Memo

To:	Argentina Election Commission
From:	Rita Kurban
Date:	12.06.2017
Re:	Decision memo that replicates and extends the findings of the paper “Voting Made Safe and Easy: The Impact of e-voting on Citizen Perceptions ” (Alvarez, Levin, Pomares, Leiras, 2015).

Executive Summary

The purpose of this memorandum is to evaluate the impact of the electronic voting on the voter experience and give a recommendation on whether it’s worth implementing in Argentina. This topic gets more and more relevant as new sophisticated technologies are being developed. Several developing countries including Brazil, India, the Philippines, and Venezuela have recently implemented e-voting in efforts to improve the legitimacy of their election process. However, only some researchers, mostly from the United States, have thoroughly investigated voter perceptions and evaluations of e-voting relative to traditional voting (Nichols & Strizek, 1995; Kohno, Stubblefield, Rubin & Wallach, 2004). In this memo, I replicate the findings of the paper “Voting Made Safe and Easy: The Impact of e-voting on Citizen Perceptions ” (Alvarez, Levin, Pomares, Leiras, 2015) which uses data from the field experiment in Salta, Argentina. The authors employed propensity score matching and sensitivity tests to measure the causal effect of e-voting on the voter experience. I replicated and extended this paper by applying multivariate and genetic matching. I recommend replacing traditional voting technologies with e-voting in Argentina. However, it’s worth mentioning that new technologies raise concerns about ballot secrecy which should be further investigated.

Dataset

The data I used for the replication comes from a questionnaire that was administered to 1,502 voters from 36 different polling stations. From the Table 1, you can see that there is a statistically significant difference in the evaluations of the voting experience for the two groups for 8 out of 9 outcome variables (excluding the speed of voting process).

	N	All Voters (%)	E-Voters (%)	Traditional Voters (%)	Diff.	p-value
Qualification of poll workers	1441	81.33241	84.99400	76.31579	8.678208	4.009402e-05
Evaluation of voting experience	1486	36.27187	46.64391	21.25206	25.391854	2.457725e-23
Ease of voting procedure	1495	24.74916	34.01361	11.41925	22.594356	4.349730e-23
Sure vote was counted	1444	82.34072	86.19883	76.74024	9.458593	5.066539e-06
Confident ballot secret	1455	80.06873	77.14959	84.15842	-7.008828	1.226635e-03
Elections in Salta are clean	1303	50.65234	57.69231	40.98361	16.708701	3.620959e-09
Speed of voting process	1468	82.90191	84.30699	80.84034	3.466651	9.662635e-02
Agree substitute TV by EV	1430	75.24476	84.08551	62.58503	21.500477	3.348404e-20
Select candidates electronically	1405	71.53025	83.99519	53.48432	30.510866	2.695479e-35

Table 1: Summary statistics for the original dataset

This table gives us an understanding of what kind of data we have and the distribution of responses between two groups. However, it doesn't guarantee reliable estimates of the causal effect of the e-voting. The observed results can be explained not only by the differences in the voting method but also by the selection bias introduced by the non-random allocation of e-voting devices to polling places. It is reasonable to assume that people who used e-voting were younger, better-educated, or wealthier than traditional voters. Especially given the fact that e-voting was introduced in places where voters generally had a higher socio-economic status. To alleviate these potential biases, I'll use matching procedures.

Matching

To start with, I implemented propensity score matching (PSM). This method is supposedly the easiest way to find reliable matches as it projects a large number of covariates to a scalar propensity score and applies a single model to produce an estimate. However, it is not always effective and in some cases even increases imbalance, model dependence, and bias. (King & Nielsen, 2016). Even after playing around with multiple parameters and introducing a caliper, I, as well as the authors of the original paper, didn't manage to achieve a balance on all the covariates:

```
Before Matching Minimum p.value: < 2.22e-16
Variable Name(s): educ  Number(s): 2

After Matching Minimum p.value: 0.12
Variable Name(s): age.group educ  Number(s): 1 2
```

