

Social pressure and voter turnout: Evidence from a large-scale field experiment (Gerber, Alan S and Green, Donald P and Larimer, Christopher W, 2008)

Hyo Won Shin*

May 17, 2017

1 Introduction

In Green, Gerber and Laimer's 2008 article, "Social pressure and voter turnout: Evidence from a large-scale field experiment," they look into the effect social pressure has on voter turnout. They conducted an experiment designed to prime voters to think about civic duty while applying different amounts of social pressure to induce them to act according to this norm. This experiment took place in the context of an actual election. Prior to the August 2006 primary election in Michigan, around 80,000 households were sent out one of the four mailings encouraging them to vote. One of the treated groups received mailing that reminded them that voting is a civic duty. The second treated group were told that researchers would be studying their turnout records based on public record. The third group received mailing, which indicated the record of turnout among those in the household. The last group were sent mailing that displayed both the household's voter turnout and their neighbor's turnout. The last two groups were also informed that a follow-up mailing after the election would report to the household or their neighborhood the subject's turnout in the upcoming election. Results showed that social pressure does indeed increase voter turnout. The influence of mailing was greatest only when social pressure was exerted in the message. Exposing the subject's voting records to their household or neighbor had a significant effect than conventional pieces of partisan or nonpartisan direct mail. It also showed that mailing infused with social pressure was as effective as door-to-door canvassers. (Gerber, Green, and Larimer 2008)

2 Overview of the article

Gerber, Green and Laimer's field experiment consisted of one control and four treatment groups. Before the individuals were randomly assigned to treatment and control groups, they eliminated candidates that were unfit to be included in the experiment. Using the "Qualified Voter File", the official state voter list, they got rid of those with missing voter history, incorrect ZIP codes, typographical errors in names or addresses, or multiple listing on the QVF. From the remaining number of candidates, they also got rid of people who could not be assigned a valid 9-digit ZIP code since this is crucial when injecting the mail treatment. They also eliminated people who lived on blocks where more than 10% of the addresses included apartment numbers. They did this because since their level of observation was neighborhood with single-family homes and for apartments, they would need to look at the physical layout of each apartment complexes in order to choose the

*hwshin2@illinois.edu, University of Illinois at Urbana-Champaign

appropriate neighborhoods. They also got rid of people who lived on streets with fewer than four address (or fewer than 10 voters) out of concern the mail would not be delivered in time for the primary vote. They then eliminated households with the following characteristics: all members of the household with a 60% probability of voting by absentee ballot if they voted or all household members had a greater than 60% chance of choosing the Democratic rather than the Republican primary. The reason for removing absentee voters was because they thought these voters would have already made up their mind about voting prior to receiving the experimental mailing. They removed those that were overwhelmingly likely to favor the Democratic primary were eliminated because they thought that, given the lack of contested primaries, these people would ignore pre-election mailing. From those left, they removed those that lived in a route where fewer than 25 households remained since the experiment depended on using a carrier-route-presort standard mail process. In order for a neighbor to qualify for this, it requires that at least 10 pieces to be mailed within each carrier route, which may not have been available after the control group were removed. Finally, the authors removed those that abstained in the 2004 general elections on grounds that those not voting in this very high turnout election were likely to be “deadwood” meaning they were thought to have moved, died or registered under more than one name.

After the sample was carefully selected, they send households assigned treatment groups one mailing 11 days prior to the primary election. They were randomly assigned to either one control or one of the four treatment groups. The authors sorted the 180,002 households into the order required by the USPS for “ECRLOT” eligibility (by ZIP, carrier route; then the order the carrier delivers the mail). The households were then divided in 10,000 cells of 18 households each, with each cell consisting of households 1-18, 19-36 and so forth of the sorted file. After sorting, each cell consisted entirely of either one or two carrier routes. Then a random number was generated and the entire 180,002 records were sorted by cell number and the random number. The purpose of doing this was to keep the cells together but just in a random order. Using the randomly sorted copy of the file, the records were then assigned treatments 1/1/2/2/3/3/4/4/c/c/c/c/c/c/c/c/c/c. Here the c indicates control group. The records were then resorted into carrier route order.

