

A Statistical Analysis of the Causality of Experimental and Non-Experimental Studies

Tristan Mesaros

University of California, Santa Cruz

Department of Economics

December 3, 2021

Abstract

This paper will compare the results of an experiment and a non-experiment to better understand which method is able to generate valid causal estimates through the use of data from a get-out-the-vote randomized control trial in Iowa. Regressions will be made using two methods, and will be compared using balance tables, sample characteristics and other statistical tools in order to first determine which method can successfully provide causal estimates. Using the experiment regression results as the best case, we will compare the results with the regression of the non-experiment case to get a better understanding of what is wrong with a non experiment where participants have a choice to take part in the experiment. We will then analyze the results by comparing them with a two-stage least squares regression in order to estimate the local average treatment effect. This will eliminate the bias that arises from abnormalities in the experiment such as non-compliers. The results were able to show that the coefficient of *contact* in the non-experiment was overstated compared to the coefficient of *treat_real* in the experiment. This is related to a positive bias which can be progressively removed by adding more controlling variables to the regression which results in the positive bias being less overstated.

1 Introduction

This study was made to increase vote turnout in the voting age population and to understand if a phone call would be able to influence that population into going out to vote using a randomized control trial and observational studies. The phone call consisted of an encouraging conversation between the caller and the callee which stated: *“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?”* In order to understand the effect the phone call has on voter turnout, an experimental and a non-experimental study must be conducted in order to really calculate the magnitude of that effect.

1.1 Experimental Study

The first of the two different study methods I will be discussing throughout this paper is the experimental study. An experiment is a study in which a treatment or procedure is intentionally given to a group and the results are observed and compared with that of the group that did not receive the treatment. This study consists of two groups, a treatment group and a control group. Participants from the voting population age were randomly assigned to either receive the phone call, which is the treatment group, or not get the phone call, which is the control group. The treatment group would receive a phone call encouraging the individuals to go vote in the upcoming 2002 Senate elections. The control group would not receive a phone call, however would still be observed to see if they voted in 2002 despite not receiving a call. Due to the balanced randomization of the treatment and control groups, this experimental study should provide a “gold standard” measurement of the impact of assigning an encouraging phone call on voter turnout. In order to calculate the effect that receiving phone call

has on voting in 2002, I will be using multi-variable regressions through the use of STATA. Using this "gold standard estimate", we are able to compare it with the results that are obtained when using a non-experimental study.

1.2 Non-Experimental Study

The second of the two different study methods I will be discussing is the non-experiment study. A non-experimental is a study in which the participants have the option to participate or not. This means the treatment group and control group are created based on the participants own will and decisions. This could cause somewhat of a complication between the two groups. Since the two groups were not chosen randomly from all the participants, there is opportunity for hidden bias to take place in between these two groups. This bias could cause our data to be unbalanced and skewed. Using the multi-variable regression I previously mentioned, I will be able to find the effect of the phone call on voting in 2002 in a non-experimental situation where the participants chose whether or not they want to be part of the treatment group.

After running the two regressions, the experimental study and the non-experimental, I was able to find large differences between the two. The first difference was the change in the balance tables of the studies which compares the different sample characteristic means between the treatment and the control group of both types of studies. The experimental study which chose the treatment and control group had a balance table which seemed very even and balanced, while the balance table of the non-experimental study, in which participants chose which group they were in, was too unbalanced to be able to deem it reasonable.

The second difference is seen in the regression tables of the two studies. The primary variable of the regressions, *treat_real* for the experimental study and *contact* for the non-experimental study, were very different. The coefficient of

the primary variable in the regression for the non-experimental study was overstated compared to that of the experimental study. This over-estimation of the coefficient on the primary variable was able to show positive bias, which signifies that the phone call had too large of an effect on the voter turnout of 2002. I will explain in further detail the reason as to why there was positive bias. However this positive bias in the non-experimental study compared to the "golden standard" that is the experimental is able to show that non-experimental studies are less able to retrieve a certain conclusion due to the unbalance in the treatment and control groups.