The problem of PSM is that it tries to approximate a completely randomized experiment, rather than a more effective fully blocked randomized experiment. To address this issue, I implemented multivariate matching which is considered to be more robust as it matches on all the independent variables in addition to the propensity scores. Using this method, I managed to get significant improvement and achieved a perfect balance on all but one variable:

```
Before Matching Minimum p.value: < 2.22e-16
Variable Name(s): educ  Number(s): 2

After Matching Minimum p.value: 0.00081374
Variable Name(s): pol.info  Number(s): 4
```

Lastly, I implemented genetic matching which uses an evolutionary search algorithm to determine the weight of each covariate. This method performed significantly better than the previous two and achieved perfect balance on all covariates. To demonstrate it, I created a table that presents the balance on all covariates before and after the genetic matching which is not possible to achieve with the PSM:

	Before Matching				After matching			
	EV	TV	Diff.	p-value	EV	TV	Diff.	p-value
Age group (1-5)	2.406302	2.337029	0.0692730	0.55600000	1.706366	1.706366	0	1
Education (1-8)	4.689884	4.152993	0.5368906	0.00000000	4.644764	4.644764	0	1
White collar (%)	29.850746	28.381375	1.4693715	0.65232145	17.043121	17.043121	0	1
Not full time worker (%)	27.197347	32.815965	-5.6186179	0.05627993	22.792608	22.792608	0	1
Male (%)	51.077944	50.110865	0.9670789	0.80378565	46.406571	46.406571	0	1
Technology count (1-6)	4.187396	3.988914	0.1984828	0.01200000	4.490760	4.490760	0	1
Political information (1-4)	1.477612	1.330377	0.1472350	0.00000000	1.098563	1.098563	0	1

Table 2. Balance statistics

Causal Effects

Table 3 presents the responses of e-voters and traditional voters before and after the genetic matching. Even though I eliminated the observable differences between e-voters and regular voters, differences in attitudes toward voting technologies as well as voter satisfaction are still relatively large and statistically significant for all the outcome variables:

	Before Matching				After matching			
	E-Voting (%)	Traditional Voting (%)	Diff.	p-value	E-Voting (%)	Traditional Voting (%)	Diff.	p-value
Elections in Salta are clean	60.03317	42.35033	17.682835	1.850974e-08	59.34292	32.23819	27.104723	3.609611e-17
Evaluation of voting experience	47.92703	23.50333	24.423706	8.954108e-16	49.07598	22.17659	26.899384	3.374914e-18
Select candidates electronically	84.07960	52.99335	31.086254	1.018295e-27	86.03696	61.39630	24.640657	4.614492e-18
Ease of voting procedure	32.00663	13.74723	18.259405	1.235051e-11	36.75565	14.37372	22.381930	2.140056e-15
Sure vote was counted	87.06468	77.82705	9.237626	1.055108e-04	86.24230	68.78850	17.453799	1.140237e-10
Agree substitute TV by EV	83.58209	62.74945	20.832644	2.483859e-14	81.93018	65.91376	16.016427	1.916627e-08
Qualification of poll workers	83.91376	75.60976	8.304008	1.040869e-03	83.57290	71.66324	11.909651	1.176339e-05
Speed of voting process	84.24544	80.93126	3.314176	1.838429e-01	83.36756	75.56468	7.802875	3.336376e-03
Confident ballot secret	77.28027	83.14856	-5.868293	2.321331e-02	71.86858	81.51951	-9.650924	4.897614e-04

Table 3. Causal Effect of e-voting

Overall, the results from the Table 3 correspond to the outcomes reported by the authors of the original paper. For example, they also found that e-voters are nearly 25% more likely than traditional voters to characterize the voting experience as 'very good' and proved that the overall perception of cleanness of elections is especially low for traditional voters. The only aspect of the voting experience that e-voters evaluate more negatively is ballot secrecy: e-voters are at least 9% less likely to say that they are confident that the confidentiality of their votes was preserved. What is more, I found a statistically significant difference in the way traditional and e-voters characterize the speed of the voting process which the original paper failed to do.