The authors did a good job of balancing the covariates between control and treatment groups. Each of the groups had a mean household size of 1.91, and 84% voter turnout rate for the 2002 general elections, 87% turnout for 2000 general elections, 42% turnout for 2004 primary elections, 41% turnout for 2002 primary elections, and 26% turnout for 2000 primary elections. The gender ratio between male and females for each group was 50:50, and the mean age was around 52 years old. Each treatment group consisted approximately 20,000 households with 99,999 households in the control group.

As aforementioned each of the treatment groups received mail that exerted different amounts of social pressure to vote. Results showed that as more social pressure was exerted, the percentage of vote turnout increased. For the first treatment (and the weakest) of reminding citizens voting is a civic duty, the vote turnout increased by 1.8%. The group that received the second treatment, they were sent mail that informed them they were being studied. This group, the vote turnout increased by 2.6%. The third treatment group received mail that informed them their voting behavior was public information and listed the recent voting record of each registered voter in the household. The voter turnout increased by 4.9%. Lastly, the neighborhood treatment group received mail informing them their voting information was public information and listed not only household’s voting records but also that of their neighbors. The voter turnout increased by 8.1%.

The authors then corrected for robust cluster standard errors since the analysis before ignored the issue of sampling variability. They worried that there was a possibility that individuals within each

household shared unobserved characteristic that could have accounted for errors when treatment was injected. Therefore, they took intrahousehold correlation into account and ran linear regression under three different models. The point of doing this was to minimize disturbance variance and improve precision of treatment estimates. The first model regressed voter turnout for individual on dummy variables marking each of the four treatments. The second model then added to this model by including fixed effects on geographic clusters within which randomization occurred. They did this because it has the potential of eliminating any imbalances within each geographic clusters, thereby improving the precision of the estimates. Lastly, they controlled for voting in five recent elections. This again was to minimize disturbance variance and improve precision of the treatment estimates. The results were very robust, with hardly any movement even to the third decimal place. The effects of each treatment were the same as above and each were significant at $p < .0001$.

They also checked to see if social pressure interacted with feelings of civic duty. The logic behind was that those with a higher motivation to vote to start off with would be more influenced by social pressure than those without motivation in the first place. They, therefore, divided the samples according to the number of votes cast in five prior elections and then further divided these subsamples into number of voters in each household because household size and past voting behavior are correlated. They then conducted a series of logistic regression and examined the treatment effect across subgroups. The results showed that treatment effect underlying voting propensities were more or less constant, regardless of whether the subgroup voted frequently or rarely. They concluded that there was no interaction between social pressure and one's sense of civic duty.

Replication code for the article, which the authors provided on <http://isps.yale.edu/research/data/d001>, show the same results as those indicated in the article. It is also in the code appendix.

3 Analysis

3.1 Bias

Here we want to check for change in the voter's behavior (whether they vote or not) when they receive the treatment of social pressure to vote versus when they are not reminded to do so at all. In this article, the authors use ordinary least squares, or `ols()` to produce estimates and the coefficient produced show that as social pressure increases, voter turnout increases also. The neighbor treatment produces the greatest treatment effect of 0.081, among the four treatments. This can be interpreted as, the pressure of your neighbors knowing your voting behavior increases the probability of voting by 8.1% compared to the control group. But can we take this value as the true effect?

In order to test this, we have to test for bias. An estimator is said to be unbiased when the estimator produces the "truth". This means that the mean of the sampling distribution of the estimator is equal to the effect produced from the population, or the true effect. In order to test to see whether the estimator is biased or not, we need to create a world in which we know the true effect of the treatment, which is the true effect of social pressure on voter outcome. Our explanatory variable is consisted into 5 parts: 1 control and 4 treatments. The first treatment, civic duty, is the lowest intensity of social pressure, which is a simple reminder to vote because it is a civic duty. Second treatment is of a higher intensity, which is the Hawthorne treatment. This reminds citizens they are being studied. The third treatment, "self", is of higher intensity reminding citizens their voter behavior is of public information. Lastly, the neighborhood treatment reminds citizens their voting records are open to their family and their neighbors.

