Decision Memo

To: German Statistical Office

From: Anu Munkhsainkhan

CC: Prof. Hadavand

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

Executive Summary

This decision memo is addressed to the German Statistical Office to advise on the implementation of electric voting (e-voting) systems. We argue that e-voting should be implemented in Germany as it has many advantages such as reduced cost, increased voter turnout, and faster election results. Moreover, we replicated and extended Alvarez, Levin, Pomares, & Leiras. (2013) to understand the effect of e-voting on voters' confidence in the results and perception of ease. First, we **replicated the propensity score matching** and got close but a slightly different balance. Secondly, we **extended the paper by running genetic matching**, and the results were almost the same as the propensity score matching except one variable that got slightly worse balance. It suggests that the propensity score model used in the paper is optimal in achieving balance.

Finally, Alvarez, Levin, Pomares, & Leiras. (2013) effectively inform the validity of the assumed benefits of e-voting and public perception. It revealed that people perceived e-voting as more convenient than traditional voting, but there was no perceived difference in the "quickness" of e-voting. Surprisingly, people who used e-voting systems had more confidence in election results, undermining the concerns raised in Germany. However, people had concerns about their privacy in e-voting. **Therefore, if the German government decides to continue the e-voting system implementation in the future, citizens' privacy concerns might be stronger than distrust in election results.**

Background

Studies on e-voting have been increasing over the years since many countries had started replacing traditional voting systems with e-voting. For example, government elections in the United Kingdom, party primary elections in the United States used internet voting systems (Yi & Okamoto, 2013), and Estonia became the first country to hold nation-wide elections using an internet voting system (e-Estonia Briefing Center, 2020). The benefits of e-voting include increased voters turnover by increasing availability, reduced costs since there is no need to print high-security ballot papers, and fast results due to automation (Dimeski, Boskoska, & Prisajganec, 2013).

However, there are some disadvantages such as privacy concerns, vulnerability to hacking, and potential distrust from citizens (Dimeski, Boskoska, & Prisajganec, 2013). Thus, there is a need to study the extend of these disadvantages. Alvarez, Levin, Pomares, & Leiras. (2013) highlights the lack of field experiments on the implication of e-voting on voting experience and aims to fill this gap, especially in the context of developing countries. The study findings can deepen the understanding of public perception on e-voting and inform further implementation strategies.

Summary of the Original Paper

Hypotheses

Alvarez, Levin, Pomares, & Leiras. (2013) developed and tested the following two hypotheses:

- “Electronic information technologies are better at increasing perceptions of the accuracy of the tally than the secrecy of the ballot, especially when the process involves transmitting information over the internet.”
- E-voting improves the perception of usability and ease of voting

Study Design

The study takes advantage of the gradual introduction of e-voting to Salta province in Argentina to test the hypotheses. Even though a randomized experiment would have been the best method to study the effect of e-voting, it is costly and impractical in this case. Some residents in the rural part of Salta does not have access to electricity, and the illiteracy rates could reach up to 10 percent in those areas while the national average is 1.9 percent.

Furthermore, the government assigned the voting system at the voting station level, and the allocation was not random. Considering the impracticality and the partial implementation of e-voting, the researchers used statistical matching to test the hypotheses. Thus, the treatment of the study is the e-voting system, and causality can be inferred by balancing the observed attributes of people who e-voted (treatment group) and people who used the traditional voting system (control group).

On election day, researchers distributed questionnaires to 1,502 voters, consisting of 887 e-voters and 615 traditional voters. The questionnaire asked the voters' opinion on "ease and speed of the voting process" and "confidence in the fairness of the election process," which will be the outcome variables.

The matched attributes or covariates are age group, education, gender, technology use, political information, full-time work status, and whether the voter has an office-setting job (white-collar job). Moreover, the research used propensity score matching and matched voters who have less than a 5 percent difference in their propensity score standard deviation (caliper). Table 1 shows that voters' attributes are reasonably balanced in the treatment and control groups after matching, which can be seen from their small mean differences and p-values.

