

Effect of Phone Calls on Iowa Voter Turnout in 2002

Wren Landis, Devin Maya, Peter Li

Econ 124 Machine Learning for Economists

University of California, Santa Cruz

Lecturer Michael P. Leung

### Effect of Phone Calls on Iowa Voter Turnout

Elections are pivotal times in the United States. Every American has the power to vote for candidates that would help shape the economy towards what they believe will be beneficial towards everyone. However, this power often goes unnoticed by the majority of Americans. During midterm elections, less people in the country care about placing in their votes, which can make it harder for politics in Congress or the White House to change. Ideally, it would be in everyone's interest to vote for a candidate that best represents them. However, in reality, most Americans don't care or lack the time to care about voting.

Many forget to vote, while some choose to vote and advocate for the causes they're passionate about. This is why political campaigns often hire workers, regardless if it's for a presidential or midterm elections, to help encourage the demographic of people in the population that haven't voted to go out and vote for their candidate. These workers usually message registered voters either through phone calls or email.

As researchers, hired by multiple political campaigns, we wanted to know whether or not registered voters can be influenced or changed by encouraging phone calls to go out and vote. Our goal is to estimate whether there exists a causal effect on Iowa voter turnout for 2002 if someone received an encouraging phone call or not. We believe that registered voters that receive an encouraging phone call will choose to go vote. Thus, increasing the Iowa voter turnout in the country. If this hypothesis is true, we will be able to attempt to make a case to political campaigns that giving phone calls to registered voters and outwardly encouraging these voters to go out will increase the amount of potential Iowa voter turnout for their respective candidate. This can result in a win-win for the country because more people will vote, during both election seasons. This also means more people will participate in the political ecosystem to

help make their communities into better places. However, if the hypothesis is found not to be true, we might need to question our data or what type of controls we may have forgotten to consider when carrying out the research.

## **Data**

Our main purpose of this research report is to find whether there exists a causal effect between receiving a phone call on Iowa voter turnout results for 2002. For our research endeavors, we obtained a subset of data from a journal article called *Comparing Experimental and Matching Methods Using a Large-Scale Voter Mobilization Experiment*, written by Arceneaux et.al, Political Analysis, 2006, 14, 37-62. In this journal article, the researchers did an experiment using three treatments, high-contact phone calls, low-contact phone calls and no phone calls, on the effect of Iowa voter turnout rates. However, for the purpose of our research report, we will only think of these three treatments as two levels, where 1 = received encouraging phone calls and 0 = didn't receive encouraging phone calls.