Since we have four treatments, I will take one treatment to test for bias of the estimator. I would not

have to test for bias for all of the four treatments, because the estimates were all generated using the `ols()`. Of the four treatments, I chose the neighborhood treatment because that shows the highest treatment effect of 8.1%. This treatment to me was the most interesting one since it returned a very high effect value, one that was even surprising to the authors themselves. I wanted to check whether this effect is really the “truth” or one due to random chance.

The code returns 1000 coefficients all generated using different sample size. As the number of sample size gets bigger, we should witness the coefficients getting closer to the true effect, which is 0.08130991. The coefficients, as predicted, returns coefficients that are smaller and greater than the true value randomly. At first, it seemed as if the coefficients were heading towards the true effect, but it jumped to a higher value again. At other times, it jumped from a high coefficient value to a way lower coefficient value, not getting any where close to the true effect. This proves that `ols()` produces unbiased but inconsistent estimators.

The true effect is 0.08130991, while the mean of 1000 coefficients is 0.08140758. The two coefficients are very close to one another, which shows that the `ols()` produces an unbiased estimator. We also plotted the observations and a regression line that passes through these points. This shows that estimator is unbiased. Also we can see that there aren’t any outliers or observations that might skew the outcome.

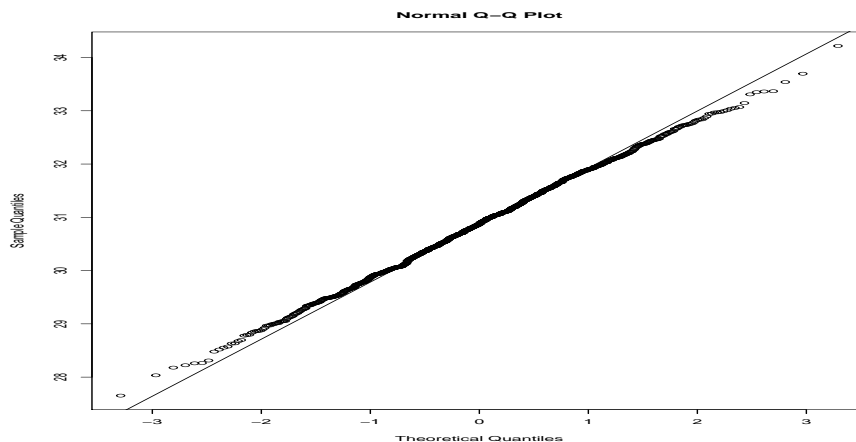


Figure 1: Bias Test qqplot

In the past explorations, we have found that `ols` produced unbiased estimates but did not produce very consistent ones. In order to test whether this is true for this case, we will check for consistency.

3.2 Consistency

Consistency of an estimator indicates that as the sample size gets larger, the estimate gets closer and closer to the true effect. To test for consistency of the estimator, I will do something very similar to what I did when testing for bias. I will first create a world of known effects and `rn` regressions on that world. In the code, that is done within the inner loop. On the outer loop, I will increase the number of samples in each run, starting with 100 and increasing by 400 until it reaches 2000 samples. If the `ols()` were to produce a consistent estimator, the coefficient each loop produces should get closer to the true effect. However, if it does not, meaning if the coefficients seem to bounce back and forth, it would prove that it produces an inconsistent estimator. Let’s see which of the two is the case.

3.3 Type I error

Now we will test to see whether the `ols()` produces an estimator that correctly rejects the null hypothesis when it is true. When it does not, meaning if it rejects a correct null hypothesis, we will face the problem of committing a type 1 error. This is problematic because in this case, we will be accepting an alternative hypothesis when the results are actually produced by chance. We want the case of this error occurring to be as small as possible. Traditionally, researchers set the chances of this happening at 0.05 - this means there is only a 5 in 100 chance that this variation (effect present) is due to chance. We, therefore, want to see whether `ols()` produces an estimator that does not commit serious type 1 errors (more than 5%).