	Before matching (N = 1,475)				After matching (N = 1,164)				
	EV	TV	Diff.	p-value*	EV	TV	Diff.	p-value*	% Imp.
Age group (1–5)	2.5	2.4	0.0	0.55	2.5	2.5	0.0	1.00	100%
Education (1–8)	4.8	4.1	0.6	0.00	4.2	4.2	0.0	0.72	98%
White collar (%)	30.3	27.6	2.7	0.29	29.2	28.4	0.9	0.80	68%
Not full time worker (%)	27.7	33.5	−5.8	0.02	30.8	32.0	−1.2	0.80	79%
Male (%)	49.7	49.1	0.6	0.87	49.0	49.0	0.0	1.00	100%
Technology count (1–6)	4.2	3.9	0.3	0.00	4.0	3.9	0.1	0.59	76%
Political information (1–4)	1.5	1.3	0.2	0.00	1.4	1.3	0.0	0.55	77%

Table 1. Balance Statistics from the Paper (Dimeski, Boskoska, & Prisajganec, 2013)

Results and Discussion

The study discovered that e-voters are 20 percent more likely to find voting “easy” compared to traditional voters and supported e-voting as a replacement of the traditional voting system. There is no significant difference between e-voters and traditional voters in the perception of “quickness” in the voting process. Furthermore, in terms of confidence in the election process, “e-voters are 16 percent more likely to report that elections are more than ‘a little’ clean.” However, e-voters seem to be less confident in their privacy as it was “7 points less likely to say that they are ‘confident’ or ‘very confident’ that the secrecy of their votes was preserved.”

The results suggest that e-voting is perceived as an easier and cleaner voting system compared to the traditional one. However, it does not seem to be perceived as a quicker alternative, which is an important finding since efficiency is one of the major reasons governments implement e-voting.

Replication

We decided to perform partial study replication using the published replication code, available in the Appendix. We focused on replicating the balance summary after propensity score matching and the treatment effect summary for each outcome variable that are displayed as Tables 2 and 3 in Alvarez, Levin, Pomares, & Leiras. (2013).

Although we have used the original code from the study, we have reached a different covariate balance after running the propensity score matching. Therefore, we added an extra section to the original Table 2, comparing our propensity score match with the original propensity score match. We do not have an exact explanation for the differences, considering that we used the exact same code. However, we guess that library and R updates could be the reason since the paper was written seven years ago.

In terms of Table 3, the treatment effects from the original paper and the replicated result are different but the difference is not significant. The biggest change in the treatment effect, a decrease of 0.7, is observed in the outcome variable "confident ballot secret." It means the perceived negative effect of e-voting on voter privacy got weaker in our replication. All the other treatment effects are relatively similar. [1]

[1] **#studyreplication**: After presenting the background research, we designed and conducted partial study replication to verify the study's treatment effect and covariate balance; extended the results by performing a genetic matching. Also, we provided the full replication R code in Appendix.

Replication of Table 2: Balance Statistics

	Before Matching (N=1475)				Propensity (2013) (N=1164)				Propensity (2020) (N=1168)			
	EV	TV	Diff.	p-value	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	2.5	2.5	0.0	1.00
Education (1-8)	4.8	4.1	0.6	0.00	4.2	4.2	0.0	0.72	4.2	4.2	0.0	0.74
White Collar (%)	30.3	27.6	2.7	0.29	29.2	28.4	0.9	0.80	29.5	28.4	1.0	0.75
Not full time worker (%)	27.7	33.5	-5.8	0.02	30.8	32.0	-1.2	0.80	30.7	31.5	-0.9	0.80
Male (%)	49.7	49.1	0.6	0.87	49.0	49.0	0.0	1.00	48.8	49.7	-0.9	0.81
Technology count (1-6)	4.2	3.9	0.3	0.00	4.0	3.9	0.1	0.59	4.0	3.9	0.1	0.44
Political information (1-4)	1.5	1.3	0.2	0.00	1.4	1.3	0.0	0.55	1.4	1.3	0.0	0.59