2 Data

The data used for these two studies was taken from a political science experiment conducted by Kevin Arceneaux, Alan S. Gerber, and Donald P. Green, that goes by the name *Comparing Experimental and Matching Methods Using a Large-Scale Voter Mobilization Experiment*. This political science paper speaks on a field experiment that was conducted in Iowa and Michigan prior to the 2002 midterm elections. For households containing one or two registered voters, that household was randomly selected into the treatment or the control group. As for the households with two people, only one of the two was selected to be assigned to the treatment or control group in order to keep the data balanced. The participants who were randomly selected into the treatment group would receive a call encouraging them to vote in the midterm elections of 2002. Two national phone banks were hired by the conductors of the experiment to read a get-out-the-vote message to the treatment group. The message reads, "*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?". The participants who were randomly selected into the control group would not receive the call in order to determine the effect that the phone call had on the treatment group's voter turnout compared to the control group's voter turnout.

Throughout this paper, I mention the two different primary variables for the experimental and non-experimental study. For the experimental study, the primary regressor is *treat_real* which signifies whether a participant is assigned to the treatment group or not, *treat_real*=0 means they are part of the control group while *treat_real*=1 means they are part of the treatment group.

When it comes to the non-experimental study, the participants are not randomly assigned to either group. Instead, all participants receive the phone call. To distinguish which participants are part of the new treatment and control group, we use *contact*, *contact*=0 is for the control group and *contact*=1 is for the treatment group. The reason why *treat_real* is no longer the primary regressor in the non-experimental study is because in the non-experimental study, technically, all participants are part of the treatment group because they all receive the call. The difference is that now the treatment group is determined by the participants picking up and listening to the phone call, that means the control group are the participants who received the call but did not pick up.

To determine the legitimacy of the treatment group and control group, during the experiment, a few sample characteristics of the participants were taken. These sample characteristics include; the participants age (*age*), if they were a newly registered voter (*newreg*), if they voted in the 1998 (*vote98*) and the 2000 elections (*vote00*), and their gender (*female*). Using the mean of the sample characteristics between the treatment and control groups we are able to determine whether the sample being randomly selected made the treatment and control groups well balanced and even.

After generating and analyzing Table 1, which is the comparison of the sample

characteristics between the control and treatment group in the experimental study, there are a few factors that stand out. The first being the very close similarities between the control and treatment group. As you can see, the difference in mean age between the two groups is about 0.165 years which corresponds to a little under two months. The difference in newly registered voters is about 0.1% which is very insignificant. As well as the difference on if the participants voted in 1998 and 2000 is less than 1% (0.4% and 0.6% respectively) between the control and treatment group. Lastly, both the control and the treatment group have around 56.3% female participants which is practically the same. The second would be the p-values of the experimental study balance table. As seen in Table 1, the p-values of the sample statistics are all very high. The very high p-values state that the difference in sample statistics between the control and treatment group are high enough to be statistically significant. This means that the differences between the control and treatment groups are not statistically significant.

Table 1: **Experimental Study**

	control mean	treat mean	difference	p-value
age	55.787	55.622	0.165	0.3249
newreg	0.048	0.047	0.001	0.7238
vote98	0.575	0.571	0.004	0.3309
vote00	0.735	0.729	0.006	0.1346
female	0.563	0.563	-0.000	0.9770
Observations	85889	14957	100846	

The balance table of the experimental study shows the difference in mean of the sample characteristics, *age*, *newreg*, *vote98*, *vote00*, and *female*, between the control and treatment group. This table also provides the p-values for the differences between the two groups.

The same sample characteristics were taken for the non-experimental study

as the experimental study. After generating and analyzing Table 2, which is difference in sample characteristics between the control and treatment group, there are once again 2 main takeaways. The first main takeaway from Table 2, is the larger difference between the mean sample statistics of the control and treatment group. To begin with, the negative unbalance in age is very high at about -5.2%, which makes the difference statistically significant. In addition, the difference in if the participants voted in 1998 and 2000 were both about -7.5% which very high when it comes to the voting age population. If we take a look at the p-value from Table 2, from the non-experimental study, we can see that all the p-values are very low. Since the p-values are all lower than .05, which is the significance level at 95%, it causes all these large differences in the sample characteristic's mean between the control and treatment group to be statistically significant which can cause positive or negative bias which we will go into deeper.