The steps to test for type 1 error are as follows: After breaking up the relationship between the independent and dependent variable by shuffling the outcome variable, we ran an `ols()` regression on that world. The code returned confidence interval and saw whether it included the null effect of 0 or not. Since we have broken up the relationship between the explanatory and outcome variables, the treatment effect should no longer have an effect, returning a 0 outcome. As a result, when running the `ols()` regression in the world of no effect, we should almost always get a confidence interval that contains the null effect of 0. Since we allowed 5% of type 1 errors to occur by chance, we expect the estimator to correctly accept the correct null hypothesis 95% of the time.

I am, however, quite confused about the return values. After running the function 1000 times, the result I get is 0.505. This, I think, is the percentage of correctly accepting the null hypothesis. This means, 50.5% of the time, the estimator correctly accepts the null hypothesis. This means that, on the other hand, the estimator is making type 1 error, or falsely rejecting the correct null hypothesis 49.5% of the time. If this were true, then the results we get are quite contrary to what I expected. This means that half of the time, the estimator is false rejecting the correct null hypothesis. I also ran a function called the “`simsim`”, which returns first 10 rate at which we accepted the correct null hypothesis. All of the 10 values returned were around 50%, which means that the probability of the estimator making type 1 error was also around 50%.

Is producing a large type 1 error rate another characteristic of using an `ols()` regression? Are they known to produce a lot of type 1 errors? If this case were true, then the `ols()` seems to produce a highly unreliable estimator that correctly accepts the correct null hypothesis almost as much as it false rejects them!

3.4 Testing for power

Power is a description of how often our tests are going to produce a significant result, given a certain sample size, confidence interval, and effect size. In this case, the significant result would indicate producing outcomes that are close to the true effect. In the code below, we will look for a power of 80% or more, where a power of 80% means that given a certain treatment effect, 80% of the time, we will be able to confidently reject the false null hypothesis, where the confidence is determined by our alpha cutoff of 5%. Whereas we tested for how often the estimator correctly reject the null hypothesis, here we are looking to see how often it rejects a false hypothesis. In this case, we want our estimator to reject the false null hypothesis 80% of the time and we are looking for the treatment effect that returns us that power rate.

To test for power, we need to determine what the confidence interval/alpha would be. Here, we will use the conventional 0.05 alpha cut off. The sample size here has been set at $n=1000$. we will be looking for a power of at least 80%. Anything that is 80% and above shows that the estimator is almost always rejecting the false null hypothesis and only accepting significant results.

The results show us that when we use the true effect, the estimator rejects false null hypotheses only 48% of the time. This is similar to what we saw when we were testing for type 1 errors. Only half the time, the estimator is rejecting a coefficient that was produced by chance and not the true null hypothesis. If the estimator is a powerful one, it should reject the false null hypothesis more often. We then test to see whether changing the treatment effect gives us a higher power rate. When we increase our treatment effect, we get higher power rates. The minimum detectable effect of 80% power rate was when it was 0.116. When the treatment effect of 0.117, the power rate increased to 0.8153, meaning 81.5% of the time, the estimator was able to correctly reject false null hypotheses.

4 Strengths of the article

One of the biggest strengths of this paper is the number of samples the researchers managed to collect. The field experiments I usually come across would have samples of thousands or at most ten thousands, but this study alone had a total of more than 300,000 samples. This was the number of individuals after the authors carefully weeded out the people that were “unfit” for the study. For example, household members that had a 60% probability of choosing the Democratic primary rather than the Republican primary were removed as injecting treatment into these individuals would not have the same effect as those that do not have a strong preference for either candidates. Also for practical purposes, the researchers removed those that could not be assigned a valid 9-digit ZIP, people who live on blocks where more than 10% of the addresses included apartment numbers, and those that lived on streets with fewer than four addresses. But even after these people were removed, the researchers were still left with a sample of more than 300,000 individuals, of which 190,000 were controls and the rest treated individuals. The larger sample size helps increase accuracy and validity of the study as the greater the number of samples, the closer we are to the population.

Another strength of this paper is the well-balanced randomized assignment of treatment and control groups. As I have already mentioned above, the covariates of each group are very similar to one another. From the average number of household size to turnout history in previous primary and general election, Each group was well balanced compared to one another making it comparable to each other. This is important especially because we are taking the control group as the baseline and comparing how voting behavior differed after injecting treatment into the treatment group. If the control and treatment groups showed stark difference in pre-treatment characteristics, we would have suspicions on whether the difference in outcome was due to the treatment or some other unobserved characteristic in that particular treatment group. Both, randomized assignment with large sample size, together makes sure that unobservable characteristics are likely to be balanced.