Replication of Table 3: Causal effect of e-voting

	Before Matching (N=1475)					Propensity (2013) (N=1164)					Propensity Matching (N=1168)				
	N	EV (%)	TV (%)	Diff.	p-value	N	EV (%)	TV (%)	Diff.	p-value	N	EV (%)	TV (%)	Diff.	p-value
Select candidates electronically	1388	83.8	53.4	30.4	0.000	1101	82.7	54.1	28.6	0	1106	82.3	54.1	28.2	0.000
Evaluation of the voting experience	1460	46.3	21.3	25.0	0.000	1151	45.6	20.9	24.7	0	1153	44.6	20.2	24.5	0.000
Agree substitute TV by EV	1469	33.6	11.5	22.1	0.000	1159	32.5	11.9	20.6	0	1161	31.4	11.2	20.2	0.000
Elections in Salta are clean	1409	84.1	62.4	21.7	0.000	1114	82.4	63.3	19.1	0	1118	83.1	63.6	19.5	0.000
Ease of voting experience	1284	58.0	41.0	17.0	0.000	1022	57.6	41.5	16	0	1024	58.0	41.4	16.6	0.000
Sure vote was counted	1416	85.1	76.3	8.9	0.000	1117	85.6	77	8.8	0	1124	85.2	75.9	9.4	0.000
Qualification of pool workers	1418	86.4	77.0	9.3	0.000	1123	84.5	76	8.5	0	1118	85.9	76.9	9.0	0.000
Speed of voting process	1443	84.1	80.9	3.2	0.130	1137	83.2	80.7	2.5	0.306	1137	83.1	80.7	2.4	0.331
Confident ballot secret	1431	77.1	84.5	-7.4	0.001	1133	76.9	84.3	-7.4	0.002	1134	77.8	84.5	-6.7	0.005

Extension

Problems with Propensity Score

Propensity score is one of the most common matching method used in research (Haukoos & Lewis, 2015). However, there are some arguments against using propensity score matching. First of all, King & Nielsen (2018) argue that random pruning in propensity score increases imbalance and ultimately bias. The paper also argues that propensity score matching approximates complete randomization while other methods, Mahalanobis Distance for example, approximates blocked randomization, which is superior in reducing model dependency and bias (King & Nielsen, 2018). The second major drawback is the need for manual iteration.

Rosenbaum and Rubin (1984) suggest iterating propensity score models to reach an optimal balance, which could be time and labor-intensive. Therefore, we decided to extend the paper by conducting genetic matching to test whether it could reach a better balance than propensity score matching. We chose genetic matching because it is a generalization of the Mahalanobis Distance matching method (Diamond & Sekhon, 2013), which approximates blocked randomization. Also, it uses an iterative genetic algorithm, removing the need for manual iteration recommended by Rosenbaum and Rubin (1984).

For the genetic matching, we have chosen MatchIt function with the arguments "genetic", population size 1000, and a total of 300 generations. Figure 1 shows the balance of the covariates using "Love" plot for an unbalanced dataset, propensity score, and genetic matching dataset. Table 1 extends the Table 2 published in the paper by adding the genetics matching results. As we can see the balance produced by genetic matching is not that different than the balance achieved by propensity score matching. In fact, the attribute "white-collar" got slightly worse for genetic matching compared to propensity score. We also tried increasing the population size to 1500 and maximum generation to 900 but it did not yield any results after 6 hours of running the code. Other attempts using different calipers made the balance worse. Thus, we conclude that the original propensity score model is effective in achieving the desired balance.

Extension Results

The right side of Figure 1 shows the p-value for the KS test that better represents the shape of the distribution than a standard p-value, which only reflects the difference in the means. Based on the results, there is no significant visible improvement of the KS test compared to the propensity score match used in the study. Propensity score matching yields better results for the variables "education" and “office job” (white-collar). However, genetic matching produces better outcomes for “political preference” and “age group.” [2]

