Supplementary Information for: Internet Voting and Turnout: Evidence from Switzerland

Micha Germann* Uwe Serdült†

March 6, 2017

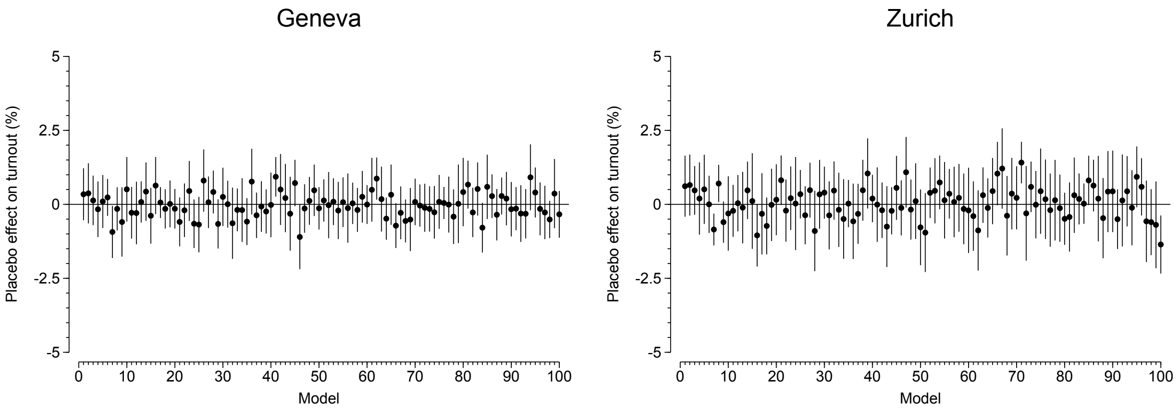
*Department of Political Science, University of Pennsylvania.
†Center for Democracy Studies Aarau, University of Zurich.

Supplement A: Variables

Variable	Description	Sources
Turnout	Percentage of votes cast as a share of the total number of citizens eligible to vote in federal referendums. If there were multiple issues to vote on, turnout is based on the referendum vote with the highest participation.	Swiss Federal Statistical Office (2015)
I-voting	Dummy indicating the availability of remote Internet voting in a municipality on a given referendum day.	Statistical Office of the Canton of Geneva (2015), Statistical Office of the Canton of Zurich (2015), Swiss Federal Statistical Office (2015)
Population (logged)	Natural logarithm of the total population of a municipality in a given year. Available up to and including 2013.	Swiss Federal Statistical Office (2015)
35- share	Percentage of people aged 35 years and under in a municipality in a given year (in the canton of Zurich: 34 years and under). Available up to and including 2013.	Statistical Office of the Canton of Geneva (2015), Statistical Office of the Canton of Zurich (2015)
65+ share	Percentage of people aged 65 years and above in a municipality in a given year. Available up to and including 2013.	Statistical Office of the Canton of Geneva (2015), Statistical Office of the Canton of Zurich (2015)
Per capita income (logged)	Natural logarithm of per capita income in constant 1993 Swiss Francs (1,000s). Available up to and including 2011.	Swiss Federal Tax Administration (2015), Swiss Federal Statistical Office (2015)
Unemployed	Percentage of unemployed as a share of persons aged 15–64 years. Available up to and including 2013.	Swiss Federal Statistical Office (2015)

Foreigners	Percentage of non-Swiss citizens as a share of the total population. Available up to and including 2013.	Swiss Federal Statistical Office (2015)
Vote share (SP)	Percentage of votes cast for the Social Democrats (SP) in the last national election (lower house).	Swiss Federal Statistical Office (2015)
Vote share (Greens)	Percentage of votes cast for the Green party in the last national election (lower house).	Swiss Federal Statistical Office (2015)
Young voter turnout	Percentage of votes cast by voters aged 35 years and under as a share of total number voters aged 35 years and under. If there were multiple issues to vote on, turnout is based on the referendum vote with the highest participation. Available only for Geneva and only up to and including September 2014.	Statistical Office of the Canton of Geneva (2015)
Turnout past 4 years	Average turnout of past four calendar years.	Swiss Federal Statistical Office (2015)
Turnout trend	Average turnout in the past two calendar years minus average turnout three and four years ago.	Swiss Federal Statistical Office (2015)
Population density (logged)	Natural logarithm of the number of residents by square kilometer. Available up to and including 2013.	Swiss Federal Statistical Office (2015)

Figure 1: Placebo tests



(a) (b)

Note: The dots represent point estimates and the spikes 95% confidence intervals.

