

Impact of voting technologies on voting experience in Argentina

Final Project Decision Memo

CLASS CODE: CS112

PhD. Alexis J Diamond

Minerva Schools at KGI

Loan, Chau, Chantsaldiimaa, Kareem

April 2020

Executive Summary	3
Background	3
Genetic Matching	5
Extension of Table 2 in the original paper	6
Extension of Table 3 in the original paper	7
Conclusion	9
References	10
Appendix	11

To : The Election Commission of Argentina
From : Loan Nguyen, Chau Le, Chantsaldiimaa Lkhagvatogtokh, Kareem Alsayed
Date : April 22nd, 2020
Re : The impact of voting technologies on voting experience in Argentina

Executive Summary

Alvarez, Levin, Pomares, and Leiras's paper "Voting Made Safe and Easy: The Impact of e-voting on Citizen Perceptions" studied the perceptions and evaluations of voters in Salta, Argentina on e-voting compared to traditional voting. Using Propensity Score Matching, the study found that Argentine voters supported replacing traditional voting technologies with e-voting as they perceived the new technology as easier to use and more likely to record votes as intended. However, there remained concerns about ballot secrecy with e-voting. In order to extend the results of the study, our group used the Genetic Matching method, and we obtained a higher degree of balance between treatment and control groups on almost all the covariates. The Genetic Matching results are consistent with the Alvarez et al. results that voters in Salta, Argentina rated e-voting more favorably than traditional voting in terms of the overall experience, ease of use and speed, but still doubted the ballot secrecy. This additional finding suggests that the Government of Argentina should take actions to improve voters' confidence in the ballot secrecy of e-voting if they decide to fully replace the traditional voting with this new technology.

Background

There has long been a debate about the usage of emerging technology in the voting process. Despite the potential outcome voting technology can bring to the national elections, little effort has been put into evaluating the impact of voting technology on the transparency and legitimacy of elections. Alvarez et al. (2013) studied the perceptions and evaluations of voters in Salta, Argentina on e-voting compared to traditional voting.

The study used a quasi-experimental design to collect data and matching methods to generate statistical inference about the effect of voting technology on voters' opinions. First, the propensity score matching is utilized to balance the relevant covariates, followed by the causal effect estimation between the matched e-voters and traditional voters. After that, a post-matching regression adjustment is implemented to control for other differences in the observed covariates.

To further account for unobservable covariates, the sensitivity testing was the last step to evaluate the robustness of the paper's results.

Alvarez et al. (2013) used propensity score caliper matching to match e-voters with traditional voters. Propensity scores were estimated with a caliper of 0.05, which means that the absolute difference between propensity scores for matched voters must be equal to or less than 5% of the standard deviation of the estimated propensity score. The caliper is meant to limit the possible difference between two data points, thus creating more balance between covariates.

The paper matched e-voters and traditional voters on 7 independent variables: age, education, while collar, not full-time worker, gender, technology count, and political information. However, considering the potential relationship between variables and the goal to maintain a high balance, the matching procedure ran across 16 covariates including quadratic and interaction terms: `age.group`, `I(age.group^2)`, `I(age.group^3)`, `I(age.group:educ)`, `I(age.group:tech)`, `educ`, `I(educ^2)`, `tech`, `I(tech^2)`, `pol.info`, `I(educ:pol.info)`, `I(age.group:pol.info)`, `I(tech:pol.info)`, `white.collar`, `not.full.time`, `male`.

As a result of the matching procedure, 27 control units and 284 treated units (~33% of all treated) were dropped, which was a significant number (Figure 1). Above all, the balance statistics before and after matching improved significantly as there was no longer any big difference between the treated and control groups after matching (Table 1).

Sample sizes:

	Control	Treated
All	609	866
Matched	582	582
Unmatched	27	284
Discarded	0	0

Figure 1. Sample sizes of control (traditional voting) and treated (e-voting) groups after Propensity Score Matching

	Before matching (N=1,475)				After matching (N=1,164)			
	EV	TV	Diff.	p-value*	EV	TV	Diff.	p-value*
Age group (1-5)	2.5	2.4	0.0	0.55	2.5	2.5	0.0	1.00
Education (1-8)	4.8	4.2	0.6	0.00	4.2	4.2	0.0	0.49
White collar (%)	30.3	27.6	2.7	0.29	29.4	27.3	2.1	0.47
Not full time worker (%)	27.7	33.5	-5.8	0.02	30.6	32.1	-1.5	0.61
Male (%)	49.7	49.1	0.6	0.87	49.0	49.8	-0.9	0.81
Technology count (1-6)	4.2	3.9	0.3	0.00	4.0	3.9	0.1	0.31
Political information (1-4)	1.5	1.3	0.2	0.00	1.4	1.3	0.0	0.63

