

**Replication for: Voting Made Safe and Easy: The Impact of e-voting on Citizen Perceptions**

Minerva University

CS130: Statistical Inference

Prof. Diamond

December 16, 2022

## **Replication for: Voting Made Safe and Easy: The Impact of e-voting on Citizen Perceptions**

### **1. Summary:**

Alvarez, R. Michael; Levin, Ines; Pomares, Julia; and Leiras, Marcelo in their 2015 paper, called *Voting Made Safe and Easy: The Impact of e-voting on Citizen Perceptions* focus exclusively on the citizen's perception of e-voting versus traditional voting. They claim that this perception directly influences voter satisfaction which in-turn dictates their opinion about the reliability of the elections.

They make this comparison by comparing opinions of citizens who casted their vote using e-voting and those who used traditional voting. They hope that this would help developing countries decide if they want to implement e-voting by shedding light on the consequences of doing so. The data was accessible on the Harvard dataverse and most variables were in binary format which made it fairly easy to work and do computation on. The data and the original study are linked in the references.

The study focuses on Salta, Argentina and uses matching specifically, propensity score matching since it wasn't possible to randomly assign voting technologies to polling stations and also it lets them compute individual level differences under each system. "We specify that the absolute difference between propensity scores for any pair of matched voters must be lower than 5 percent of the standard deviation of the estimated propensity score (this constraint is usually termed a 'caliper'. (Alvarez, 2015)

This is a causal inference, since we would like to know whether using e-voting or traditional voting causes the perception of voters to change which could influence voter satisfaction and the democracy overall. If e-voting is perceived better then there is a greater reason for developing countries to implement this technology.

The next part of the paper replicates propensity score matching as the original study did while part 3 using genetic matching and goes one step further to see how the results differ.

## **2. Replication:**

I replicated Table 2 from the original study which analyzed the balance before and after matching. The code for the table can be found in the appendix. The original study used propensity score matching to balance on the covariates. This reduced the bias that might be caused by confounding factors. For example, if we have our e-voting polling station in an area where the population is technologically savvy, then that would change their subjective evaluation of the voting process. This could include bias in our paper and skew our results.

Ideally we would reduce bias by using a randomized study. Which would achieve balance on all known and unknown confounding and would lead to the best causal inference. Unfortunately, randomization is not always possible. For example, in this study we have polling places put up in districts by the government and we are analyzing results from that data. The researchers in this case don't have the power to randomly assign voters to e-voting or traditional voting districts.

Matching can help us solve this problem since it achieves balance by comparing outcomes of units that received the treatment versus the ones that didn't.

If we see our results in the replicated table below, we can see that the standard mean difference between each of our covariates has decreased. This means that they are closer to being equal which would make our study free of known confounders. Ideally, the standard mean difference would be zero, like in variable 5 and 6 after matching. The closer the difference is to zero the more confident we can feel in our causal inference. Similarly we can see that for most of our variables the variance got lower after matching and p-value from the t-statistic got higher which is exactly what we were aiming for.

We can also see that we got less p-values for our binary variables, and that is because KS test p-value is identical to T-test p-value for binary variables, since for binary variables, the variance is the mean.