Supplement B: Pre-intervention trends

In the paper we showed graphical evidence that pre-intervention trends between i-voting (treated) and non-i-voting (untreated) municipalities are parallel. The conclusion remains the same under a more formal test of the equivalence of pre-intervention trends based on randomly generated placebo treatments. The placebo test proceeds as follows: First, for each municipality that had i-voting at least once we randomly draw a date from the pre-intervention phase. Second, we define a fake treatment indicator taking the value 1 for all referendum days after the randomly generated date, 0 otherwise.[1] Municipalities that never had i-voting are always coded 0. Third, we re-estimate the main model including municipality and referendum day fixed effects as well as quadratic municipality time trends, but this time using the fake treatment indicator instead of the i-voting identifier. Finally, we repeat this procedure 99 times. Figure 1 plots the resulting 200 placebo treatment estimates (100 for Geneva and 100 for Zurich). Statistically insignificant placebo effects can be seen as evidence that pre-intervention trends are parallel, whereas statistically significant placebo effects speak against the equivalence of pre-intervention trends. In both cantons only four out of the 100 placebo models are statistically significant at the 95% level. This is within what we would expect if the null hypothesis of parallel

[1]For treated units the first referendum day is by definition 0 and the last always 1.

Table 2: Predicting municipalities that adopted i-voting

	(1) Geneva	(2) Geneva	(3) Zurich	(4) Zurich
Phase	I	II	I	II
Year(s) of i-voting adoption	2004	2008–2012	2005	2008
Turnout past 4 years	−0.045	0.321*	0.207	0.48***
	(0.232)	(0.192)	(0.222)	(0.165)
Turnout trend	0.489	−0.454	−0.018	−0.338
	(0.605)	(0.504)	(0.461)	(0.311)
Per capita income (logged)	2.956	−1.694	−11.383*	1.656
	(2.133)	(1.968)	(6.552)	(2.253)
Population (logged)	0.176	−0.491	1.035	2.23***
	(1.069)	(0.846)	(1.219)	(0.773)
65+ share	0.011	−0.131	0.000	−0.523***
	(0.209)	(0.152)	(0.313)	(0.197)
Population density (logged)	0.636	1.96**	0.747	0.035
	(1.147)	(0.957)	(1.465)	(0.931)
Constant	−15.478	−22.343	11.547	−45.303***
	(13.15)	(15.23)	(22.786)	(13.061)
Observations	33	37	161	168
Pseudo R2	0.309	0.246	0.207	0.371

Note: This table shows the results from logit regressions explaining whether a municipality adopted i-voting or not. Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

pre-intervention trends holds true.

Supplement C: Matching

This section describes the procedure used to re-estimate the DID models based on matched samples. In the paper the results are reported in Table 2 (columns 3 and 4).

As indicated in the paper, we matched treated and untreated municipalities based on the propensity (probability) of the introduction of i-voting in a municipality (Rosenbaum & Rubin 1983). In both cantons, we derived matches based on two separate samples, one for early adopters and one for late adopters. We then used the combined phase I and II matching weights to re-estimate the DID models. For a similar approach see Galiani, Gertler & Schargrodsky (2005) and Ryan, Burgess & Dimick (2015).

