Final Project: Replicating "Voting Made Safe and Easy: The Impact of e-voting on Citizen Perceptions"

Minerva University

CS130 - Statistical Modeling: Prediction and Causal Inference

Prof. Alexis Diamond

April 23, 2022

Completed by: Georgi Sokolov and Helen Prykhodko

Replicating "Voting Made Safe and Easy: The Impact of e-voting on Citizen Perceptions"

Executive summary

We replicated the findings of Alvarez et al. (2013) and extended their analysis using a more accurate statistical method. We were able to exactly replicate the findings of Alvarez et al. (2013) using the statistical methods also used by the authors of the original paper. However, we were only able to replicate the 2 out of 8 findings using a more accurate genetic matching method. Thus, we invite Alvarez and colleagues to re-evaluate their findings using better statistical tools and rethink the conclusions of their research.

Introduction

In their paper Voting Made Safe and Easy: The Impact of e-voting on Citizen Perceptions, Alvarez et al. (2013) evaluated the question "What is the causal effect of e-voting on the voting experience in Salta, Argentina?" Specifically, the authors had three types of questions:

1. Questions about the overall voting experiences such as assessing the preference for selecting candidates electronically.
2. Questions about the perception of ease of use such as assessing the quickness of the process.
3. Questions about the confidence in the election process such as assessing the perception of "cleanness" of the electoral process.

The authors acknowledge that a random assignment of voters to either e-voting or traditional polling sites would have produced the most accurate results for the purpose of causal inference. However, they claim that a randomized experiment was unfeasible given financial and other practical constraints. Instead, they performed a quasi-experiment. To ease the transition to e-voting, the decision was made to place e-voting locations in places, with on average more educated and digitalized demographics, inevitably introducing bias to the data. Salta has 850,000 voters, the e-voting was introduced to 36 locations, which accounts for approximately 13,000 voters. On the election day, across the locations, a questionnaire was administered to 1,502 voters (887 of whom used e-voting) (Alvarez et al., 2013). To correct for the selection bias, the authors decided to use Propensity Score matching - they found a match with similar relevant observed independent variables for an e-voter from the sample of traditional voters. However, the authors acknowledge that there might be unobserved confounds. To assess the sensitivity of matching estimates, they conducted the Rosenbaum bounds test. The authors made the following conclusions:

1. E-voting positively impacts the evaluation of the voting system.
2. E-voting does not significantly impact the perception of the speed of the voting process.
3. E-voting positively affects the perception of the cleanliness of the elections but has a small negative effect on ballot secrecy.

Replication

In this section, we replicate Table 3 from Alvarez et al. (2013) (see Table A). This table illustrates the proportion of e-voters and traditional voters who have positively responded to questions on the left before and after propensity score matching.

TABLE 3 *Causal Effect of e-voting*

	Before matching (N = 1,475)					After matching (N = 1,164)				
	N	E-Voting (%)	Traditional Voting (%)	Diff.	p-value*	N	E-Voting (%)	Traditional Voting (%)	Diff.	p-value*
Select candidates electronically	1,388	83.8	53.4	30.4	0.000	1,101	82.7	54.1	28.6	0.000
Evaluation of voting experience	1,460	46.3	21.3	25.0	0.000	1,151	45.6	20.9	24.7	0.000
Ease of voting procedure	1,469	33.6	11.5	22.1	0.000	1,159	32.5	11.9	20.6	0.000
Agree substitute TV by EV	1,409	84.1	62.4	21.7	0.000	1,114	82.4	63.3	19.1	0.000
Elections in Salta are clean	1,284	58.0	41.0	17.0	0.000	1,022	57.6	41.5	16.0	0.000
Sure vote was counted	1,418	86.3	77.0	9.3	0.000	1,117	85.7	77.0	8.8	0.000
Qualification of poll workers	1,416	85.1	76.2	8.9	0.000	1,123	84.5	76.0	8.5	0.000
Speed of voting process	1,443	84.1	80.8	3.2	0.130	1,137	83.2	80.7	2.5	0.306
Confident ballot secret	1,431	77.1	84.5	−7.4	0.001	1,133	76.9	84.3	−7.4	0.002

Note: sample sizes given in the first column (before matching) differ from those given in Table 1 because Table 3 omits respondents with missing values in covariates.
*Test of difference in proportions.

Table A. Table 3 from Alvarez et al. (2013).