Table 1. Replication of Table 2 in Alvarez et al. (2013)

Genetic Matching

We decided to extend using genetic matching because of its superiority over propensity score matching. More specifically, propensity score matching involves representing multiple dimensions of covariates as a unidimensional metric of the propensity score; therefore, it may conduct worse balance across key confounding variables and is sensitive to model specification. On the other hand, the genetic matching method is nonparametric and does not depend on knowing or estimating the propensity score. “It automatically finds the set of matches that minimize the discrepancy between the distribution of potential confounders in the treated and control groups” (Sekhon, 2009). As a result, the genetic matching reduces biases when there are significant imbalances and is less sensitive to model specification (Stuart, 2010).

In order to maintain a comparable model with the original paper, we ran Genetic Matching on exactly the same covariates and their interactions that were previously included in Propensity Score Matching. They were `age.group`, `I(age.group^2)`, `I(age.group^3)`, `I(age.group:educ)`, `I(age.group:tech)`, `educ`, `I(educ^2)`, `tech`, `I(tech^2)`, `pol.info`, `I(educ:pol.info)`, `I(age.group:pol.info)`, `I(tech:pol.info)`, `white.collar`, `not.full.time`, `male`.

We experimented with different matching parameters, including changing the caliper, ties, population size (from 2000 to 3000) and maximum generations (from 50 to 60). Our goal is to minimize the number of dropped units, yet still improve the treatment/control group balance. We found that the higher the pop.size and max.generations, the higher the balance. We also found out that with the caliper, many data points were discarded (i.e. 311 units were dropped in total if caliper = 0.05 as in the original paper), while without setting calipers, no data points were missing. When we tried without calipers and ties = TRUE, the results were acceptable with no units discarded and improved balance. However, the balance as well as the run time was the best when there was no caliper and ties = FALSE, so we ultimately decided to continue with this model.

Our final matching algorithm was running Genetic Matching on 16 covariates and 40 wait generations, each of which has a size of 3000 (pop.size = 3000). Maximum generations were set to 60. Ties = FALSE, and no caliper was set, so no units were dropped, which is more likely to produce less biased causal inference.

Extension of Table 2 in the original paper

After running genetic matching on the dataset, the balance statistics for multiple covariates are computed as shown in Table 2 (the result from R is shown in Appendix A). Even though there is a slight difference between the treated and control units (as evidenced in the Diff. column of “White collar”, “Not full time worker” and “Male” covariates), it can be seen that the overall balance obtained from genetic matching is still better than that from the initial propensity score matching. In other words, the difference between e-voters and traditional voters is slightly reduced with genetic matching. For example, the difference in “White collar” between e-voters and traditional voters was 2.1 in propensity score matching but reduced to 1.2 in genetic matching. Similarly, the difference of the “Not full time worker” variable went down from 1.5 to 0.5.

Furthermore, the p-value also shows that the corresponding covariates in treated and control samples are less statistically different than those from the propensity score matching method. For example, the p-value for “Technology count” and “Political information” improves

from 0.31 and 0.63 under propensity score matching to 0.94 and 0.88 respectively under genetic matching. Note that, here we are referring to the K-bootstrap p-value.

	Before propensity score matching (N=1,475)				After propensity score matching (N=1,164)				After genetic matching (N=1,475)			
	EV	TV	Diff.	p-value _{e*}	EV	TV	Diff.	p-value _{e*}	EV	TV	Diff.	p-value _{e*}
Age group (1-5)	2.5	2.4	0.0	0.55	2.5	2.5	0.0	1.00	2.5	2.5	0.0	1.00
Education (1-8)	4.8	4.2	0.6	0.00	4.2	4.2	0.0	0.49	4.8	4.8	0.0	0.52
White collar (%)	30.3	27.6	2.7	0.29	29.4	27.3	2.1	0.47	30.3	31.4	-1.2	0.64
Not full time worker (%)	27.7	33.5	-5.8	0.02	30.6	32.1	-1.5	0.61	27.7	28.2	-0.5	0.87
Male (%)	49.7	49.1	0.6	0.87	49.0	49.8	-0.9	0.81	49.7	51.7	-2.1	0.41
Technology count (1-6)	4.2	3.9	0.3	0.00	4.0	3.9	0.1	0.31	4.2	4.2	0.0	0.94
Political information (1-4)	1.5	1.3	0.2	0.00	1.4	1.3	0.0	0.63	1.5	1.5	0.0	0.88

Table 2. Balance statistics before matching, after propensity score matching and after genetic matching

Extension of Table 3 in the original paper

With a higher balance from genetic matching, the causal effect of e-voting is calculated as in Table 3 (the R-coded result is shown in Appendix A). The “Diff.” column presents the effect of switching to e-voting by calculating the difference in the response of e-voters and traditional voters. The computing method is the same as that in the original paper, except for the dataset being sorted by genetic matching instead of propensity score matching.