Table 2: **Non-Experimental Study**

	Control Mean	Treat Mean	difference	p-value
age	53.204	58.425	-5.221	0.0000
newreg	0.051	0.043	0.007	0.0401
vote98	0.536	0.611	-0.075	0.0000
vote00	0.694	0.770	-0.076	0.0000
female	0.547	0.581	-0.035	0.0000
Observations	8030	6927	14957	

The balance table of the non-experimental study shows the difference in mean of the sample characteristics, *age*, *newreg*, *vote98*, *vote00*, and *female*, between the control and treatment group. This table also provides the p-values for the differences between the two groups.

After assessing the balances of the treatment and control group in the experimental study, based on their sample characteristics, we were able to show that the treatment and control group were similar enough to be able to provide relevant and trustworthy results. However the unbalances in the sample characteristics of the treatment and control group in the non-experiment were able to show that the differences were too high, which

could in the end result possibly cause some type of bias. The differences in the balance table of the non-experimental study were too statistically significant to provide a trustworthy estimation of the effect of receiving a phone call has on voter turnout in the 2002 midterm elections.

3 Empirical Methods

In order to gain a better understanding of the causal effects, we must analyze the primary variables being regressed on both studies. For the experimental study, the main variable is *treat_real*, which, along with its coefficient is able to determine the effect that the treatment has on whether the participant will vote in the upcoming midterm election. As for the non-experimental study, due to the fact that the participants chose if they want the treatment, the primary variable is now if they pick up the phone call, *contact*.

3.1 Experimental Study Regression

A regression will be made on the *vote02* which includes the primary variable, *treat_real*, and the sample characteristic variables, *age*, *newreg*, *vote98*, *vote00*, and *female*. Multiple regressions will be made where a sample characteristic variable is added after each regression to determine the effect that these covariates have on the primary variable *treat_real*. Since this is an experimental study based on a randomized control trial, the treatment is *treat_real* because the participants are randomly assigned. A table of regressions will be made using this regression equation first;

$$vote02_i = \beta_0 + \beta_1 treat_real_i + u_i$$

In that same regression table, we will add more regressions with all the covariates one by one to find the change in coefficients as more are added. The final

regression equation should look like;

$$vote02_i = \beta_0 + \beta_1 treat_real_i + \beta_2 age_i + \beta_3 newreg_i + \beta_4 vote98_i + \beta_5 vote00_i + \beta_6 female_i + u_i$$

After generating Table 3, which is the table of regression for the experimental study, we can now move on to the results of our findings.

Table 3: **Experimental Study Regression Table**

	1	2	3	4	5	6
Variables	vote02	vote02	vote02	vote02	vote02	vote02
treat_real	0.00474 (0.00435)	0.00454 (0.00431)	0.00547 (0.00422)	0.00652 (0.00423)	0.00903 (0.00379)	0.00906 (0.00366)
newreg		-0.306 (0.00718)	-0.214 (0.00717)	-0.212 (0.00720)	0.138 (0.00720)	0.164 (0.00661)
age			0.00529 (8.07e-05)	0.00582 (8.19e-05)	0.00361 (7.49e-05)	0.00175 (7.54e-05)
female				-0.0310 (0.00304)	-0.0271 (0.00273)	-0.0235 (0.00264)
vote00					0.527 (0.00342)	0.397 (0.00363)
vote98						0.271 0.271
constant	0.595 (0.00167)	0.610 (0.00167)	0.311 (0.00486)	0.315 (0.00505)	0.0209 (0.00492)	0.0626 (0.00477)
Observations	100,846	100,846	100,846	98,284	98,284	98,284
R-Squared	0.000	0.018	0.058	0.066	0.248	0.300

The regression table from the experimental study shows change in the effect of the primary variable, *treat_real* has on the voter turnout of the 2002 midterm election as we add more sample characteristics covariates. The standard errors can be found in parentheses.