Green, Gerber, and Lainer were also smart to test for more than one type of treatment. The effect of social pressure on voter behavior was their topic of interest, and to test this, they see how increasing pressure would change the voter turnout rate. If they just compared between a control group and one treatment group, we would see the effect of social pressure of only one kind of social pressure. The concept of social pressure is such a broad definition that it is hard to capture it through just one measurement. The authors did a good job of testing out various social pressure of different levels. The first treatment, which reminded the people that voting is a civic duty, defined social pressure as being indirect pressure from the state to perform their duty as a member of the society. The second treatment, which informed citizens they are being studied, defined social pressure as that being monitored by some study center, from those who you do not have any acquaintance. The third treatment, which is a relatively stronger level of social pressure, informed citizens that their voting records were shown to the public and to their household members. This treatment defined social pressure as not only pressure from the society of performing one’s civic duty but that of the family members who you

share intimate details of your life. This is stronger social pressure because compared to the former two treatments, family members have the ability to criticize and shame that person for not performing their civic duty. The last treatment is the strongest level of social pressure. This not only involves the society and family members but also that of your neighbors. This puts more pressure on you because not only can you be shamed by family members but also your neighbors. While your family members can shame you within the household, being shamed within your neighborhood involves a bigger geographical and social space. This means that more people may criticize and shame you, and this may last a longer time, which to the person would be more severe compared to shaming within the household.

5 Propose changes in the approach to the statistical inference

I believe that the study over all was a very well done one conducted on a scale that not everyone is capable of doing so, statistically and experimentally. As a result, it was quite hard to come up with areas of improvement for this particular studies. In order to look for areas it could be improved on, I looked at studies that were published after this article was written, which was in 2008, and searched for any other factors the studies considered.

In Sinclair, McConnell and Green's study on spillover effects, they used a multilevel design to consider how communication among subjects not only within households but within neighbors could blur the line between treatment and control conditions. It may be the case that two individuals in different households and different zipcodes may influence one another in the workplace, church, school, etc. There is a network present that Green, Gerber and Laimer have not really looked at closely. It is something the authors should take into account as this kind of unusual message is bound to be the subject of conversation between household members and neighbors. The literature has proven that interpersonal communication is thought to influence voter turnout when the voter receives information about a friend's or neighbor's voting rates. In the multilevel model Sinclair, McConnell and Green use, the potential outcomes for each voter would be a function of whether they themselves were treated, whether a member of their household was treated, and whether others in the neighbor were treated. To see the treatment effect of social pressure, via mailing, increased voter turnout, we would compare those whose zip code neighbor, housemates and the voter themselves were treated to untreated voters that live in an untreated zip code neighbor with untreated housemates using a multilevel model. By accounting for not only household spillover but also that of the neighborhood, we would be able to test for a more precise effect of social pressure via mailing on the voter outcome. In order to conduct this experiment, we would need more information on the zip code of each individuals. (Sinclair, McConnell, and Green 2012)

Another possible weakness of this kind of field experiment is the absence of information after treatment injection and before measuring voter turnout. In this field experiment, citizens in the experiment received mail, of which had a message with different levels of social pressure. The experimenters in this kind of experiment have limited control on how they can conduct the experiment. They have control of randomizing who goes into what group and how the messages are written on the mail for each treatment group. After that, however, the experimenters have to leave it to fate that those individuals assigned to each group get properly treated, meaning they read the message properly and then decide how they will vote. Unfortunately, this is not always the case. There may be those that do not pick up their mail immediately but rather open them after one or two weeks or months after they receive it. This means that treatment was not properly injected into that individuals. They, even though were assigned to treatment, would not have received it since they read it too late or did not see the mail at all. Also some may have been assigned to a control group but since they have friends in