To replicate the table, we first load the data into the R environment, load necessary libraries, and declare outcome variables:

```
# loading libraries
library(MatchIt)
install.packages("Zelig")
library(Zelig)
library(rbounds)

# loading the data
load("C:/Users/Georgi PC/Downloads/dataverse_files/datamatch.RData")

# declaring outcome variables
outcomes <- datamatch[10:18]
outcomes.lbls <- names(outcomes)
n.outcomes <- dim(outcomes)[2]
```

Then, we calculate the proportions of e-voters and traditional voters that responded to questions positively before matching and store the results in a matrix that will later be turned into a table.

```
# Drop observations with missing values in covariates
datamatch[, 10:18][is.na(datamatch[, 10:18]) == "TRUE"] <- 99999
datamatch <- na.omit(datamatch)
datamatch[datamatch == 99999] <- NA
```

```
outcomes.pre <- datamatch[10:18]

# constructing Table 3 before matching
tab3.pre <- matrix(NA,nrow = n.outcomes,ncol = 5)
rownames(tab3.pre) <- outcomes.lbls
colnames(tab3.pre) <- c("N", "prop.ev", "prop.tv", "diff", "pvalue")

for (i in 1:n.outcomes) {
  tab3.pre[i, 1] <- length(na.omit(outcomes.pre[, i]))
  tab3.pre[i, 2:3] <- rev(prop.table(table(outcomes.pre[,i],datamatch$EV),2)[2,])*100
  tab3.pre[i, 4] <- tab3.pre[i, 2] - tab3.pre[i, 3]
  tab3.pre[i, 5] <- prop.test(table(outcomes.pre[, i], datamatch$EV)[2, ], n =
apply(table(outcomes.pre[, i], datamatch$EV), 2, sum))$p.value
}

datamatch[, 10:18][is.na(datamatch[, 10:18]) == "TRUE"] <- 99999
```

Then, we perform Propensity Score Matching. Propensity Score Matching is a way to approximate the Randmozied Control Trial by constructing an artificial control group by matching each treated unit to control with similar covariates. Such an approach, in theory, should allow us to make causal inferences about the data. When performing Matching, it is impossible to get perfect matches for each of the covariates in a real-world setting. In the case of Propensity Score Matching, we calculate the probability that a set of covariates relates to a treated unit (Propensity Score) and match based on this score. In our case, we take variables such as gender and education, calculate the probability that a unit with these specific scores was exposed to e-voting, and find a control unit with the same Propensity Score.

```
set.seed(36466)
m.out <- matchit(EV ~ age.group + I(age.group^2) + I(age.group^3) + age.group:educ +
age.group:tech + educ + I(educ^2) + tech + I(tech^2) + pol.info + educ:pol.info +
age.group:pol.info + tech:pol.info + white.collar + not.full.time + male, caliper = 0.05,
data = datamatch, method = "nearest")

print("Balance Improvement")
print(summary(m.out))
```

```
# matched sample
datamatched <- match.data(m.out)
datamatched[datamatched == 99999] <- NA
```

Finally, we populate the table with after matching proportions and print it:

```
outcomes.post <- datamatched[10:18]

tab3.post <- matrix(NA, nrow = n.outcomes, ncol = 5)
rownames(tab3.post) <- outcomes.lbls
colnames(tab3.post) <- c("N", "prop.ev", "prop.tv", "diff", "pvalue")

for (i in 1:n.outcomes) {
  tab3.post[i, 1] <- length(na.omit(outcomes.post[, i]))
  tab3.post[i, 2:3] <- rev(prop.table(table(outcomes.post[, i], datamatched$EV), 2)[2, ]) *
100
  tab3.post[i, 4] <- tab3.post[i, 2] - tab3.post[i, 3]
  tab3.post[i, 5] <- prop.test(table(outcomes.post[, i], datamatched$EV)[2, ], n =
apply(table(outcomes.post[, i], datamatched$EV), 2, sum))$p.value
}

tab3 <- cbind(tab3.pre, tab3.post)

tab3 <- tab3[rev(order(tab3[, 9])), ]

### Table 3 ###

print(tab3, digits = 4)
```