***** (V1) age.group *****		***** (V2) educ *****		***** (V3) tech *****		***** (V4) pol.info *****	
Before Matching		Before Matching		Before Matching		Before Matching	
After Matching		After Matching		After Matching		After Matching	
mean treatment.....	2.4063	mean treatment.....	4.6899	mean treatment.....	4.1874	mean treatment.....	1.4776
0.24063		0.46899		0.41874		0.14776	
mean control.....	2.337	mean control.....	4.153	mean control.....	3.9889	mean control.....	1.3304
0.23715		0.4638		0.42056		0.14494	
std mean diff.....	5.2169	std mean diff.....	23.717	std mean diff.....	12.362	std mean diff.....	18.521
0.26227		0.22905		-1.1362		0.35464	
mean raw eQQ diff.....	0.066519	mean raw eQQ diff.....	0.56098	mean raw eQQ diff.....	0.19512	mean raw eQQ diff.....	0.14412
0.051607		0.13632		0.025316		0.022395	
med raw eQQ diff.....	0	med raw eQQ diff.....	0	med raw eQQ diff.....	0	med raw eQQ diff.....	0
0		0		0		0	
max raw eQQ diff.....	1	max raw eQQ diff.....	2	max raw eQQ diff.....	1	max raw eQQ diff.....	1
1		2		1		1	
mean eCDF diff.....	0.013855	mean eCDF diff.....	0.063231	mean eCDF diff.....	0.03308	mean eCDF diff.....	0.036809
0.010321		0.01278		0.0042194		0.0055988	
med eCDF diff.....	0.006468	med eCDF diff.....	0.067363	med eCDF diff.....	0.030706	med eCDF diff.....	0.025819
0.011685		0.011685		0.0038948		0.0063291	
max eCDF diff.....	0.031965	max eCDF diff.....	0.13596	max eCDF diff.....	0.084463	max eCDF diff.....	0.095597
0.021422		0.03408		0.0087634		0.0097371	
var ratio (Tr/Co).....	1.0625	var ratio (Tr/Co).....	1.3905	var ratio (Tr/Co).....	1.0079	var ratio (Tr/Co).....	1.2709
1.0728		1.153		1.0274		1.1053	
T-test p-value.....	0.39417	T-test p-value.....	3.4716e-05	T-test p-value.....	0.046848	T-test p-value.....	0.0015439
0.074618		0.13632		0.40841		0.00081374	
KS Bootstrap p-value..	0.548	KS Bootstrap p-value..	< 2.22e-16	KS Bootstrap p-value..	0.014	KS Bootstrap p-value..	0.002
0.556		0.292		0.98		0.78	
KS Naive p-value.....	0.95468	KS Naive p-value.....	0.00014402	KS Naive p-value.....	0.050373	KS Naive p-value.....	0.017901
0.97251		0.58985		1		1	
KS Statistic.....	0.031965	KS Statistic.....	0.13596	KS Statistic.....	0.084463	KS Statistic.....	0.095597
0.021422		0.03408		0.0087634		0.0097371	

  

***** (V5) white.collar *****		***** (V6) not.full.time *****		***** (V7) male *****	
Before Matching		Before Matching		Before Matching	
After Matching		After Matching		After Matching	
mean treatment.....	0.29851	mean treatment.....	0.27197	mean treatment.....	0.51078
0.29851		0.27197		0.51078	
mean control.....	0.28381	mean control.....	0.32816	mean control.....	0.50111
0.29851		0.27197		0.50746	
std mean diff.....	3.2084	std mean diff.....	-12.616	std mean diff.....	1.933
0		0		0.66295	
mean raw eQQ diff.....	0.013304	mean raw eQQ diff.....	0.05765	mean raw eQQ diff.....	0.0088692
0		0		0.0029211	
med raw eQQ diff.....	0	med raw eQQ diff.....	0	med raw eQQ diff.....	0
0		0		0	
max raw eQQ diff.....	1	max raw eQQ diff.....	1	max raw eQQ diff.....	1
0		0		1	
mean eCDF diff.....	0.0073469	mean eCDF diff.....	0.028093	mean eCDF diff.....	0.0048354
0		0		0.0014606	
med eCDF diff.....	0.0073469	med eCDF diff.....	0.028093	med eCDF diff.....	0.0048354
0		0		0.0014606	
max eCDF diff.....	0.014694	max eCDF diff.....	0.056186	max eCDF diff.....	0.0096708
0		0		0.0029211	
var ratio (Tr/Co).....	1.0296	var ratio (Tr/Co).....	0.89759	var ratio (Tr/Co).....	0.99898
1		1		0.99976	
T-test p-value.....	0.60342	T-test p-value.....	0.049884	T-test p-value.....	0.75632
1		1		0.15713	

Replication table for propensity score Matching from Table 2 of the original study.

### 3. Extension:

I am changing the matching method used in the paper to genetic matching. While propensity score matching is not as robust, using genetic matching is a much better option since instead of calculating propensity scores of each individual, the method provides a weight to each variable and then uses that to achieve optimal balance after matching. “The Genetic Matching algorithm searches amongst a range of distance metrics to find the particular measure that optimizes post-matching covariate balance. Each potential distance metric considered corresponds to a particular assignment of weights  $W$  for all matching variables. The algorithm weights each variable according to its relative importance for achieving the best overall balance.”