3.2 Non-Experimental Study Regression

For the non-experimental study, the primary variable is no longer *treat_real* due to the fact that now the treatment and control group are chosen by the participants themselves. Since the non-experimental study is based on an observational study, this means that the new treatment group is not if they received the phone call but if they listened to the phone call, which is why the new

primary variable is now *contact*. A second table of regressions will be made for the non-experimental study using this regression equation first;

$$vote02_i = \beta_0 + \beta_1 contact_i + u_i$$

In that same regression table, we will add more regressions with all the covariates one by one to find the change in coefficients as more are added. The final regression equation should look like;

$$vote02_i = \beta_0 + \beta_1 contact_i + \beta_2 age_i + \beta_3 newreg_i + \beta_4 vote98_i + \beta_5 vote00_i + \beta_6 female_i + u_i$$

After generating Table 4, which is the table of regression for the non-experimental study, we can now move on to explain the results that accompany our findings.

Table 4: Non-Experimental Study Regression Table

	1	2	3	4	5	6
Variables	vote02	vote02	vote02	vote02	vote02	vote02
contact	0.120 (0.00797)	0.118 (0.00789)	0.0920 (0.00782)	0.0773 (0.00784)	0.0597 (0.00705)	0.0575 (0.00678)
newreg		-0.330 (0.0185)	-0.239 (0.0186)	-0.239 (0.0186)	0.104 (0.0177)	0.136 (0.0171)
age			0.00506 (0.000211)	0.00566 (0.000214)	0.00356 (0.000196)	0.00166 (0.000196)
female				-0.0172 (0.00784)	-0.0151 (0.00705)	-0.0129 (0.00678)
vote00					0.516 (0.00877)	0.391 (0.00920)
vote98						0.276 (0.00809)
constant	0.545 (0.00543)	0.561 (0.00545)	0.288 (0.0126)	0.287 (0.0131)	0.00798 (0.0127)	0.0474 (0.0122)
Observations	14,957	14,957	14,957	14,565	14,565	14,565
R-Squared	0.015	0.035	0.071	0.079	0.256	0.311

The regression table from the non-experimental study shows change in the effect of the primary variable, *contact* has on the voter turnout of the 2002 midterm election as we add more sample characteristics covariates. The standard errors can be found in parentheses.

In the experimental study regression table (Table 3) as the sample statistic covariates are added to the regression, the coefficient on *treat_real* slowly increases meaning the effect that the treatment has on the voter turnout of the midterm elections of 2002 is increasing. The effect of the treatment increases because the error term u is holding less effect because the covariates are now being introduced into the regression. In the non-experimental study regression table (Table 4). As more sample characteristic covariates are added, the coefficient on the primary variable *contact* decreases meaning that the effect that the primary variable *contact* has on the voter turnout of the 2002 midterm election is also decreasing. This is quite strange because it is behaving in the complete opposite of the experimental study.

3.3 Instrumental Variable Analysis

Instrumental variables help get a better understanding of the effect of the treatment on *vote02*, because it accounts for the people from the treatment group that did not take the treatment and the people from the control group that took the treatment. For calculating the instrumental variables, I will calculate the intention to treat (ITT) and the local average treatment effect (LATE). To do so we must first intention to treat which is also known as the reduced form, then the first stage, then the IV estimate by dividing the reduced form by the first stage. To find the reduced form or ITT, we regress *treat_real* on *vote02*,

$$vote02_i = \phi_0 + \phi_1 treat_real_i + \epsilon_i$$

Using this regression, the greek letter phi 1 gives us the intention to treat which in this case is about 0.00474. Now that we have found the intention to treat, we must find the first stage estimate using a regression where we regress *treat_real* on *contact*.

$$contact_i = \phi_0 + \phi_1 treat_real_i + \epsilon_i$$

Using this regression, phi 1 in this case is the first stage. I was able to find that the first stage estimate is about 0.463, this means there is a 46.3% difference in proportion that got the treatment between the treatment and control group. Now in order to find the local average treatment effect or the LATE, we must divide the intention to treat by the first stage estimate.

$$LATE = \frac{reducedform}{firststage} = \frac{0.00474}{0.463} = 0.0102$$

The local average treatment effect is about 0.0102 or 1.02%. After generating Table 5, which is a regression table using the two-stage least squares IV, it was also able to show that the local average treatment effect is in fact 0.0102. This estimate shows a much more accurate estimate of the treatment on voting in 2002, because it accounts for abnormalities such as non-compliers.