	N	prop.ev	prop.tv	diff	pvalue	N	prop.ev	prop.tv	diff	pvalue
eselect.cand	1388	83.84	53.42	30.428	1.237e-34	1107	82.14	53.75	28.395	7.444e-24
eval.voting	1460	46.33	21.30	25.035	1.833e-22	1155	44.73	20.66	24.073	4.925e-18
agree.evoting	1409	84.14	62.44	21.705	2.864e-20	1120	83.10	62.72	20.372	2.821e-14
easy.voting	1469	33.64	11.53	22.111	5.420e-22	1163	31.15	11.34	19.813	2.644e-16
how.clean	1284	57.97	40.99	16.980	2.561e-09	1025	58.08	41.22	16.862	9.550e-08
sure.counted	1418	86.35	77.02	9.332	7.444e-06	1121	85.92	77.32	8.597	2.740e-04
capable.auth	1416	85.14	76.25	8.889	2.954e-05	1126	84.88	76.60	8.278	5.845e-04
speed	1443	84.06	80.85	3.209	1.298e-01	1139	83.10	80.88	2.216	3.701e-01
conf.secret	1431	77.11	84.53	-7.417	6.506e-04	1136	77.68	84.55	-6.870	3.918e-03

Table B. The replication of Table 3 from Alvarez et al. (2013).

The replicated results match the results presented by Alvarez et al. (2013) and are in line with the conclusions stated by the researchers: while e-voting does not significantly impact the perception of voting speed, it positively impacts the overall voting experience and its perceived "cleanliness."

Extension

The authors used propensity score matching - a method that 'summarizes' all the covariates into one probability number between 0 and 1 for each unit. In our case, it is the probability the voter is assigned to an e-voting location based on their observable attributes, such as demographics and socioeconomic status. According to pseudo-randomization, E-voters are paired up with traditional voters: pairs are created based on confounders' similarities, increasing the conditional probability and allowing for a sufficiently similar propensity score.

While this method is supposed to reduce selection bias and create balanced data for causal inference, it happens that propensity matching produces the opposite. When balancing covariates, we can use complete randomization or fully blocked experiments. The weakness of propensity matching comes from accomplishing complete randomization, which estimates the balance only on average; thus, it is 'blind' to increased imbalance, model dependence, and bias (King & Nielsen, 2019). In contrast, the fully blocked experiment, used in other matching methods, allows for more effective results since it achieves the exact observed covariate balance rather than its approximation.

Hence, to replicate the study on e-voting, we decided to change the matching method used in the paper and proceed with genetic matching. We use this method because it is much better in obtaining balanced groups than propensity matching - genetic matching allows for matching with the worst balance to be as good as it can be, getting the lowest p-value before matching as high as possible after matching. The genetic matching algorithm improves propensity scores by using a mix of propensity score matching and Mahalanobis distance matching - a greedy match, where the order in which we match control to treatment rules the next matches (Diamond & Sekhon, 2013). We conducted genetic matching on nine outcome variables, following these steps:

1. Run *Genmatch()* function to find the best scaling factors on all the covariates and store the values in *genout* variable.
2. Match with a *Match()* function using the previously obtained weights, and store values in *mout*.
3. Check the balance of covariates after matching with *MatchBalance()*.
4. Run *Match()*, including outcome variables to find treatment effect estimates, standard errors, and p-values for results.

The results are presented in Table C.

Variable	**Estimate**	**Standard Error**	**p-value**
Select candidates electronically	95.547	1572.5	0.95155
Evaluation of voting experience	-1028.7	763.38	0.17779
Ease of voting procedure	208.09	408.69	0.61064
Agree substitute TV by EV	313.14	1362.2	0.81819
Elections in Salta are clean	3733.7	2302.3	0.10487

Sure vote was counted	-1273	1393.3	0.3609
Qualification of poll workers	4373.6	1333.5	**0.0010386**
Speed of voting process	-238.93	859.83	0.7811
Confident ballot secret	2685.3	1137.1	**0.018198**

Table C. The extension of Alvarez et al. (2013) results using Genetic Matching.