(Diamond-Sekhon, 2012)

Genetic matching is also superior because it “can reduce biases in case of great imbalances and are not sensitive to model specification.” (Donze, 2011)

```
> summary(mout_eselect.cand)
Estimate... 0.28038
AI SE..... 0.033664
T-stat..... 8.3287
p.val..... < 2.22e-16

Original number of observations..... 1054
Original number of treated obs..... 603
Matched number of observations..... 603
Matched number of observations (unweighted). 1010

> summary(mout_conf.secret)
Estimate... -0.089055
AI SE..... 0.031408
T-stat..... -2.8354
p.val..... 0.0045768

Original number of observations..... 1054
Original number of treated obs..... 603
Matched number of observations..... 603
Matched number of observations (unweighted). 1010

> summary(mout_sure.counted)
Estimate... 0.089552
AI SE..... 0.029406
T-stat..... 3.0454
p.val..... 0.0023236

Original number of observations..... 1054
Original number of treated obs..... 603
Matched number of observations..... 603
Matched number of observations (unweighted). 1010

> summary(mout_agree.evoting)
Estimate... 0.16645
AI SE..... 0.033016
T-stat..... 5.0414
p.val..... 4.6225e-07

Original number of observations..... 1054
Original number of treated obs..... 603
Matched number of observations..... 603
Matched number of observations (unweighted). 1010

> summary(mout_speed)
Estimate... 0.036097
AI SE..... 0.030832
T-stat..... 1.1708
p.val..... 0.2417

Original number of observations..... 1054
Original number of treated obs..... 603
Matched number of observations..... 603
Matched number of observations (unweighted). 1010

> summary(mout_how.clean)
Estimate... 0.15945
AI SE..... 0.039595
T-stat..... 4.0271
p.val..... 5.6466e-05

Original number of observations..... 1054
Original number of treated obs..... 603
Matched number of observations..... 603
Matched number of observations (unweighted). 1010

> summary(mout_eval.voting)
Estimate... 0.28242
AI SE..... 0.03654
T-stat..... 7.7291
p.val..... 1.088e-14

Original number of observations..... 1054
Original number of treated obs..... 603
Matched number of observations..... 603
Matched number of observations (unweighted). 1010

> summary(mout_easy.voting)
Estimate... 0.19486
AI SE..... 0.031422
T-stat..... 6.2014
p.val..... 5.5948e-10

Original number of observations..... 1054
Original number of treated obs..... 603
Matched number of observations..... 603
Matched number of observations (unweighted). 1027
```

Results from genetic matching of each interest variable.

Within our results we can see that all of our treated observations got matched but a higher number of unweighted matched observations indicate that we have multiple ties from each treatment, and thus that treatment was repeated. Multiple ties mean that each tie is weighted less

and the combined weight of all ties add upto one, this happens when there are two similar controls to match with.

Our p-values for most variables are fairly low indicating that our results are statistically significant.