Table 5: **Two-Stage Least Squares Table**

	1	2	3	4	5	6
Variables	vote02	vote02	vote02	vote02	vote02	vote02
contact	0.0102 (1.09)	0.00980 (1.05)	0.0118 (1.30)	0.0139 (1.54)	0.0192 (2.38)	0.0193 (2.48)
newreg		-0.306 (-42.62)	-0.214 (-29.87)	-0.212 (-29.51)	0.138 (20.09)	0.164 (24.77)
age			0.00528 (65.32)	0.00581 (70.82)	0.00360 (48.04)	0.00174 (23.05)
female				-0.0310 (-10.19)	-0.0272 (-9.95)	-0.0235 (-8.92)
vote00					0.527 (154.22)	0.397 (109.27)
vote98						0.271 (85.34)
Constant	0.595 (355.71)	0.610 (359.88)	0.311 (64.37)	0.315 (62.77)	0.0217 (4.43)	0.0633 (13.36)
Observations	100846	100846	100846	98284	98284	98284

This table shows the new regression estimates using two-stage least square instrumental variables.

4 Results

Now that we have provided an explanation of the treatment and control group of both the experimental and non-experimental study, as well as an analysis of the regression tables for both the experimental and non-experimental study, we are able to gather the information to further explain the results.

After going over the findings from Table 3, the experimental study regression table, there are a few points to be made about the coefficient on the primary variable as we add covariates. In the first column which is the regression including just *treat_real* the coefficient is about .00474, this means that if a participant takes the treatment, their predicted chance of voting in the 2002 election will increase by about 0.474%. As we add more and more covariates, the coefficient on *treat_real* slowly increases. As the primary variable coefficient increases, the treatment has a larger effect on the voter turnout of 2002. The largest jump is explained by the addition of the covariate *vote00*. This can be explained by the fact that the coefficient on *vote00* is no longer hidden in the error term u . The coefficient on *vote00* is about 0.527 which signifies that if the participant voted in 2000, their predicted probability of voting in the upcoming 2002 election will increase by 52.7%. This makes sense because if someone voted in 2000, it probably means they are politically involved which makes it more likely for them to vote in 2002. As all the covariates are added into the regression, the coefficient on the primary variable *treat_real* increases from 0.00474 all the way to 0.00906. This is because the effect of all the covariates, the sample characteristics, was not explained and was hidden in the error term u .

After going over the findings from Table 4, the non-experimental study regression table, I was able to deduce the following. The first main takeaway from this regression table (Table 4) on the non-experimental study is the very high coefficient on primary variable which is in this case *contact*. As you can see, this coefficient is about 0.12, this means that if a participant were to listen

to the phone call, their predicted probability of voting in the upcoming 2002 midterm election will increase by 12%. The t-stat for the regression on just the primary variable is very high, $0.12/0.00797 = 15.06$. A very high t-stat means that there is an effect being made, however this effect can be related to the unbalance in the treatment and control group. As we add more covariates such as *age*, *newreg*, *vote98*, *vote00*, and *female*, the coefficient on the primary variable decreases substantially from about 0.12 all the way to 0.0575. The largest difference in the coefficient on *contact* can be seen when the covariate *age* is added. The change in the coefficient is about $0.118 - 0.0920 = .026$, which is about a 2.6% change on the effect of the treatment. The large difference in the primary variable coefficient once the covariate *age* is added can be explained by the difference in the balance table (Table 2). As the rest of the covariates are added, the primary variable coefficient decreases which is quite strange.

As we have seen in past few paragraphs, the difference in the two regressions, the experimental regression and the non-experimental regression, were omnipresent. To start off, the experimental regression's primary variable, *treat_real*, slowly increases as we add more sample characteristic covariates. This can be explained by the fact that the covariates being added reduces the error term u which increases the effect that the treatment has on the voter turnout of the 2002 midterm election. When it comes to the non-experimental regression, the coefficient on the primary variable, *contact*, increases at a fast rate as we add the covariates. There are two reasons as to why the primary coefficient decreases. The first reason is because due to the large differences between the control and treatment group in Table 2. Since the treatment group is substantially older, the treatment group participants were most likely already planning on voting due to their political involvement in the previous elections of 1998 and 2000. If we use the experimental study as a "gold standard", like we previously mentioned, the coefficient on *contact* in the non-experimental study is over-estimated meaning that a bias takes place. The