Sensitivity Analysis

To control for any differences in unobservable attributes that might skew the results mentioned above, I conducted a Rosenbaum bounds sensitivity analysis with matched data and constructed the following table:

	Qualification of poll workers		**Evaluation of voting experience**		**Ease of voting procedure**		**Sure vote was counted**		**Confident ballot secret**		**Elections in Salta are clean**		**Speed of voting process**		**Agree substitute TV by EV**		**Select candidates electronically**	
Gamma	**U**	**L**	**U**	**L**	**U**	**L**	**U**	**L**	**U**	**L**	**U**	**L**	**U**	**L**	**U**	**L**	**U**	**L**
1	2e-04	2e-04	0	0	0	0	0	0	2e-04	2e-04	0	0	0.0017	0.0017	0	0	0	0
1.1	0	0.0015	0	0	0	0	0	0	0	0.0015	0	0	2e-04	0.0102	0	0	0	0
1.2	0	0.0079	0	0	0	0	0	0	0	0.0079	0	0	0	0.0391	0	0	0	0
1.3	0	0.0284	0	0	0	0	0	0	0	0.0284	0	0	0	0.1053	0	0	0	0
1.4	0	0.0757	0	0	0	0	0	0	0	0.0757	0	0	0	0.217	0	2e-04	0	0
1.5	0	0.1589	0	0	0	0	0	0	0	0.1589	0	0	0	0.3645	0	0.0011	0	0
1.6	0	0.2768	0	0	0	0	0	1e-04	0	0.2768	0	0	0	0.5243	0	0.0041	0	0
1.7	0	0.4164	0	0	0	0	0	5e-04	0	0.4164	0	0	0	0.6714	0	0.0119	0	0
1.8	0	0.5589	0	0	0	1e-04	0	0.0016	0	0.5589	0	1e-04	0	0.7896	0	0.0286	0	0
1.9	0	0.6873	0	1e-04	0	2e-04	0	0.0042	0	0.6873	0	3e-04	0	0.8744	0	0.0589	0	0
2	0	0.7913	0	3e-04	0	7e-04	0	0.0097	0	0.7913	0	0.001	0	0.9296	0	0.1067	0	0
2.1	0	0.8684	0	0.0011	0	0.0018	0	0.0197	0	0.8684	0	0.003	0	0.9627	0	0.1733	0	0
2.2	0	0.9211	0	0.0029	0	0.0043	0	0.0364	0	0.9211	0	0.0076	0	0.9813	0	0.257	0	1e-04
2.3	0	0.9549	0	0.0069	0	0.0091	0	0.0615	0	0.9549	0	0.0171	0	0.991	0	0.353	0	3e-04
2.4	0	0.9753	0	0.0147	0	0.0177	0	0.0966	0	0.9753	0	0.0341	0	0.9958	0	0.4548	0	8e-04
2.5	0	0.987	0	0.0283	0	0.0314	0	0.1422	0	0.987	0	0.0616	0	0.9982	0	0.5553	0	0.0017
2.6	0	0.9933	0	0.0498	0	0.0519	0	0.1979	0	0.9933	0	0.1018	0	0.9992	0	0.6488	0	0.0034
2.7	0	0.9967	0	0.0811	0	0.0804	0	0.2622	0	0.9967	0	0.1558	0	0.9997	0	0.731	0	0.0063
2.8	0	0.9984	0	0.1235	0	0.1178	0	0.3328	0	0.9984	0	0.223	0	0.9999	0	0.7999	0	0.011
2.9	0	0.9993	0	0.177	0	0.164	0	0.4071	0	0.9993	0	0.3011	0	0.9999	0	0.8552	0	0.0181
3	0	0.9997	0	0.2408	0	0.2184	0	0.4823	0	0.9997	0	0.3864	0	1	0	0.898	0	0.0283

Table 4. Sensitivity Analysis. This table gives Rosenbaum bounds for p-values of the effect of e-voting on voters' satisfaction with the voting experience. 'L' stands for lower bounds, while 'U' indicates upper bounds.