## Appendix:

```
#Replication:
```

```
load("datamatch.Rdata") #set working directly to downloads
```

```
summary(datamatch)
```

```
data<- na.omit(datamatch)
```

```
#matching:
```

```
library(Matching)
```

```
covariates <- data.frame(data$age.group, data$educ, data$white.collar, data$not.full.time,  
data$male, data$tech, data$pol.info)
```

```
EV <- data[,2]
```

```
glm1=
```

```
glm(EV~eval.voting+easy.voting+sure.counted+conf.secret+how.clean+speed+agree.evoting+ese  
lect.cand, data = data)
```

```
datamatched<-Match(Y=glm1$fitted,Tr=EV, X=covariates)
```

```
mb <-
```

```
MatchBalance(EV~eval.voting+easy.voting+sure.counted+conf.secret+how.clean+speed+agree.e  
voting+eselect.cand, data = data, match.out=datamatched, nboots=500)
```

```
datamatched<-Match(Tr=EV, X=covariates)
```

```
capable.auth <- data$capable.auth
```

```
mout_capable.auth<-Match(Y=capable.auth,Tr=EV, X=covariates)
```

```
eval.voting <- data$eval.voting
```

```
mout_eval.voting<-Match(Y=eval.voting,Tr=EV, X=covariates)
```

```
easy.voting <- data$easy.voting
```

```
mout_easy.voting<-Match(Y=easy.voting,Tr=EV, X=covariates)
```

```
sure.counted <-data$sure.counted
```

```
mout_sure.counted<-Match(Y=sure.counted,Tr=EV, X=covariates)
```

```
conf.secret<- data$conf.secret
```

```
mout_conf.secret<-Match(Y=conf.secret,Tr=EV, X=covariates)
```

```
how.clean<- data$how.clean
```

```
mout_how.clean<-Match(Y=how.clean,Tr=EV, X=covariates)
```

```
speed<- data$speed
```

```
mout_speed<-Match(Y=speed,Tr=EV, X=covariates)
```

```
agree.evoting<- data$agree.evoting
```

```
mout_agree.evoting<-Match(Y=agree.evoting,Tr=EV, X=covariates)
```

```
eselect.cand<- data$eselect.cand
```

```
mout_eselect.cand<-Match(Y=eselect.cand,Tr=EV, X=covariates)
```

```
Extention GenMatch:
```

```
covariates <- data.frame( data$age.group, data$educ, data$white.collar, data$not.full.time,  
data$male, data$tech, data$pol.info)
```

```
genout <- GenMatch(X = covariates, Tr = EV,pop.size=100,max.generation=50,  
wait.generations=4)
```

```
# DEP VAR 1
```

```
eval.voting <- data$eval.voting
```

```
mout_eval.voting <- Match(Y = eval.voting,
```

```
                  X = covariates,
```

```
                  Tr = EV,
```

```
                  Weight.matrix = genout)
```

```
summary(mout_eval.voting)
```

```
# DEP VAR 2
```

```
easy.voting <- data$easy.voting
```

```
mout_eval.voting <- Match(Y = easy.voting,
```

```
                  X = covariates,
```

```
                  Tr = EV,
```

```
                  Weight.matrix = genout)
```

```
summary(mout_easy.voting)
```

```
# DEP VAR 3
```

```
capable.auth <- data$capable.auth
```

```
mout_capable.auth <- Match(Y = capable.auth,
```

```
                  X = covariates,
```

```
                  Tr = EV,
```

```
Weight.matrix = genout)  
summary(mout_capable.auth)
```

```
# DEP VAR 4  
sure.counted <- data$sure.counted  
mout_sure.counted <- Match(Y = sure.counted,  
X = covariates,  
Tr = EV,  
Weight.matrix = genout)  
summary(mout_sure.counted)
```

```
# DEP VAR 5  
conf.secret <- data$conf.secret  
mout_conf.secret <- Match(Y = conf.secret,  
X = covariates,  
Tr = EV,  
Weight.matrix = genout)  
summary(mout_conf.secret)
```

```
# DEP VAR 6  
how.clean <- data$how.clean  
mout_how.clean <- Match(Y = how.clean,  
X = covariates,  
Tr = EV,  
Weight.matrix = genout)  
summary(mout_how.clean)
```

```
# DEP VAR 7  
speed <- data$speed  
mout_speed <- Match(Y = speed,  
X = covariates,  
Tr = EV,  
Weight.matrix = genout)  
summary(mout_speed)
```

```
# DEP VAR 8  
agree.evoting <- data$agree.evoting  
mout_agree.evoting <- Match(Y = agree.evoting,  
X = covariates,  
Tr = EV,
```



```
Weight.matrix = genout)  
summary(mout_agree.evoting)
```

```
# DEP VAR 9  
eselect.cand <- data$eselect.cand  
mout_eselect.cand <- Match(Y = eselect.cand,  
X = covariates,  
Tr = EV,  
Weight.matrix = genout)  
summary(mout_eselect.cand)
```

## 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>
- Alvarez, R. M., Levin, I., Pomares, J., & Leiras, M. (2015). *Replication data for: Voting Made Safe and Easy: The Impact of e-voting on Citizen Perceptions* [Data set]. Harvard Dataverse. <https://doi.org/10.7910/DVN/24896>
- Matching Recording.mov*. (n.d.). Google Docs. Retrieved December 16, 2022, from [https://drive.google.com/file/u/1/d/0B3ECwzWL2AHgbWFLTm9saW1oOXc/view?usp=sharing&resourcekey=0-JazdynM1OA6stUfsTgY5Kw&usp=embed\\_facebook](https://drive.google.com/file/u/1/d/0B3ECwzWL2AHgbWFLTm9saW1oOXc/view?usp=sharing&resourcekey=0-JazdynM1OA6stUfsTgY5Kw&usp=embed_facebook)
- 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](https://doi.org/10.1162/REST_a_00318)