	Before propensity score matching (N=1,475)					After propensity score matching (N=1,164)					After genetic matching (N=1,475)				
	N	E-Voting (%)	Traditional Voting (%)	Diff.	p-value*	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	1,102	81.7	53.4	28.3	0.000	1,631	83.8	57.5	26.3	0.000
Evaluation of voting experience	1,460	46.3	21.3	25.0	0.000	1,151	45.2	21.4	23.8	0.000	1,713	46.3	22.8	23.5	0.000
Ease of voting procedure	1,469	33.6	11.5	22.1	0.000	1,159	32.0	11.4	20.6	0.000	1,726	33.6	10.3	23.3	0.000
Agree substitute TV by EV	1,409	84.1	62.4	21.7	0.000	1,113	82.8	62.6	20.2	0.000	1,657	84.1	68.5	15.7	0.000
Elections in Salta are clean	1,284	58.0	41.0	17.0	0.000	1,021	57.0	41.7	15.3	0.000	1,512	58.0	45.6	12.4	0.000
Sure vote was counted	1,416	85.1	76.3	8.9	0.000	1,123	84.7	76.5	8.2	0.001	1,660	86.4	76.9	9.5	0.000
Qualification of poll workers	1,418	86.4	77.0	9.3	0.000	1,117	85.6	77.5	8.0	0.001	1,670	85.1	76.3	8.9	0.000
Speed of voting process	1,443	84.1	80.9	3.2	0.130	1,135	83.0	81.0	2.0	0.412	1,708	84.1	78.1	5.9	0.002
Confident ballot secret	1,431	77.1	84.5	-7.4	0.001	1,133	76.6	84.7	-8.1	0.001	1,686	77.1	87.4	-10.3	0.000

Table 3. Causal effect of e-voting before matching, after propensity score matching and after genetic matching

In general, the results from genetic matching are not exactly the same numbers as those from propensity score matching, but the overall trend and conclusion about how e-voters perceive their voting experience remains unchanged compared to the original one.

Regarding e-voting preference and e-voting substitution, the results from genetic matching are slightly lower than those from propensity score matching (26.3% under genetic matching vs 28.3% under propensity score for “Select candidates electronically” and 15.7% vs 20.2% respectively for “Agree substitute TV by EV”). The same pattern is observed for the clarity perception, with e-voters 12.4% more likely to describe elections more than “a little” clean (for propensity score method the proportion stands at 15.3%).

On the other hand, the group of variables whose proportion increases with genetic matching includes “Ease of voting procedure” (2.7% increase), “Sure vote was counted” (1.3%

increase) and “Qualification of poll workers” (0.9% increase). Furthermore, it is also reported that e-voters are 5.9% more likely to consider the voting process as “quick” and “very quick”, compared to only 2% more using the dataset matched by propensity score method. Though the difference is relatively small, the result from genetic matching is statistically significant (p -value = 0.002) while that from propensity score matching is not, indicating an improvement when using genetic matching method. Finally, there is a greater concern over the ballot secrecy. With the genetic matched data, e-voters are 10.3% less likely to express confidence in this domain, while with propensity score matching, they are only 8.1% less likely.¹

Overall, despite several over-evaluated and under-evaluated points, the outcome from genetic matching supports the hypothesis from the original paper that e-voting has a positive impact on the confidence on vote being counted, but undermines the confidence on ballot secrecy in Argentina. However, the genetic matching method makes our findings more robust and significant because:

- The p -value obtained from genetic matching indicates the statistical significance of results all given variables
- A better balance between treated units and control units is obtained
- No units are dropped during the genetic matching process, while there are 311 dropped units in propensity score matching, ensuring the significance of treatment effect obtained from the matched treated units.

Conclusion

Compared to propensity score, genetic matching has two main advantages - no dropped treated units and better balance between treated and control units. Therefore, this method helps generate more statistically significant and robust results. In this case, the outcome gathered from genetic matching consolidates the impact of e-voting versus traditional voting on people's confidence and perception about their voting experience.

¹ **#significance:** We computed the causal effect of voting technology on multiple variables in Table 3 and used p -value to prove the statistical significance of the effect. Also, the P -value was a powerful tool to measure the match balance of the datasets for both using GenMatch and propensity score matching.