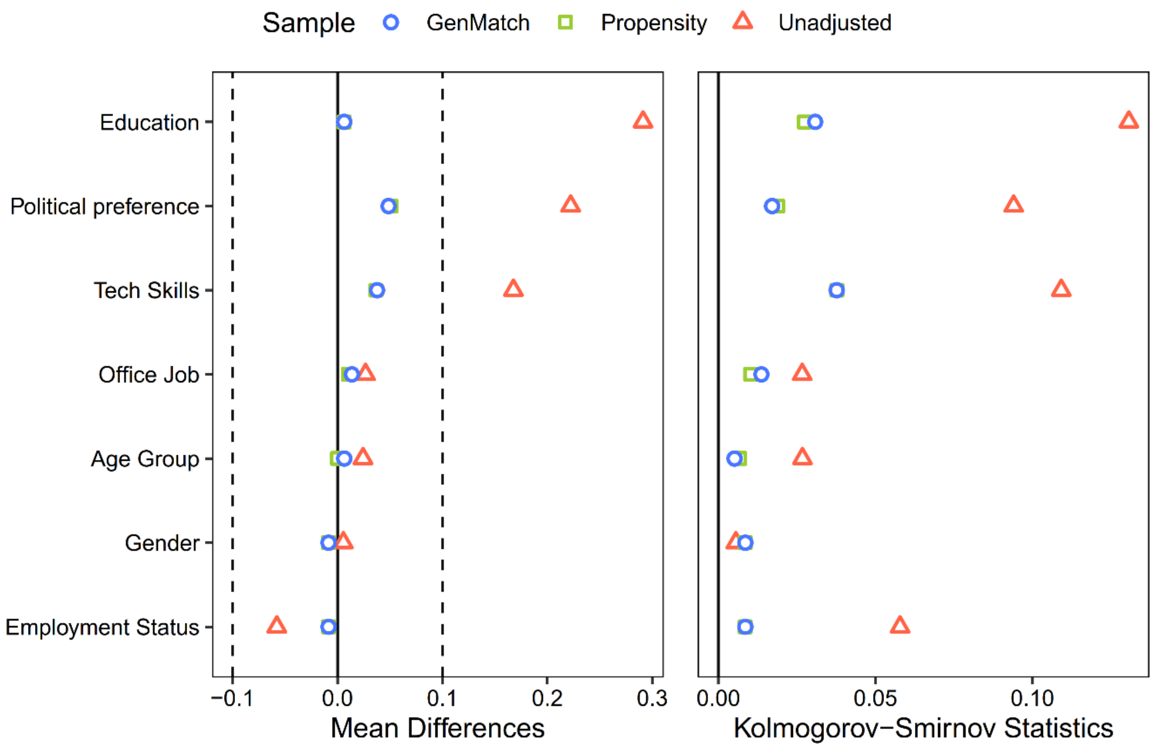

Figure 1. Covariate Balance

Figure 2 summarizes the comparison of each covariate’s distribution. Both methods achieve reasonably good covariate balance, and the distributions are overall similar. Slight differences in the distributions are visible on Figure 2 for “tech skills” and “education”. Both variables are ordered, discrete, categorical variables. Even though these variables are matched in terms of their means, a few categories are not equally represented in the after-match sample. For example, people who have tech skill score of four are more represented in the control group. On the other hand, the treatment group has more of people who have tech skill score of three and five. Thus, the balance for “tech-skill” is not as strong as some variables that have small mean difference and same distribution. [3]

[2] **#gapanalysis**: Identified limitations of propensity score based on evidence-based arguments and proposed an existing solution (genetic matching) to fill the limitations; evaluated that propensity score is effective by comparing the results from propensity score and genetic matching.
[3] **#distributions:** Prepared multiple visualizations to compare the balance of covariates distribution as balanced distributions are key for the interpretation of causal studies. We compared both mean differences and KS test, to quantify the quality of match of two distributions.