Lower (L) and upper (U) bound columns indicate the p-values of a test of the difference between responses reported by e-voters and traditional voters for a series of Γ values. To interpret these results, consider the outcome variable titled "Speed of voting Process." We see that for an increase of 0.3 in Γ the p-value increases to 0.1053, which is above the usual 0.05 threshold. It suggests that even a reasonably small unobserved difference in a covariate might change our inference. However, for the rest of the variables, the levels of Γ are much higher — more than 1.8 for 6 out of 9 variables. These results indicate that most of my findings are robust to the presence of unobservable covariates, which further reinforces the idea that e-voting is associated with a positive voting experience and supports replacing traditional voting with electronic voting.

Recommendation

One of the main concerns of observational studies is that it is impossible to directly compare the outcomes for those who received treatment and those who received control, as differences could be attributed to inherent differences between the groups. Genetic matching is one of the most sophisticated methods that copes with this problem. This approach gave me an opportunity to achieve perfect balance on the covariates and improve the results of the original paper. However, the estimates of treatment effect based on genetic matching are unbiased only in case there are no unobserved confounders. To account for this concern, I used the sensitivity analysis. The results of this replication suggest that voters who used the e-voting evaluated the voting procedure more favorably than people who used the traditional system for 8 out of 9 outcomes including the cleanliness of elections, ease of the voting procedure, and the overall experience. In case of Salta, the new process raises some concerns about ballot secrecy. Despite the fact that this negative effect is small, it should be subject to larger scrutiny. Overall, this study strongly supports e-voting and encourages the Election Commission to implement it in other parts of Argentina.

Technical Appendix

The code can also be found under:

https://gist.github.com/ritakurban/3407599f9badd789ed41db76364928dc

```
# Installing all the necessary libraries
library(Matching)
library(rbounds)
library(tableHTML)

# Loading the data
load("datamatch.Rdata")

# Outcome variables
outcomes <- datamatch[10:18]

# Variable names
names <- c("Qualification of poll workers","Evaluation of voting
experience","Ease of voting procedure","Sure vote was
counted","Confident ballot secret","Elections in Salta are
clean","Speed of voting process","Agree substitute TV by EV","Select
candidates electronically")

# Number of outcomes
number <- dim(outcomes)[2]

#___________________________ Table 1 __________________________#

# The first table visualizes the results of the survey

# Creating the matrix
tab1 <- matrix(NA, nrow = number, ncol = 6)
rownames(tab1) <- names
colnames(tab1) <- c("N", "All Voters (%)", "E-Voters (%)",
"Traditional Voters (%)", "Diff.", "p-value")

for (i in 1:number) {
  # The number of responses
  tab1[i, 1] <- length(outcomes[, i])
  # Proportion of all voters
  tab1[i, 2] <- prop.table(table(outcomes[, i]))[2] * 100
  # Proportion of electronic and traditional separately
  tab1[i, 3:4] <- rev(prop.table(table(outcomes[, i], datamatch$EV),
2)[2, ]) * 100
```

```
  # The difference
  tab1[i, 5] <- tab1[i, 3] - tab1[i, 4]
  # Test of difference in proportions to find the p-value
  tab1[i, 6] <- prop.test(table(outcomes[, i], datamatch$EV)[2, ], n =
apply(table(outcomes[, i], datamatch$EV), 2, sum))$p.value
}
tab1 <- tab1[rev(order(tab1[, "Diff."])), ]
outcomes[, i]
datamatch <- na.omit(datamatch)

#_____________________ Table 2, pre-matching ______________________#

# This part of the second table calculates the differences in the
covariates before matching

EV <- datamatch[2]

covar <- datamatch[c("age.group", "educ", "white.collar",
"not.full.time", "male", "tech", "pol.info")]
covar.names <- c("Age group (1-5)", "Education (1-8)", "White collar
(%)", "Not full time worker (%)", "Male (%)", "Technology count
(1-6)", "Political information (1-4)")

number.covar <- dim(covar)[2]

# Creating the table
tab2.pre <- matrix(NA, nrow = number.covar, ncol = 4)
rownames(tab2.pre) <- covar.names
colnames(tab2.pre) <- c("EV", "TV", "Diff.", "p-value")

# Calculating the values for electronic and traditional voters and the
difference between them
tab2.pre[, 1:2] <- cbind(apply(covar[EV == 1,], 2, mean),
apply(covar[EV == 0,], 2, mean))
tab2.pre[, 3] <- tab2.pre[, 1] - tab2.pre[, 2]

# Finding p-values of Kolmogorov-Smirnov tests for ordinal variables
for (i in c(1, 2, 6, 7)){
  tab2.pre[i, 4] <- ks.boot(covar[, i][EV == 1], covar[, i][EV == 0],
nboots = 500)$ks.boot.pvalue
}
# Finding p-values using a difference in proportions test for binary
variables
for (i in c(3, 4, 5)){
  tab2.pre[i, 4] <- prop.test(table(covar[, i], EV$EV), n =
apply(table(covar[,i],EV$EV),2, sum))$p.value}
```

```
#_____________________ Table 3, pre-matching ____________________#

# This table presents the naive initial findings based on the
unmatched dataset

outcomes.pre <- datamatch[10:18]

# Creating the table
tab3.pre <- matrix(NA,nrow = number,ncol = 4)
rownames(tab3.pre) <- names
colnames(tab3.pre) <- c("E-Voting (%)", "Traditional Voting (%)",
"Diff.", "p-value")

for (i in 1:number) {
  # The proportions of electronic and traditional voters
  tab3.pre[i, 1:2] <-
rev(prop.table(table(outcomes.pre[,i],datamatch$EV),2)[2,])*100
  # The difference in the outcomes
  tab3.pre[i, 3] <- tab3.pre[i, 1] - tab3.pre[i, 2]
  # P-value using the difference in proportions test
  tab3.pre[i, 4] <- prop.test(table(outcomes.pre[, i],
datamatch$EV)[2, ], n = apply(table(outcomes.pre[, i], datamatch$EV),
2, sum))$p.value
}

#_______________________ Matching (with Match) _____________________#

# Propensity Score Matching (PMS)

# Estimating a propensity score
prop_score  <- glm(EV ~ age.group + educ + tech + + pol.info +
white.collar + not.full.time + male, data=datamatch)
X  <- prop_score$fitted
Tr  <- datamatch$EV

rr  <- Match(Y = datamatch$how.clean, Tr=Tr, X=X, estimand = "ATT",
caliper = 0.05);
summary(rr)
# Checking the balance
mb  <- MatchBalance(EV ~ age.group + educ + tech + + pol.info +
white.collar + not.full.time + male, data = datamatch,match.out=rr,
nboots=500)

# Multivariate matching
cov_names <- c("age.group", "educ", "tech", "pol.info",
"white.collar", "not.full.time", "male" )
```

```
all_covs <- datamatch[, cov_names]