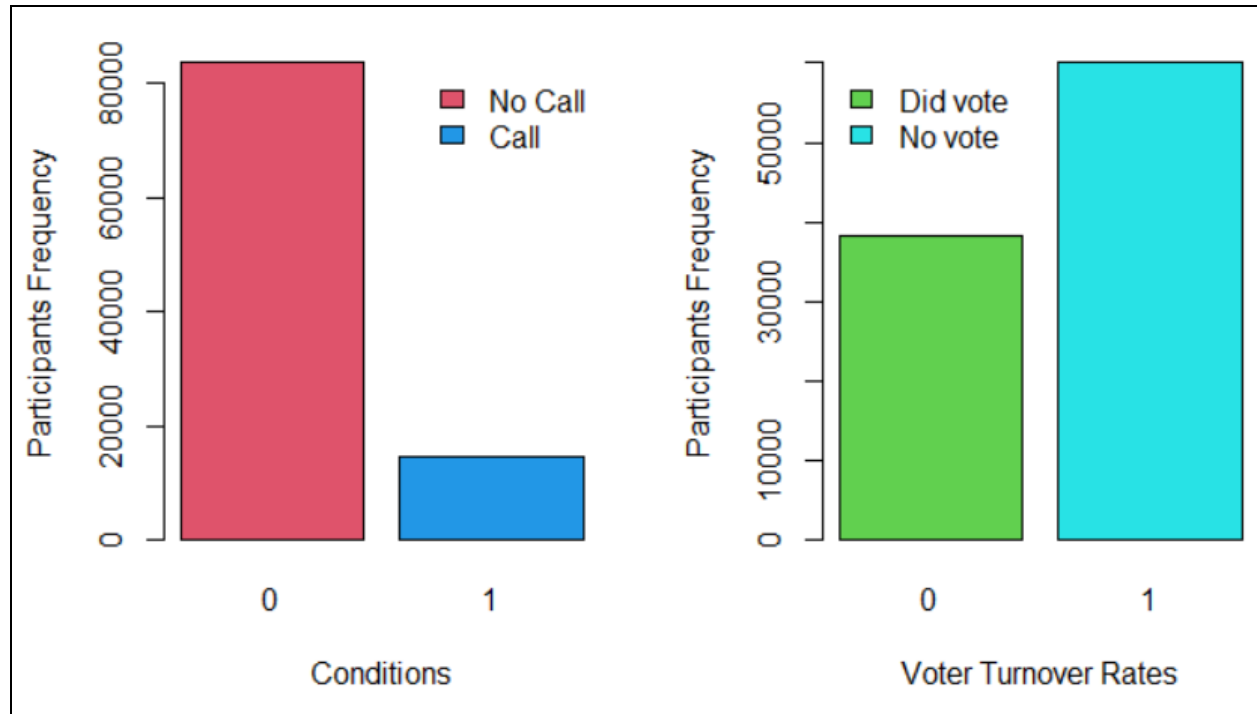
From our dataset, we intend to use the treatment variable called "treat\_real", which randomly assigns participants in the experiment in our two conditions (assignment to receive encouraging phone call = 1 and assignment to not get encouraging phone call = 0). We omit observations that were in our treatment group (1 = getting a phone call), but didn't answer the phone for our variable "contact" (1 = answered phone call). This is done in order to avoid the issues with compliance in the conducted study, thus allowing us to assume 100% compliance when being randomly assigned to the treatment group and control group. Given the randomization, we will achieve our "gold standard" estimate of the effect of getting assigned to receive a phone call on Iowa voter turnout in 2002.

If the participants received the phone call (treat\_real = 1), they'd encounter this message:

*“Hello, may I speak with (name of person) please? Hi. This is (caller's name) calling from Vote 2002, a nonpartisan effort working to encourage citizens to vote. We just wanted to remind you that elections are being held this Tuesday. The success of our democracy depends on whether we exercise our right to vote or not, so we hope you'll come out and vote this Tuesday. Can I count on you to vote next Tuesday?”*

Otherwise, participants didn't receive any message ( $\text{treat\_real} = 0$ ). For our outcome variable, we used a variable called “vote02”, which indicates a participant's voter outcome for the 2002 election. Furthermore it is a binary variable that determines whether an individual voted for the given year 2002.

Because the dataset is a randomized control trial, we don't believe our dataset has any limitations as it should satisfy all the conditions for the valid control assumption. However, if this dataset wasn't a randomized control trial, a limitation that may exist could be race because we believe that people who identify as a minority member will be more inclined to choose to not vote compared to its white counterparts. The reasoning behind this belief is that some minority members may face racial discrimination for trying to vote in America. So even if minorities were contacted to go vote, they may try to go vote but be deterred from actually voting when at the election polls. Which can deviate our results from its causal effect and weaken the effectiveness of the treatment. But, we don't need to worry about these confounded variables as we have a randomized control trial.



In the subset of the RCT dataset on Idaho voters, we found that in our treatment variable (“treat\_real”), there were 14,679 participants in our treatment condition while 83,688 participants in our control condition. For our control variable (“vote02”), we found that 60,066 participants ended up voting while 38,301 participants didn’t vote for the 2002 election.

## Methodology

For our research analysis, we will use an experimental design with treatment and control groups to analyze the causal effect of encouraging phone calls on voter turnout. This design allows us to isolate the effect of the independent variable phone call encouragement on the dependent variable voter turnout by comparing the outcomes of the treatment group (those who received a phone call) with the control group (those who did not receive a phone call). This approach relies heavily on the principle of random assignment, ensuring each participant has an equal chance of being assigned to either the treatment or control group. This helps to control for

any potential confounding variables or bias, resulting in groups that are expected to be similar in all respects except for the treatment.