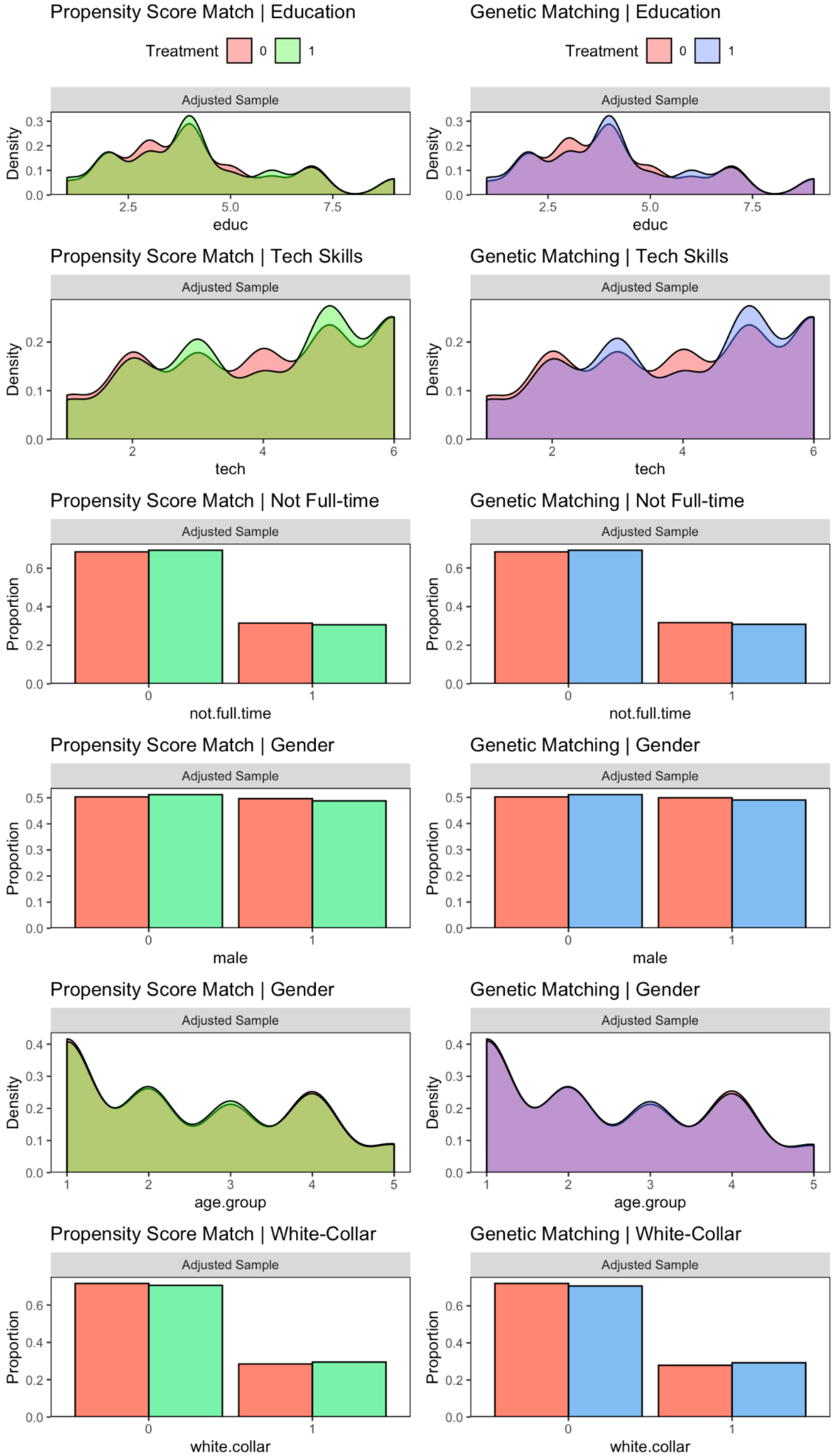

Figure 2: Visual Summary of Covariates' Distribution

Extension of Table 2: Balance Statistics

	Before Matching (N=1475)				Propensity (2013) (N=1164)				Propensity (N=1168)				Genetic Matching (N=1168)			
	EV	TV	Diff.	p-value	EV	TV	Diff.	p-value	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	2.5	2.5	0.0	1.00	2.5	2.4	0.0	1.00
Education (1-8)	4.8	4.1	0.6	0.00	4.2	4.2	0.0	0.72	4.2	4.2	0.0	0.74	4.2	4.2	0.1	0.61
White Collar (%)	30.3	27.6	2.7	0.29	29.2	28.4	0.9	0.80	29.5	28.4	1.0	0.75	29.3	27.9	1.3	0.65
Not full-time worker (%)	27.7	33.5	-5.8	0.02	30.8	32.0	-1.2	0.80	30.7	31.5	-0.9	0.80	30.8	31.7	-0.9	0.80
Male (%)	49.7	49.1	0.6	0.87	49.0	49.0	0.0	1.00	48.8	49.7	-0.9	0.81	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	0.59	4.0	3.9	0.1	0.44	4.0	3.9	0.1	0.43
Political information (1-4)	1.5	1.3	0.2	0.00	1.4	1.3	0.0	0.55	1.4	1.3	0.0	0.59	1.4	1.3	0.0	0.64