treated households, they would indirectly be treated, even though the experimenters did not intend to treat them. The researchers, therefore, must leave it up to fate and pray that these treatment effects are reflected on the voter outcomes, the only source of information they can use to interpret the treatment effect. I think this problem is one that all researchers who conduct field experiments would suffer from. Since the world we live in is so complex and at time very hard to predict due to unforeseen variables, we can only hope that these unusual circumstances are rare and not the case for most of the samples. I think that Green, Gerber and Laimer tried to handle this problem by collecting a large sample, so that the “normal” cases, where the treatment is properly injected in that particular group, are the majority and the cases where the treatment is not injected make up only a few cases, in which will not be significant enough to distort the treatment effect. In order to strengthen their validity of findings, they can repeat this experiment in the general elections. The treatment may have a different effect on individuals voting in the primaries versus the general elections. The underlying logic behind this is that as the salience of election grows, social pressure on one’s voter behavior will also increase. In the general elections, compared to the primaries, the effect of social pressure on voter behavior may be stronger because the stakes of voting are higher and the costs of voting is lower. The choice of deciding who to vote for general elections is easier because they need less information in order to make their decision, whereas in the primaries, you would need relatively more information on deciding who to voter for. Also there is evidence that some people think primaries are reserved for party members only, and believe they should not vote if they do not strongly identify with that party or if they might not vote for the party’s eventual nominee in the general election (Gerber et al. 2017). In the general elections there will be more people coming out to vote and because of this, the social pressure would be greater in this case. I would suspect that the treatment effect would be higher as the electoral salience increases.

When the authors adjusted for standard errors, they took into account past voting behaviors of each individuals. This is important as those who voted more often may have the propensity to vote again in the future compared to those that seldomly go out to vote. I, however, suspect that not all of these elections would have the same weight in terms of how important it was at the time of the event. In other words, some elections may have been more salient than others. For example, the 2017 presidential elections in Korea was much more salient than the past presidential elections due to the scandal President Park caused. The scandalous events that led to her impeachment and arrest caused many people to take interest in Korean politics, which then led to the record voter turnout. This kind of election was an exception to many other elections in Korea. This one involved a scandalous event that caused even those who were formerly uninterested in Korea politics to take interest and cast their vote. As a result, not every election should be weighed differently in terms of how salient it was during that period of time. By weighing the past elections differently, we will get a more accurate voter propensity, whether this person is someone who always casts their votes for whatever reason or someone who only votes when the election is significantly salient like the 2017 Korean presidential election.

The article used the individual’s voting propensity to see whether one’s sense of civic duty had an influence on the treatment’s effect on voting outcome. To calculate the voting propensity, the authors used voting records in five prior elections and the number of voters in each household. Voting records were include since those who voted in past elections are thought to be more likely to vote in the future. Thus, the authors divided the sample into six subsamples based on five past elections; and they further divided the subsamples according to the number of voters in each household. They did this because household size and past voting behavior are correlated (Gerber, Green, and Larimer 2008). When they tested this by conducting a series of logistic regression and examined the treatment effects across subgroups, they found that the treatment effect on underlying voting propensities were more or less

constant regardless of whether the target group votes often or rarely. They, therefore, found that there were no appreciable interactions between social pressure and one's sense of civic duty.

6 Implement proposed change

In this article, Green, Gerber and Lainer adjust for standard error by taking into account past voting behavior. They regard all five past elections, two general and three primary elections, as having the same electoral salience. I, however, think that electoral salience would play a role in how effective the treatment is on the individual's voting behavior. I hypothesize that treatment is likely to have a greater effect on those that did not vote for primaries but voted in general elections compared to those that did not vote for generals but voted in primary elections. I believe this would be the case because general elections compared to primaries have more political salience. According to Gerber, Huber, Biggers and Hendry, a sizable segment of electorate consider the stakes for primary elections, compared to general, to be lower and costs of the voting greater. People tend to feel less social pressure to vote and hold exclusionary beliefs about who should participate, and are more willing to turn to those that know and care more about the contests.(Gerber et al. 2017) Since people are less likely to vote in primaries, I hypothesize that treatment will have a greater effect on those that do not vote in the primaries versus those who do. Those who do in the primary election may have a higher propensity to vote than those who do not. Therefore, those that do not will be more likely to be affected by the treatment. In order to see how this may change the treatment effect on voter turnout, I divide the sample into four groups: 1) voted at least once in primary and general elections, 2) voted at least once in primary but never in general elections, 3) voted at least once in general but never in primary election, 4) never voted on either primary or general election. The treatment will not have that big of an impact on the always takers and never takers since they will vote or not vote no matter what. The comparison of interest would be those that vote for primaries but not for general and those that vote for general but not for primary elections. In order to do this, I made a new column categorizing the sample into four groups as indicated above. 1 is for the always takers, 2 for those that only voted in general elections, 3 for those that only voted in primary elections, and 4 for the never takers. I will compare the average treatment effect between 2 and 3 to see whether the treatment effects are different or not.