To make this clearer, consider the case of Geneva in more detail. As a first step, we separated the sample of treated municipalities (i.e. those that had i-voting at least once) into those running their first trials in 2004 (a total of 8 municipalities) and those running their first trials in 2008 or after (12 municipalities). We then conducted two separate searches for matching control units, one for the early adopters and one for the late adopters. In what follows we consider the case of the early adopters in some more detail. To identify matching control units for Geneva's early adopters, we ran a logistic regression explaining the adoption of i-voting in 2004, using the following predictors: past turnout levels (i.e. average turnout in 2000-2003), past turnout trends (i.e. average turnout in 2002-2003 minus average turnout in 2000-2001), and lagged (i.e. 2003) values of per capita income, population size, the 65+ share, and the population density. Since we aim to eliminate possible confounding of the effect on turnout, the choice of predictors was simultaneously driven by expected correlations with both i-voting adoption and turnout. Note that since we want to match treated municipalities with control municipalities, we dropped all 12 municipalities that adopted i-voting in the second phase (i.e. 2008 onwards). The first column in Table 2 shows the results of the logistic regression explaining i-voting adoption in 2004. None of the predictors is statistically significant, suggesting that there are no or at least very limited differences between municipalities that introduced i-voting in 2004 and control units.

Based on this model, we then predicted the propensity that a municipality introduced i-voting in 2004 and matched treated and control municipalities based on the propensity score using 1:1 nearest neighbor matching with replacement, imposing common support. This means that for each treated municipality we identified the control municipality with the closest propensity score after removing treated municipalities whose propensity score is higher than the maximum or less than the minimum propensity score of the control units. Note that we match control units to treated units with replacement, meaning that the same control unit can be associated with multiple treated municipalities. This procedure results in a frequency weight for each municipality indicating the number of times it was matched. By definition, this weight is one for treated units on common support and zero for treated units off support. For control units the weight can vary from zero (if the control unit was not matched to any treated unit) to, at least in theory, the number of treated units on common support (in the unlikely case that the same control unit turns out to be the nearest neighbor for all treated units). In the present case, one of the eight treated municipalities was dropped because it was off common support (Cologny). The remaining seven treated municipalities received a weight of one (Anières, Carouge (GE), Collonge-Bellerive, Meyrin, Onex, Vandoeuvres, and Versoix). Out of the 25 control units, 20 were not matched to any treated unit and thus received a weight of zero, four were matched to one treated unit and thus received a weight of one (Choulex, Genthod, Puplinge, and Veyrier) while one was matched to three treated units and received a weight of three (Lancy).

Having identified matching control units for the early adopters, we repeated this process for the late adopters. That is, we matched municipalities that adopted i-voting in 2008 or later with control municipalities based on the propensity score derived from a logistic regression of i-voting adoption in 2008 on past turnout levels (i.e. average turnout in 2004-2007), past turnout trends (i.e. average turnout in 2006-2007 minus average turnout in 2004-2005), and lagged (i.e. 2007) values of per capita income, population size, the share of 65+, and population density. Municipalities that had run their first i-voting trials before 2008 were excluded. This again resulted in a set of matching weights. Note that unlike in the early adopter model, some of the treatment predictors are statistically significant in the late adopter treatment model (see column 2 in Table 2). Specifically, municipalities that adopted i-voting in 2008 or later tend to be more densely populated compared to municipalities that never had i-voting, and had slightly higher turnout in 2004-2007 (though past turnout is significant only at the 10% level). Nearest neighbor matching decreases such heterogeneity between treated and untreated municipalities, thus minimizing potential biases that could emerge.

As a final step, we combined the matching weights of the late adopter analysis with those from the early adopter analysis by adding them up and re-estimated the DID model using the combined matching weights as frequency weights. This leads to the results reported in Table 2 (column 3) in the paper.