Extension of Table 3: Causal effect of e-voting

	Before Matching (N=1475)					Propensity (2013) (N=1164)					Propensity Matching (N=1168)					Genetic Matching (N=1168)				
	N	EV (%)	TV (%)	Diff.	p-value	N	EV (%)	TV (%)	Diff.	p-value	N	EV (%)	TV (%)	Diff.	p-value	N	EV (%)	TV (%)	Diff.	p-value
Select candidates electronically	1388	83.8	53.4	30.4	0.000	1101	82.7	54.1	28.6	0	1106	82.3	54.1	28.2	0.000	1106	82.3	54.4	27.9	0.000
Evaluation of the voting experience	1460	46.3	21.3	25.0	0.000	1151	45.6	20.9	24.7	0	1153	44.6	20.2	24.5	0.000	1156	44.7	21	23.8	0.000
Agree substitute TV by EV	1469	33.6	11.5	22.1	0.000	1159	32.5	11.9	20.6	0	1161	31.4	11.2	20.2	0.000	1120	83.1	62.7	20.4	0.000
Elections in Salta are clean	1409	84.1	62.4	21.7	0.000	1114	82.4	63.3	19.1	0	1118	83.1	63.6	19.5	0.000	1163	31.3	11.2	20.2	0.000
Ease of voting experience	1284	58.0	41.0	17.0	0.000	1022	57.6	41.5	16	0	1024	58.0	41.4	16.6	0.000	1024	58.1	41.1	17	0.000
Sure vote was counted	1416	85.1	76.3	8.9	0.000	1117	85.6	77	8.8	0	1124	85.2	75.9	9.4	0.000	1120	85.9	76.9	9	0.000
Qualification of pool workers	1418	86.4	77.0	9.3	0.000	1123	84.5	76	8.5	0	1118	85.9	76.9	9.0	0.000	1126	85.1	76.3	8.8	0.000
Speed of voting process	1443	84.1	80.9	3.2	0.130	1137	83.2	80.7	2.5	0.306	1137	83.1	80.7	2.4	0.331	1139	83.1	80.7	2.4	0.331
Confident ballot secret	1431	77.1	84.5	-7.4	0.001	1133	76.9	84.3	-7.4	0.002	1134	77.8	84.5	-6.7	0.005	1136	77.9	84.55	-6.7	0.000

Conclusion

Overall, the partial study replication and extension of genetic matching strengthen the validity of Dimeski, Boskoska, & Prisajganec (2013). The slight differences in the replication results are negligible. Moreover, the similar results from genetic matching prove that the study methods are robust. However, genetic matching automates the process of iteration to find the optimal weights, while it is unclear how the researchers settled on the propensity score model they used in the study.

Furthermore, Germany had been successfully running e-voting trials from 1998 to 2005 until two citizens filed a case to the German Constitutional Court, stating that 2005 election results are unreliable, and the use of voting machines is unconstitutional (Sebes, 2009). The court denied the argument against the election result but agreed that e-voting is against the constitutional principle that "all essential steps in the elections are subject to public examinability."

As a result, e-voting was banned in Germany unless it allowed citizens to scrutinize the process without expert knowledge (Palmer, 2009). However, if further technological advancements align with the constitutional principle and Germany decides to continue e-voting, the study result emphasizes the need to address the voters' concern for their privacy. Other than that, the voters perceived e-voting as easy and trust-worthy in terms of election results.

Strengthening data protection technologies, policies and informing the voters on the government's safety measures can smoothen the transitions to e-voting systems. Lastly, one should keep in mind that Dimeski, Boskoska, & Prisajganec (2013) was conducted in Argentina, and country differences should be considered when drawing any generalized conclusions.

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. doi: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. *The Review of Economics and Statistics*, *95*(3), 932–945.

Dimeski, Branko & Boshkoska, Meri & Prisaganec, Milco. (2010). ELECTRONING VOTING: NECESSITY OR FICTION.

E-Estonia Briefing Center. (2020, September 21). Estonia - We have built a digital society and we can show you how. Retrieved December 14, 2020, from https://e-estonia.com/

Haukoos, J. S., & Lewis, R. J. (2015). The propensity score. *Jama*, *314*(15), 1637–8. https://doi.org/10.1001/jama.2015.13480

King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political Analysis*, *27*(4), 435–454. https://doi.org/10.1017/pan.2019.11

Palmer, E. (2009, March 25). Global Legal Monitor. Retrieved from https://www.loc.gov/law/foreign-news/article/germany-constitutional-court-decision-on-electronic-voting/

Rosenbaum, Paul R. and Donald B. Rubin. 1984. "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score." Journal of the American Statistical Association 79 (387): 516–524.

Sebes, J., & A National Call to Link Arms for Democracy - Humor Times. (2009, September 02). Electronic Voting Banned in Germany. Retrieved from https://trustthevote.org/blog/2009/09/02/electronic-voting-banned-in-germany/

Yi, X., & Okamoto, E. (2013). Practical Internet voting system. *Journal of Network and Computer Applications, 36*(1), 378-387. doi:10.1016/j.jnca.2012.05.005

Word count: 1714