bias that takes place is a positive bias because of the difference in age and if the participants voted in 1998 and 2000 make it so that there were already going to be more participants voting in the treatment group which will make the effect of the phone call on the voter turnout of 2002 overestimated. The difference in age is what changes the coefficient on *contact* the most because as we added the covariate *age* to the regression, we had the biggest change of the effect that the phone call has on the voter turnout of 2002. Adding covariates in Table 2 reduced the bias. The positive bias meant that the coefficient on the treatment in the non-experiment was overstated compared to the experimental study. The bias was reduced because as we added more covariates, the coefficient on the treatment (*contact*) decreased which means it was being less and less overstated. However, adding covariates to the regression did not eliminate all the bias in the estimates because there are always more covariates that one can add to a regression to reduce the bias. We can also find in the table that besides the effect of *age*, the negative coefficient estimate of newly registered voters (*newreg*) means that if the participant is a newly registered voter, the chance that they are going to vote in the upcoming 2002 elections are lower than if they were not a newly registered voter. There are a very large amount of covariates that can be added that will most likely carry a very small amount of bias. We can also say that it did not eliminate all the bias because the coefficient on *treat_real* in the experimental study was at about .009 after adding the covariates, and the coefficient on *contact* in this assignment was at .0575 after adding the covariates. There is still a little bit of positive bias left to be removed. When comparing the non-experimental study and the experimental study with the two-stage least squares regression (Table 5), we are able to see that the coefficient estimates from the experimental study's regression table keeps increasing and ends up almost reaching the 0.0102 from the two-stage least squares IV estimate.

4.1 Balance Table Results

As previously described, there were a few differences between the sample characteristics mean of the treatment and control group of the experimental and non-experimental study. The first would be the strong unbalance in age of the treatment and control group in the non-experimental study. The 5.2 year difference in age between the two too large which is why it had a p-value near zero. This makes the difference too high to be statistically insignificant. The strong difference in age can cause bias in the way that younger people such as millennials are less likely to vote rather than seniors. Younger people might not have as much time on their hand or political interest as older people. This can cause the effect of the treatment in the non-experimental study to be overestimated. The difference in age between the two groups also affects the differences in the other sample characteristics such as if the participants voted in the 1998 and 2000 election. In Table 2, which is balance table for the non-experimental study, the difference between the treatment and control on if the participants voted in 1998 and 2000 is about 7.5% for both. This difference in `vote00` and `vote98` can be explained by the difference in age due to the fact that the control group is younger which means they have most likely not voted in the previous elections as well.

5 Conclusion

To conclude this combination of the two experiments, the results that were found made it able to deduce a few conclusions. To begin, as a whole, a non-experimental study is not a very trustworthy experiment. With the help of the balance table (Table 2) and the regression table (Table 4), I was able to find that when the treatment group and control group are not randomly assigned to a certain group problems arise. The issue with the participants choosing what group they would want to be in creates selection bias. The selection bias comes from the fact that since the treatment group are the people that listen to the

phone call in the non-experimental study, these participants were most likely already planning on listening to the phone call and voting in the upcoming 2002 midterm elections. As shown previously, the non-experimental study is not able to provide causal effects of the treatment on the voter turnout of 2002 due to the large positive selection bias that was associated with it. As we added more covariates to the non-experimental regression, the omitted variable bias that came from the lack of the sample characteristic variables started to decrease because more of the bias was being explained by the covariates that were hiding in the error term u .

6 References

- Arceneaux, Kevin, et al. "Comparing Experimental and Matching Methods Using a Large-Scale Voter Mobilization Experiment." *Political Analysis*, vol. 14, no. 1, [Oxford University Press, Society for Political Methodology], 2006, pp. 37–62, <http://www.jstor.org/stable/25791834>.
- "Voter Turnout in Presidential Elections." Voter Turnout in Presidential Elections, The American Presidency Project, <https://www.presidency.ucsb.edu/statistics/data/voter-turnout-in-presidential-elections>.