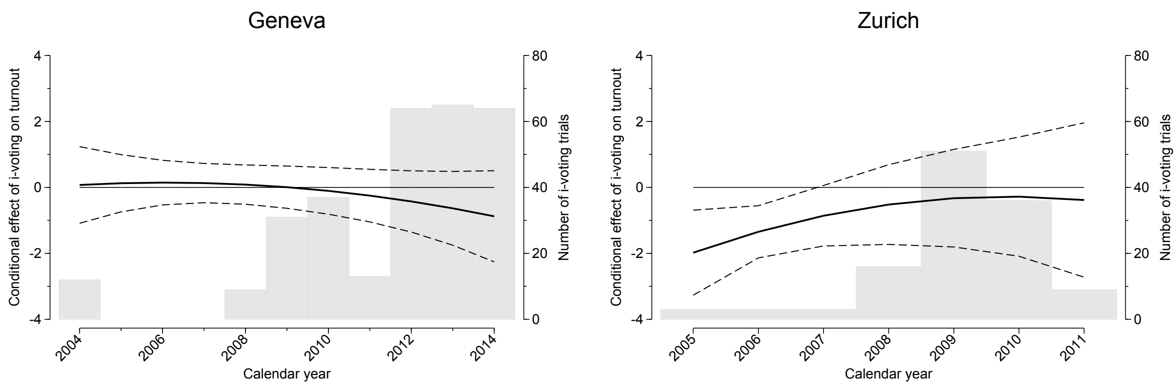
The procedure used for the canton of Zurich is analogous. We separated the sample of i-voting municipalities into those running their first i-voting trials in 2004 (early adopters) and 2008 (late adopters), respectively. Again proceeding separately for early and late adopters, we then searched matching control units using the propensity score. The propensity score derives from the same predictors, and we again employed 1:1 nearest neighbor matching with replacement, imposing common support. Again, the matching weight used in the paper results from adding up the matching weights for early and late adopters. The results are reported in Table 2 (column 3) in the paper. Note that while none of the treatment predictors is statistically significant in the early adopter model, past turnout, population size, and the 65+ share have a statistically significant relationship with i-voting adoption in the late adopters model. Re-estimation based on matched samples should minimize potential biases due to this.

Supplement D: Additional robustness tests

This section reports a number of additional robustness checks. The main focus lies on the possibility of heterogeneous treatment effects. As mentioned in the paper, two possibilities for heterogeneous effects are a novelty effect and changes in the salience of security aspects. Both mechanisms would suggest that the treatment effect decreases with time. To address this, we interacted the treatment indicator with a linear function of time, which suggested that the treatment effect is reasonably stable over time. In Figure 2 we show that the conclusion remains unchanged if we allow for more complex interactions with time (i.e. with a quadratic function of time). Specifically, for Geneva we again find that the effect of i-voting on turnout is decreasing with time but always statistically insignificant (see the left panel). For Zurich, we find that i-voting negatively affected turnout in the two initial years, but has no statistically significant relationship with turnout in all other years. Again, the finding for the two initial years can probably be discarded, given the very low number of i-voting trials on which the estimate is based (see the histogram in the background).

As a next step, we consider an alternative specification testing for a possible novelty effect. Thus far we assumed that novelty is a function of time. The main rationale for this

Figure 2: Interaction with time squared



(a) (b)

Note: The solid line indicates the effect of i-voting on turnout conditional on the calendar year and the dashed lines give the 95% confidence intervals. The underlying histogram gives the number of i-voting trials per calendar year. The coefficients on the product terms are 0.198/-0.016 (clustered standard errors=0.352/0.018) and 1.441/-0.073 (1.264/0.077) for the Genevan and the Zurich model, respectively.