Genetic Matching Code:

```
library(Matching)
set.seed(2324);

# determining covariate weights
genout <- GenMatch(Tr = datamatch$EV, X = cbind(datamatch$age.group,
datamatch$educ, datamatch$tech, datamatch$pol.info, datamatch$white.collar,
datamatch$not.full.time, datamatch$male), data=datamatch)
summary(genout)

mout <- Match(Y = datamatch$eselect.cand, Tr = datamatch$EV, X =
cbind(datamatch$age.group, datamatch$educ, datamatch$tech, datamatch$pol.info,
datamatch$white.collar, datamatch$not.full.time, datamatch$male), Weight.matrix =
genout)

# checking covariate balance after matching
MatchBalance(EV ~ age.group + educ + tech + pol.info + white.collar + not.full.time +
male, match.out = mout, data=datamatch)

# determimning estimates and standard errors for all of the Y-variables
mout1 <- Match(Y = datamatch$eselect.cand, Tr = datamatch$EV, X =
cbind(datamatch$age.group, datamatch$educ, datamatch$tech, datamatch$pol.info,
datamatch$white.collar, datamatch$not.full.time, datamatch$male), Weight.matrix =
genout)
mout2 <- Match(Y = datamatch$eval.voting, Tr = datamatch$EV, X =
cbind(datamatch$age.group, datamatch$educ, datamatch$tech, datamatch$pol.info,
datamatch$white.collar, datamatch$not.full.time, datamatch$male), Weight.matrix =
genout)
mout3 <- Match(Y = datamatch$easy.voting, Tr = datamatch$EV, X =
cbind(datamatch$age.group, datamatch$educ, datamatch$tech, datamatch$pol.info,
```

```
datamatch$white.collar, datamatch$not.full.time, datamatch$male), Weight.matrix =
genout)
mout4 <- Match(Y = datamatch$agree.evoting, Tr = datamatch$EV, X =
cbind(datamatch$age.group, datamatch$educ, datamatch$tech, datamatch$pol.info,
datamatch$white.collar, datamatch$not.full.time, datamatch$male), Weight.matrix =
genout)
mout5 <- Match(Y = datamatch$how.clean, Tr = datamatch$EV, X =
cbind(datamatch$age.group, datamatch$educ, datamatch$tech, datamatch$pol.info,
datamatch$white.collar, datamatch$not.full.time, datamatch$male), Weight.matrix =
genout)
mout6 <- Match(Y = datamatch$sure.counted, Tr = datamatch$EV, X =
cbind(datamatch$age.group, datamatch$educ, datamatch$tech, datamatch$pol.info,
datamatch$white.collar, datamatch$not.full.time, datamatch$male), Weight.matrix =
genout)
mout7 <- Match(Y = datamatch$capable.auth, Tr = datamatch$EV, X =
cbind(datamatch$age.group, datamatch$educ, datamatch$tech, datamatch$pol.info,
datamatch$white.collar, datamatch$not.full.time, datamatch$male), Weight.matrix =
genout)
mout8 <- Match(Y = datamatch$speed, Tr = datamatch$EV, X =
cbind(datamatch$age.group, datamatch$educ, datamatch$tech, datamatch$pol.info,
datamatch$white.collar, datamatch$not.full.time, datamatch$male), Weight.matrix =
genout)
mout9 <- Match(Y = datamatch$conf.secret, Tr = datamatch$EV, X =
cbind(datamatch$age.group, datamatch$educ, datamatch$tech, datamatch$pol.info,
datamatch$white.collar, datamatch$not.full.time, datamatch$male), Weight.matrix =
genout)

# printing summaries
summary(mout1)
summary(mout2)
summary(mout3)
summary(mout4)
summary(mout5)
summary(mout6)
summary(mout7)
summary(mout8)
summary(mout9)
```

Discussion

Our paper provides new methodological contributions to the results on e-voting. Although we were able to replicate the exact results obtained by Alvarez et al. (2013) using

Propensity Score Matching, the Genetic Matching approach produced significantly different outcomes, challenging the prior findings. In the original paper, Alvarez et al. (2013) observed a statistically significant effect of e-voting on 8 out of 9 measured dependent variables as seen in Table A and Table B. However, after constructing an artificial control group using Genetic Matching, we only observed statistically significant differences between e-voters and traditional voters on 2 out of 9 variables. Specifically, as seen in Table C, the e-voters were significantly more likely to state that poll workers were highly qualified and they were more likely to believe in ballots' secrecy. Furthermore, the above findings that we were able to replicate, replicated with much lower statistical significance compared to the original research.

As most of the results obtained by Alvarez et al. (2013) did not replicate using a more accurate matching method, we invite the authors of the original paper to re-evaluate their findings and conclusions after using Genetic Matching.

Contribution Statement

Georgi and Helen worked collaboratively on all of the parts of the assignment, reviewing each others' work and making valuable contributions to the project.

References

Alvarez, R. M., Levin, I., Pomares, J., & Leiras, M. (2013). Voting made safe and Easy: The impact of e-voting on Citizen Perceptions. *Political Science Research and Methods*, *1*(1), 117–137. https://doi.org/10.1017/psrm.2013.2

Diamond, A., & Sekhon, J. S. (2013). Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics*, *95*(3), 932–945. https://doi.org/10.1162/rest_a_00318

King G. and Nielsen R. (2019). Why Propensity Scores Should Not Be Used for Matching. *Political Analysis*, 27, 4, Pp. 435-454. https://doi.org/10.1017/pan.2019.11