In order to compare the treatment effect for four groups, I categorized the sample into four groups depending on how they voted in the past. I then found the average treatment effect for each category by simply subtracting the mean of the voter outcome for those in the control group from the treatment group. The four outcomes are how much more likely the groups who were injected treatment (in this case, neighbor treatment) to vote compared to those in the control group. As expected, the treatment effect was the highest for the always voters/takers and lowest for the never voters/takers, with 9.3% and 2.4% for each. The treatment effect for those that voted in the general election but not in the primaries was 5.9% and those that voted in the primaries but not general elections, the effect was 8.3%. According to the results, those who voted in the primaries but not general elections showed a higher treatment effect compared to those that voted in generals but not the primaries.

I, however, am a little sceptical on the results this produced because when I looked at the number of observations for each subgroups, the sample size differed from one another. The subgroup that had the biggest number of sample was the always voters/takers group with 160,132 observations. The next biggest was the the general election voters with 49,335 and third biggest was the never voters/takers group with 14,488 observations. The smallest group was the primary voters with 5,489 observations. I, therefore, wonder if it would be the case with other treatment groups. Perhaps the subcategories are more balanced in terms of observation size. I will therefore try this again with the

“Self” treatment group.

The results using the “Self” treatment subset showed different results. First of all, the treatment effect was again greatest for the always voters/takers and lowest for the never voters and takers with 5.5% and 1.4% each. The treatment effect, however, was higher for those that voted in the general elections than those that voted in the primaries with 4.1% and 2.6% each. The number of observations again was different for all four subcategories with 160,099 observations for the always voter group, 49,317 for the general voters, 5471 for primary voters and 14574 for the never voters.

I am guessing there are generally fewer cases whether individuals vote in the primaries but not generals due to electoral salience. Presidential election would have higher electoral salience than the primaries therefore most people would vote in the general election but less so for the primaries. Therefore, the category where people vote for generals but not primaries is more frequent than the case where they vote for primaries but not for the general elections. But looking back on the number of observations for each subgroup, I wondered having around 5500 observations would be a problem or not. If it were a few hundred cases, I might worry that they may not be representative of the population. But with a few thousand cases, I’m wondering if I have to worry or not. Would I have to? I am uncertain. I thought about doing bootstrapping, but as I learned it in class, it was used for a dataset that had less than 100 observation. Therefore, I was contemplating whether this was necessary or not for a sample this big. Even if I increased the sample size to one comparable to the second subcategory, I wondered if it would make a big difference in the effect it produces.

7 Reference

Haiyan Huang.(2004) Multiple Hypothesis Testing and False Discovery Rate (STATC141). [<https://www.stat.berkeley.edu/~hhuang/STAT141/Lecture-FDR.pdf>]

Gerber, Alan S, Donald P Green, and Christopher W Larimer. 2008. “Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment.” *American Political Science Review* 102 (01). Cambridge Univ Press: 33–48.

Gerber, Alan S, Gregory A Huber, Daniel R Biggers, and David J Hendry. 2017. “Why Don’t People Vote in Us Primary Elections? Assessing Theoretical Explanations for Reduced Participation.” *Electoral Studies* 45. Elsevier: 119–29.

Sinclair, Betsy, Margaret McConnell, and Donald P Green. 2012. “Detecting Spillover Effects: Design and Analysis of Multilevel Experiments.” *American Journal of Political Science* 56 (4). Wiley Online Library: 1055–69.