: Leaners are not independents", Petrocik (2009).

- For voters who did not vote, we create an additional level of difficulty: 6. We selected reasons with the criteria of whether the Democratic Party can take an action that would help correct that difficulty. We assign evel 6 to respondents who gave a positive response to these variables:
  - V202114a: did not meet registration deadlines; V202114b: did not know where or how to register; V202114e: did not have required id; V202114f: not interested in the election; V202114g: my vote would not make a difference; V202114i: difficulty with English; V202114k: other reason.

Another source of information for Difficulty is V202123, main reason respondent did not vote. We map some of the responses to our new level 6: - 1. I forgot; 2. I'm not interested; 6. I did not have correct form of id; 11. The line at the polls was too long; 13. I requested but did not receive an absentee ballot; 14. I did not know where to vote; 15.I did not feel that I knew enough about the choices

We exclude voters without a positive Difficulty rating, resulting in 5616 observations.

As alternative to the above, we considered defining "voters" as only those respondents who stated they voted, and directly taking their self-reported difficulty rating. The issue with this approach was that we were not including people who intended to vote but found difficulties that prevented them from doing so, as explained earlier.

Table 1: Voting Difficulty by Party allegiance (count)

	1	2	3	4	5	6	Sum
Democratic	2541	229	85	19	12	83	2969
Republican	2319	146	53	23	18	88	2647
Sum	4860	375	138	42	30	171	5616

As we report in Table 1, we have slightly more Democratic respondents than Republicans (2969 vs 2647). We see that the majority of voters reported having no difficulty at all, which was to be expected. To have a first approximation on which group seemed to have higher difficulty in voting, we ran the proportions within each political group in Table 2. This shows that the proportion of Democratic voters with difficulty ratings 2 and 3 is higher than Republicans for the same ratings (8% and 3% versus 6% and 2%, respectively). For any rating that implies some difficulty (2 to 6), the proportion of Democratic voters in those ratings is equal or higher than Republicans except for rating 5. This rating is the one with the smallest absolute number of voters across ratings as shown in Table 1.

Table 2: Voting Difficulty by Party allegiance

	1	2	3	4	5	6
Democratic	0.86	0.08	0.03	0.01	0.00	0.03
Republican	0.88	0.06	0.02	0.01	0.01	0.03

To answer our research question, first we need to select the most appropriate test. The data are unpaired (each respondent in our analysis data is either Democratic or Republican and has a single difficulty rating associated). The Difficulty rating is not metric: we cannot define a "distance" that would guarantee that differences between ratings 1 and 2 mean the same thing as say, distance between 5 and 6. The scale is ordinal. Therefore, we need a non-parametric test. We could have chosen to convert this scale to binary (for example saying Difficulty ratings above 2 or 3 means "having difficulty" with a value of 1 in the new binary variable, rest are 0). The resulting binary scale would be metric, and we could use a t-test. We would however lose information contained in difficulty scale, which we feel is important for a political party to assess further action. We decide to use a Wilcoxon rank-sum test. Since the Difficulty scale is only ordinal, we will use Hypothesis of Comparisons Version.

The null hypothesis of our Wilcoxon rank-sum test can be phrased as follows:

Null Hypothesis: *The probability that a Democratic voter had more difficulty voting in 2020 than a Republican voter is the same as the probability that a Republican voter had more difficulty voting in 2020 than a Democratic voter*

Assumptions for Wilcoxon rank-sum test with Hypothesis of Comparisons: 1) Ordinal scale in the outcome variable. This holds because Difficulty rating is ordinal. 2) IID data. Each Democratic sample is drawn from the same distribution, each Republican sample is drawn from the same distribution, and all are mutually independent. We believe this assumption is plausibly true given that respondents were drawn from the USPS address lists, and household members were taken to a screening instrument to randomly select one person from the ones living at the address, thus reducing household effect.

3 Results

```
# ?wilcox.test
wilcox.test(anes_lab1[anes_lab1$Party == 'Democratic',]$Difficulty,
            anes_lab1[anes_lab1$Party == 'Republican',]$Difficulty, alternative = "greater")
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  anes_lab1[anes_lab1$Party == "Democratic", ]$Difficulty and anes_lab1[anes_lab1$Party == "Repu
## W = 4000544, p-value = 0.02406
## alternative hypothesis: true location shift is greater than 0
```

The test is significant and provides evidence that we can reject the null hypothesis. Therefore we accept the alternative hypothesis, which is that the probability that Democratic voters had a greater difficulty in voting than Republicans is higher than the probability that the Republicans had more difficulty (p=0.024).

To study if this statistically significant result is of *practical significance*, we can use common language effect size (valid for ordinal data according to [4]: there are $7,858,943$ pairs (Democrat,Republican) in our data. If we divide the W value $4,000,544$ by that number, we get 0.509: in 50.9% of pairs, Democratic voters have more difficulty voting than Republican voters. This effect size is considered large enough to be meaningful in real life. Further, Table 2 shows that the percentage of Democratic voters that found some difficulty (ratings 2 and above) add up to 15% of voters, while the figure is 13% for Republicans. This difference of 2 percentage points is meaningful for the Democratic Party, because swing states have been decided, as seen in the Introduction section, by margins even smaller than 1 percentage point.

Out study does have some limitations. We cannot attribute causality between party allegiance and difficulty voting. We can only measure association between self-reported difficulty and allegiance to the two parties.

4 Discussion

The study found evidence that Democratic voters had more difficulty voting in 2020 than Republican voters. The effect appears statistically significant, and the difference of a few percentage points of Democratic voters saying they experienced difficulty is or practical significance because differences of that order of magnitude have decided elections recently. Even a voter that experienced difficulty and finally voted could decide to give up the next time, so this is an issue the Party should address.

The Democratic Party could take multiple actions from this result: increase efforts to inform potential voters about upcoming elections, improve awareness of candidates, push legislative reforms that make it easier to vote. Future work may address specific sources of difficulty or socioeconomic segmentation to improve targeting of the measures suggested to reduce voting difficulty.

[4] "A Critique and Improvement of the CL Common Language Effect Size Statistics of McGraw and Wong", Vargha, András; Delaney, Harold D. (2000).