rr1  <- Match(Tr=Tr, X=all_covs, estimand = "ATT", caliper = 0.05);
summary(rr1)
mb1  <- MatchBalance(EV ~ age.group + educ + tech + + pol.info +
white.collar + not.full.time + male, data=datamatch, match.out=rr1,
nboots=500)

# Genetic Matching
# Finding the weights
genout <- GenMatch(Tr=Tr, X=all_covs, estimand="ATT", caliper=0.05,
                   pop.size=200, max.generations=100,
wait.generations=10)
mout <- Match(Y = datamatch$conf.secret, Tr=Tr, X=all_covs,
estimand="ATT", caliper=0.05, Weight.matrix=genout)
mb_gen <- MatchBalance(EV ~ age.group + educ + tech + + pol.info +
white.collar + not.full.time + male, data=datamatch,
                       match.out=mout, nboots=500)

#______________________ Sensitivity Analysis ______________________#
psens(mout, Gamma=3, GammaInc=.1)

# Calculating the sensitivity for the first outcome
sensitivity <- data.frame(psens(mout, Gamma=3, GammaInc=.1))
sensitivity <- sensitivity[,5:7]

# Calculating the bounds for all the outcome variables with the help
of a loop
for (i in 11:18){
  mout <- Match(Y = datamatch[,i], Tr=Tr, X=all_covs, estimand="ATT",
caliper=0.05, Weight.matrix=genout)
  temporal <- data.frame(psens(mout, Gamma=3, GammaInc=.1))
  temporal <- temporal[,6:7]
  sensitivity = data.frame(sensitivity,temporal)
}
colnames(sensitivity) = c("Gamma","U", "L","U", "L","U", "L", "U",
"L","U", "L", "U", "L","U", "L","U", "L","U", "L")

# Creating a fancy table
n <- c("","Qualification of poll workers","Evaluation of voting
experience","Ease of voting procedure","Sure vote was
counted","Confident ballot secret","Elections in Salta are
clean","Speed of voting process","Agree substitute TV by EV","Select
candidates electronically")
tableHTML(sensitivity,
          rownames = FALSE,
```

```
          second_header = list(c(1,2, 2, 2, 2, 2, 2, 2, 2, 2), n),
          widths = c(rep(1, 19)),
          theme = "scientific")

# Creating a matched dataset
matched.treat <- datamatch[genout$matches[,1], ]
matched.control <- datamatch[genout$matches[,2], ]
datamatched <- rbind(matched.treat, matched.control)


#______________________ Table 2, post-matching ____________________#

# Adding matched dataset to the original table to enable comparison
EV.post <- datamatched[2]

covar.post <- datamatched[, c("age.group", "educ", "white.collar",
"not.full.time", "male", "tech", "pol.info")]
tab2.post <- matrix(NA, nrow = number.covar, ncol = 4)
rownames(tab2.post) <- covar.names
colnames(tab2.post) <- c("EV", "TV", "Diff.", "p-value")

tab2.post[, 1:2] <- cbind(apply(covar.post[EV.post == 1, ], 2, mean),
apply(covar.post[EV.post == 0,], 2, mean))
tab2.post[, 3] <- tab2.post[, 1] - tab2.post[, 2]
for (i in c(1, 2, 6 , 7)){
  tab2.post[i,
4]<-ks.boot(covar.post[,i][EV.post==1],covar.post[,i][EV.post==0],
nboots = 500)$ks.boot.pvalue
}
for (i in c(3, 4, 5)){
  tab2.post[i, 4] <- prop.test(table(covar.post[, i], EV.post$EV), n =
apply(table(covar.post[, i], EV.post$EV),2 , sum))$p.value
}
# Creating the final table
tab2 <- cbind(tab2.pre, tab2.post)
tab2[3:5, c(1:3, 5:7)] <- tab2[3:5, c(1:3, 5:7)] * 100


#____________________ Table 3, post-matching ______________________#

# The outcomes for the matched dataset
outcomes.post <- datamatched[10:18]
tab3.post <- matrix(NA, nrow = number, ncol = 4)
rownames(tab3.post) <- names
colnames(tab3.post) <- c("E-Voting (%)", "Traditional Voting (%)",
"Diff.", "p-value")
```

```
for (i in 1:number) {
  tab3.post[i, 1:2] <- rev(prop.table(table(outcomes.post[, i],
datamatched$EV), 2)[2, ]) * 100
  tab3.post[i, 3] <- tab3.post[i, 1] - tab3.post[i, 2]
  tab3.post[i, 4] <- prop.test(table(outcomes.post[, i],
datamatched$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[, 7])), ]
```

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.*

Keele, L. (2010). An overview of rbounds: An R package for Rosenbaum bounds sensitivity analysis with matched data. *White Paper. Columbus, OH, 1-15.*

Kohno, T., Stubblefield, A., Rubin, A. D., & Wallach, D. S. (2004, May). Analysis of an electronic voting system. *In Security and Privacy, 2004. Proceedings. 2004 IEEE Symposium on (pp. 27-40). IEEE.*

Nichols, S. M., & Strizek, G. A. (1995). Electronic voting machines and ballot roll-off. *American Politics Quarterly, 23(3), 300-318.*

Replication data for: Voting Made Safe and Easy: The Impact of e-voting on Citizen Perceptions - Political Science Research and Methods (PSRM) Dataverse. (2017). *Dataverse.harvard.edu.*