Specifically, we will employ a two-step process combining Lasso and Ordinary Least Squares (OLS) regression models, commonly known as post-Lasso. This approach is beneficial when dealing with high dimensional data and allows us to select relevant variables and estimate the effect size accurately.

Our primary step includes Cross Validation with Lasso, a regression analysis method that performs both variable selection and regularization. It is most useful when dealing with a high number of potential regressors. By imposing a penalty term that is dependent on the data, Lasso minimizes or zero out some coefficients of certain regressors to decrease the regression model's complexity. Lasso uses this method to find the best penalty and optimal bias-variance trade-off for a regression model. The goal of satisfying this condition with LASSO is to find the best regression model that doesn't under-fit or over-fit our data. By balancing the amount of excess variance and bias we may incur for just implementing Ordinary Least Squares method, we will have a more reliable regression model that better predicts out-of-sample data. Thus, we implement our Lasso cross-validation method on both our controls and treatment variables, separately, to generate our reliable regression models that balances bias and variance.

Following this Lasso procedure, we will apply Ordinary Least Squares regression on each optimal regression model found. The rationale behind using OLS post-Lasso is that while Lasso is a powerful tool for variable selection, it may introduce bias due to its regularization component. In contrast, OLS is an unbiased estimator that provides us with more interpretable model parameters. Using post-Lasso, we are combining the strengths of both methods. Lasso's ability to efficiently deal with high-dimensional data and perform variable selection, and OLS's

less biased estimations on the selected subset of predictors. In terms of our research analysis, using post-Lasso will be especially useful if we have a multitude of potential predictor variables influencing voter turnout. The Lasso will allow us to select the most relevant variables, and then OLS will provide us with precise estimates of the effect of these predictors on voter turnout.

The main assumptions of this method include Independence of observations, Linearity of log odds, and Absence of Multicollinearity. Each individual's choice to vote is assumed to be independent of the choices of others. In our case, the random assignment in our experiment helps to ensure this condition is met. If the observations are not independent, it implies that the observations are influencing each other in some way. This can lead to misleading results as the model would attribute changes in the dependent variable to the independent variable, when in reality they are due to dependencies between observations. Logistic regression assumes a linear relationship between the logit of the outcome variable and the predictor variables. It may not be entirely plausible in our dataset if the effect of phone calls doesn't linearly translate to increased probability of voting. Meeting the assumption of multicollinearity requires that predictor variables are not highly correlated with each other. In our case, since we have only one primary predictor variable which is receiving a phone call, this assumption is likely to hold.

Large sample size is vital since logistic regression typically requires such a condition. Based on the data source, we expect this condition to be met. The model also assumes we have included all relevant variables and correctly specified the functional form.

The randomized control design helps to control for omitted variable bias, but there's always a risk that some relevant factors like political attitudes or awareness levels might not be accounted for. In the context of this study, these assumptions are plausible due to the

experimental design and the nature of the variables. Nevertheless, the potential limitations such as unaccounted factors influencing the decision to vote, or a non-linear relationship between the intervention and the outcome, will need to be considered when interpreting the results. Another key assumption for casual interpretation is representativeness. In order for a broader casual estimate to be drawn it must be assumed that the data in question, Iowan voters in the 2002 election, is representative of all voters across all time periods. Since there are significant population characteristic differences such as differences in regions and time-trends, in the real world this assumption almost certainly will not hold. So, any causal effects measured should be interpreted as applicable only to Iowan voters in the year 2002.

### Main results

OLS (Coefficient of Treatment)	Estimate Std. Error t value Pr(> t )			
	0.010326844	0.003687542	2.800468290	0.005103847

Given our results, we find that the encouragement of phone calls has a positive effect on voter turnout. More specifically, our models indicate that this intervention, receiving an encouraging phone call to go vote, will increase voter turnout rates by approximately 1.03 percentage points according to the OLS regression, and by approximately 0.89 percentage points according to the post-Lasso model.