By using a more rigorous methodology (genetic matching), our results suggest that e-voting generally has a positive impact on people's voting experience compared to traditional voting. Additionally, the outcome further supports the notion that e-voting technology positively affects the perception of the cleanliness of elections, but reduces the confidence in the ballot secrecy in Argentina. A sensitivity analysis can be implemented to account for potential sources of bias and unobserved attributes from voters. Insights gathered from this activity further inform the robustness of the obtained results in the presence of unobserved covariates. As the study is implemented in Argentina, the Election Commission of this country can use the results as a scientific methodology to prove that actions should be enacted to boost people's confidence in ballot secrecy.

Word Count: 2550 words

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.
- Sekhon, J (2009). Opiates for the Matches: Matching Methods for Causal Inference. *Annual Review of Political Science. *Vol. 12: 487-508. Retrieved 4 November 2016 from:
Retrieved from <http://sekhon.berkeley.edu/papers/opiates.orig.pdf>
- Stuart E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science : a review journal of the Institute of Mathematical Statistics*, 25(1), 1–21. <https://doi.org/10.1214/09-STS313>

[1] "# Replicating Table 3									
	N	prop.ev	prop.tv	diff	pvalue	N	prop.ev	prop.tv	diff
eselect.cand	1388	83.84	53.42	30.428	1.237e-34	1102	81.69	53.39	28.293
eval.voting	1460	46.33	21.30	25.035	1.833e-22	1151	45.23	21.43	23.805
easy.voting	1469	33.64	11.53	22.111	5.420e-22	1159	31.95	11.38	20.572
agree.evoting	1409	84.14	62.44	21.705	2.864e-20	1113	82.76	62.59	20.175
how.clean	1284	57.97	40.99	16.980	2.561e-09	1021	57.00	41.65	15.349
capable.auth	1416	85.14	76.25	8.889	2.954e-05	1123	84.67	76.52	8.150
sure.counted	1418	86.35	77.02	9.332	7.444e-06	1117	85.56	77.52	8.044
speed	1443	84.06	80.85	3.209	1.298e-01	1135	83.04	80.99	2.047
conf.secret	1431	77.11	84.53	-7.417	6.506e-04	1133	76.57	84.67	-8.104
		pvalue							
eselect.cand		1.918e-23							
eval.voting		1.866e-17							
easy.voting		3.449e-17							
agree.evoting		7.122e-14							
how.clean		1.288e-06							
capable.auth		7.512e-04							
sure.counted		7.025e-04							
speed		4.117e-01							
conf.secret		7.294e-04							

[1] "# Table 3 After Using GenMatch									
	N	prop.ev	prop.tv	diff	pvalue	N	prop.ev	prop.tv	
eselect.cand	1388	83.84	53.42	30.428	1.237e-34	1631	83.84	57.49	
eval.voting	1460	46.33	21.30	25.035	1.833e-22	1713	46.33	22.83	
easy.voting	1469	33.64	11.53	22.111	5.420e-22	1726	33.64	10.30	
agree.evoting	1409	84.14	62.44	21.705	2.864e-20	1657	84.14	68.47	
how.clean	1284	57.97	40.99	16.980	2.561e-09	1512	57.97	45.60	
sure.counted	1418	86.35	77.02	9.332	7.444e-06	1660	86.35	76.85	
capable.auth	1416	85.14	76.25	8.889	2.954e-05	1670	85.14	76.29	
speed	1443	84.06	80.85	3.209	1.298e-01	1708	84.06	78.13	
conf.secret	1431	77.11	84.53	-7.417	6.506e-04	1686	77.11	87.38	
		diff			pvalue				
eselect.cand		26.349	2.735e-31						
eval.voting		23.499	2.675e-24						
easy.voting		23.342	2.172e-31						
agree.evoting		15.669	1.009e-13						
how.clean		12.377	1.909e-06						
sure.counted		9.499	8.089e-07						
capable.auth		8.850	6.495e-06						
speed		5.928	2.168e-03						
conf.secret		-10.275	4.617e-08						

Appendix B: Team member contributions

Loan Nguyen: I wrote the executive summary, background and genetic matching sections

Chau Le: I formatted the tables for replication and extension in excel and wrote the extension parts for Table 2 and Table 3 from the original data, as well as the conclusion.

Chantsaldiimaa Lkhagvatogtokh: I mainly worked on the replication and extension part of this paper by running the code in R, comparing results with the original paper's, and collaborating with Kareem on using GenMatch. Also, I wrote a justification comparing genetic matching and propensity score matching for the extension of this paper.

Kareem Alsayed: I mainly worked on the jupyter notebook that runs the code. I implemented the different matching trials to find the best balance using the GenMatch function. Also, I handled the different issues that existed during the table replication part.