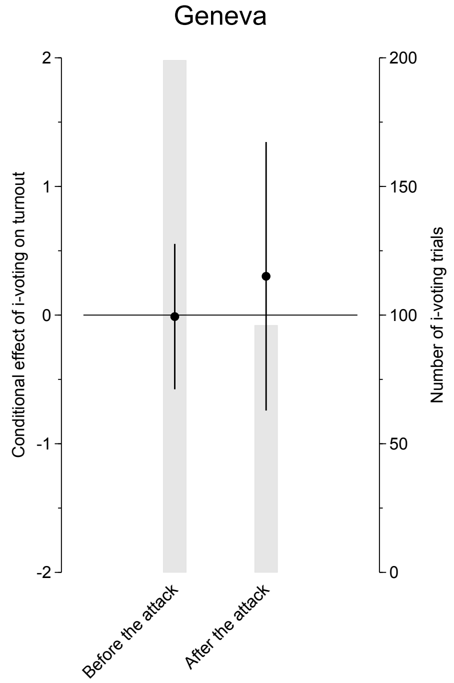
Table 3: Additional robustness checks

	(1) Geneva	(2) Zurich	(3) Zurich
I-voting			−0.332
			(0.54)
First three trials	−0.191	−0.895	
	(0.523)	(0.563)	
Other trials	0.016	−0.477	
	(0.266)	(0.676)	
Municipality FEs	✓	✓	✓
Referendum day FEs	✓	✓	✓
Quadratic municipality time trends	✓	✓	✓
Municipalities	45	171	169
Years	2001–2014	2001–2014	2001–2011
Observations	1980	5472	5408

Note: This table shows the results from two-way fixed effects regressions including municipality-level quadratic time trends. The outcome is turnout in federal referendums. Column 3 drops Winterthur and Zurich from the analysis. FEs stands for fixed effects. Standard error clustered at the municipality level in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

is that media attention was probably highest in the early years. Yet, it is also possible that novelty plays out independently of time, at least to some extent. Due to the stepwise implementation of i-voting in both cantons, there is significant variation as to the points in time when municipalities first trialled i-voting. Geneva's Carouge, for example, first trialled i-voting in 2004, while Aire-la-Ville's first trial was in 2010. It may be the case that novelty does not depend on time as such, but on the number of trials. Continuing with our example, novelty may have increased turnout in Carouge in 2004, and in Aire-la-Ville in 2010. To test for this, we coded two separate treatment indicators and reran the main models by plugging in the two alternative treatment indicators instead of the variable indicating all trials. The first alternative treatment indicator captures the first three i-voting trials in a municipality, while the second captures all other trials. Columns 1 and 2 of Table 3 give the results. Again, we find no evidence for a novelty effect with this alternative specification. Both indicators are statistically insignificant in both cantons. While the dummy indicating the first three trials yields a lower point estimate in both models, Wald tests suggest that the two variables are statistically indistinguishable in

Figure 3: Treatment effect before and after the 2013 hacking event (Geneva only)



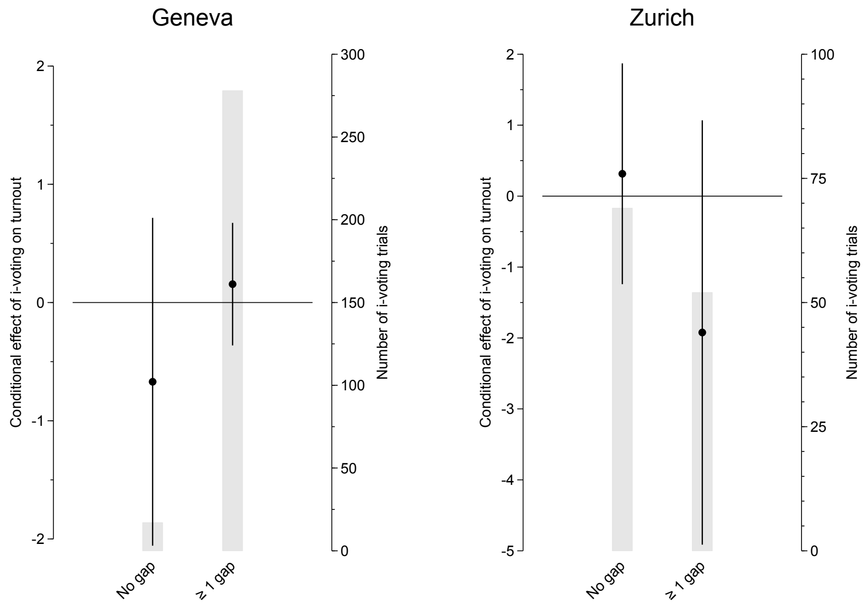
Note: The dots represent the effect of i-voting on turnout and the spikes the 95% confidence intervals. The underlying histogram gives the number of i-voting trials. The coefficient on the product term is 0.313 (clustered standard error=0.484).

either model.