Post-Lasso (Coefficient of Treatment)	Estimate Std. Error t value Pr(> t )
	0.00889887 0.00361652 2.46061713 0.01387153
Counterfactual (Suppose everyone receive an Engaging Phone Call)	[1] 0.00757092



The average of the counterfactual estimate suggests that when everyone receives an encouraging phone call to go out and vote, we will see a rise in voter turnout rates by approximately 0.76 percentage points. The differences between these estimates could stem from different variable selection and bias-variance trade-offs in the respective models. In addressing the research question: Does encouraging phone calls increase voter turnout? We have applied a variety of statistical models to estimate the causal effect. These models include Ordinary Least Squares regression, post-Lasso regression, and a counterfactual estimate.

For our OLS regression, a standard approach for estimating linear relationships between variables, provided an estimate of about 0.010 with a standard error of about 0.004. Statistically speaking, this result is significant, given a t-value of about 2.80 and a p-value of about 0.005, which is less than the conventional threshold of 0.05. In terms of an economic interpretation, the estimated coefficient suggests that for every 100 people who received an encouraging phone call to vote, approximately one more person voted compared to those who did not receive such a call. To address potential high dimensional covariates and control for potential bias-variance trade-offs, we also employed a post-Lasso regression model. The post-Lasso model yielded an estimate of about 0.009 with a standard error of about 0.004, a significant t-value of about 2.461, and p-value of about 0.014. The post-Lasso model, therefore, concurs with the OLS model, indicating that encouraging phone calls does indeed increase voter turnout rates. However, it suggests a slightly smaller effect than our OLS model, stating that voter turnout rates only increase by approximately one additional voter for every 112 individuals who received an encouraging phone call.

Finally, we considered a counterfactual scenario to understand the potential impact had there been phone call encouragement. The counterfactual estimate provided an average effect of

about 0.008, implying that with the phone call encouragement, the voter turnout rates would likely increase by around 0.76 percentage points.

The statistically significant results from all three models reaffirm the positive impact of phone call encouragement on voter turnout. The implications of these results suggest that governmental and non-governmental organizations could consider such phone call campaigns as part of their strategies to boost voter participation. However, given the relative magnitude of the effect, substantial outreach would be necessary to induce a large-scale impact on voter turnout. Further, while these findings provide an average effect, they do not capture possible variations in the effect across different demographic groups or over time, which may be an interesting avenue for future research.

Lastly, a counterfactual prediction exercise could be conducted to simulate voter turnout scenarios without phone call intervention. Such an exercise could further illuminate the potential reach and impact of phone call encouragement under varying assumptions about the baseline underlying characteristics to vote.

## **Conclusion**

In this paper, we rigorously examined the assumptions and critical components necessary when conducting research pertaining to a causal interpretation through the implementation of the ‘Gold Standard’ RCT design. The goal of our analysis was to estimate the effect of being randomly assigned to receive encouraging phone calls on voter turnout, which is a question of vital importance for democratic societies worldwide. However, given the significant regional differences and time trends that exist within real world populations, our representativeness assumption almost certainly does not hold. Political, social, and technological landscapes evolve over time, and these changes can significantly impact voter behavior and the effectiveness of

mobilization strategies. Therefore, the causal effects measured in this study should be interpreted as applicable only to the specific context of Iowa voters in 2002.

The actual economic significance of the effect found, a less than 1% increase in voter turnout, also warrants discussion. While statistically significant, the practical implications of this effect size may vary depending on factors such as the resources available for mobilization efforts, the relative costs, and the presence of alternative mobilization strategies.

In conclusion, while our findings point to a positive impact of phone call encouragement on voter turnout, these findings must be viewed within the context and timeframe they pertain to. Caution should be exercised in extrapolating these results to broader populations or more recent time periods. Future research should aim to replicate and expand on these findings within contemporary electoral contexts, taking into account the evolving dynamics of voter mobilization and the changing sociopolitical landscape. Through such endeavors, a more nuanced and generalizable understanding of effective voter mobilization strategies can be achieved.