Next, we explore an alternative opportunity to test for treatment effect heterogeneity related to changes in the salience of security issues. In June 2013, on the occasion of a hacking night in Paris, an IT specialist publicly demonstrated that software modelled after the Genevan i-voting solution can be manipulated by third parties. Rightly or wrongly, this strengthened the perception that the Genevan system is susceptible to third party manipulation. If security issues negatively affect i-voting's effect on turnout, we would therefore expect that the effect estimates before and after the hacking event differ. To test this, we interacted a dummy variable indicating the period after the hacking event with the treatment indicator. Figure 3 shows the treatment effect estimates before and after the event, which are nearly identical and both statistically indistinguishable from zero.

An additional potential source for treatment heterogeneity that we have not considered thus far is gaps in implementation. In most treated municipalities, the Internet channel

Figure 4: Treatment effect depending on presence of gaps



(a) (b)

Note: The dots represent the effect of i-voting on turnout and the spikes the 95% confidence intervals. The underlying histogram gives the number of i-voting trials. The coefficients on the product terms are 0.826 (clustered standard error=0.684) and -2.237 (1.906) for the Genevan and the Zurich model, respectively.

was not constantly available from the time it was first introduced. That is, there were gaps in implementation, with i-voting available at t1, unavailable at t2 (a gap), and then again available at t3. There are several reasons for the emergence of these gaps, including the suspension of i-voting in Geneva from 2005–2008, the rotation in 2008/2009, and simultaneous elections that the i-voting programs were unable to accommodate (see the data section in the paper for more details). Gaps raise the possibility that voters lose awareness of the option to cast their vote online. Thus the presence of gaps may dampen i-voting's effect on turnout. Note that the previously reported tests interacting the treatment indicator with time provide some initial evidence against this hypothesis, given that the presence of gaps is naturally related to time. A more direct way to test for possible treatment heterogeneity as a result of gaps involves interacting the treatment indicator with a dummy variable indicating whether a municipality experienced a gap or not. Figure 4 gives the results. We find no support for treatment heterogeneity due to gaps also with this tailored test: The effect of i-voting on turnout is statistically indistinguishable from zero irrespective of the presence of gaps.

Finally, we consider potential implications of two special cases: the municipalities of Winterthur and Zurich. Unlike in all other cases, i-voting was not made available throughout these two municipalities, but limited to selected districts (the old town in case of Winterthur, districts 1 and 2 in case of Zurich). This was done to comply with the federal rule that not more than 20% (later 30%) of a canton participate in the trials: Winterthur and Zurich are by far the canton's largest cities. Jointly, they make up about a third of the canton's population. In column 3 of Table 3 we investigate potential implications for our effect estimate by dropping both municipalities from our main model. The substantive conclusion that i-voting does not affect turnout remains unchanged.

References

Galiani, Sebastian, Paul Gertler & Ernesto Schargrodsky. 2005. "Water for Life: The Impact of the Privatization of Water Services on Child Mortality." *Journal of Political Economy* 113(1):83–120.

Rosenbaum, Paul R. & Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70(1):41–55.

Ryan, Andrew M., James F. Jr. Burgess & Justin B. Dimick. 2015. "Why We Should Not Be Indifferent to Specification Choices for Difference-in-Differences." *Health Services Research* 50(4):1211–1235.

Statistical Office of the Canton of Geneva. 2015. `http://www.ge.ch/statistique/` [October 15, 2015].

Statistical Office of the Canton of Zurich. 2015. `http://www.statistik.zh.ch/internet/justiz_inneres/statistik/en/home.html` [October 15, 2015].

Swiss Federal Statistical Office. 2015. `http://www.bfs.admin.ch/bfs/portal/en/index.html` [October 15, 2015].

Swiss Federal Tax Administration. 2015. `https://www.estv.admin.ch/estv/en/home.html` [October 15, 2015].