

Applied Statistical Analysis II

POP77003:

Replication Study

Marcus Ó Faoláin

“Reducing the Cost of Voting: an Evaluation of Internet Voting’s Effect on Turnout”

Nicole Goodman and Leah C. Stokes, 2018

As found in the British Journal of Political Science

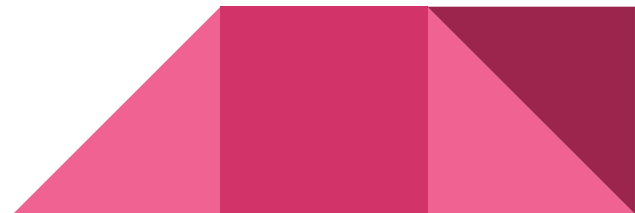
Sources:

1. Paper freely available at:

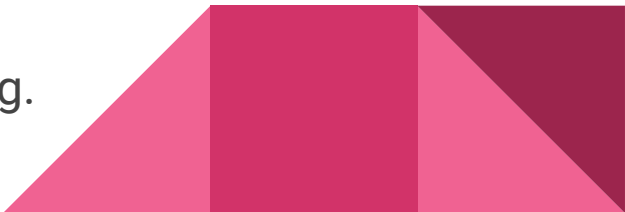
https://www.researchgate.net/publication/325197837_Reducing_the_Cost_of_Voting_An_Evaluation_of_Internet_Voting's_Effect_on_Turnout

2. Data available at:

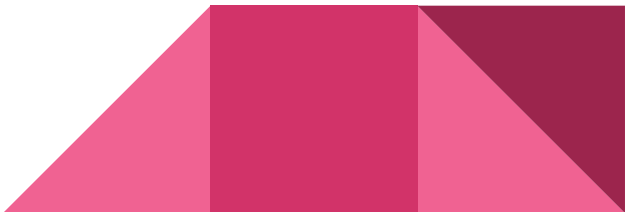
<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/KMMN1B>



What is this paper about?

1. Paper seeks to study the effect of internet voting on turnout in elections
 2. Common electoral reforms to increase voter turnout aim to decrease “voting costs”.
 3. Some examples of this include early voting, extended voting and postal voting
 4. This example looks at a case study of the effect of internet voting on elections in Ontario, Canada.
 5. “The technology was used in four election periods by ninety eight local governments, for a total of 173 elections.”
 6. Main outcome variable of the study: Voter turnout
 7. Main explanatory variable of the study: Internet voting.
- 

What is this paper about?

1. Municipal elections of mayor, deputy mayor, councilors and other representatives.
 2. Election years: 2000, 2003, 2006, 2010 and 2014.
 3. 2000 - baseline year without IV.
 4. Full population of ninety-eight Ontario municipalities that have used internet voting.
 5. 173/490 elections using internet voting since 2003
 6. Data gathered through contact of municipalities (up to five times), local news media and archives.
 7. Primary motives for implementation: improve voting accessibility and reduce voting costs.
- 

What model is used in this paper?

1. A fixed effects model
2. Fixed effects models estimate the relationship between the outcome and explanatory variables while also accounting for time specific or individual specific factors.
3. In this study, time-specific factors refer to years the elections took place.
4. Individual specific factors refer to the municipalities (geography) in which the internet voting took place.



Original Model

$$Y_{it} = \gamma_i + \delta_t + \alpha D_{it} + \beta X_{it} + \epsilon_{it}$$

Where:

Y_{it} = Voter turnout in a given municipality, i , in time, t

γ_i = municipality fixed effect

δ_t = election year fixed effect

α = treatment effect of internet voting

βX_{it} = matrix of covariate controls.



What were some potential confounding factors in this study?

1. Population (log)
2. Population density (log)
3. Unemployment rate
4. Median income (log)
5. Population with a university degree (%)
6. Immigrant population (% , log)



What data is included in the study and what are we hoping to replicate?

1. There are three tables in the paper:
 - a. Table 1, “Internet Voting’s effects on Turnout”:
 - i. Column 1: treatment effect of IV using fixed effects estimators with election year and municipal fixed effects.
 - ii. Column 2: model including linear time trend for each municipality.
 - iii. Column 3: Model including census variables as controls.
 - b. Table 2, “IV’s effect on turn out controlling for other electoral reforms, such as mail voting and elimination of paper ballots”:
 - i. Three columns as above
 - c. Table 3, containing one column and using a smaller dataset of just two years (2000 and 2014)
 - i. No online registration barrier
 - ii. Paper ballots eliminated
 - iii. Advance internet voting only.



Replicating the data: Table 1, column 1

TABLE 1 *Internet Voting's Effect on Turnout*

Internet voting (Treatment)	0.035*	0.036†	0.036*
	(0.016)	(0.019)	(0.016)
Population (log)			-0.195
			(0.178)
Population density (log)			0.362*
			(0.136)
Unemployment rate			0.002
			(0.003)
Median income (log)			0.013
			(0.088)
Population with university degree			0.131
			(0.204)
Population aged 65+ (%)			0.114
			(0.665)
Immigrant Population (% , log)			-0.019
			(0.037)
N	96	96	96
Fixed effects	Y	Y	Y
Unit-specific linear time trends	N	Y	N

Note: robust standard errors, clustered at municipal level. Intercepts are not reported.
Significant at † $p < 0.10$; * $p < 0.05$

```
> ## Table 1 column 1
> mod1 <- lm(turnout ~ intvoting + as.factor(csd_code) + as.factor(year), data=munis_dropped)
> round(coefest(mod1, vcov=vcovCluster(mod1, cluster=munis_dropped$csd_code)))[1:
5,], 3)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.401	0.013	3.122300e+01	0.000
intvoting	0.035	0.016	2.154000e+00	0.032
as.factor(csd_code)1011	0.025	0.000	1.349076e+13	0.000
as.factor(csd_code)1020	0.146	0.000	7.741402e+13	0.000
as.factor(csd_code)1030	0.053	0.000	2.800003e+13	0.000

1. Was the replication successful?

- Yes, we we were able to replicate the data from the paper.

Replicating the data: Table 1, column 2

TABLE 1 *Internet Voting's Effect on Turnout*

Internet voting (Treatment)	0.035*	0.036†	0.036*
	(0.016)	(0.019)	(0.016)
Population (log)			-0.195
			(0.178)
Population density (log)			0.362*
			(0.136)
Unemployment rate			0.002
			(0.003)
Median income (log)			0.013
			(0.088)
Population with university degree			0.131
			(0.204)
Population aged 65+ (%)			0.114
			(0.665)
Immigrant Population (%), log)			-0.019
			(0.037)
N	96	96	96
Fixed effects	Y	Y	Y
Unit-specific linear time trends	N	Y	N

Note: robust standard errors, clustered at municipal level. Intercepts are not reported.
Significant at † $p < 0.10$; * $p < 0.05$

```
> ### Table 1, column 2
> mod1 <- lm(turnout ~ intvoting + as.factor(muni_id) + as.factor(year) + as.factor(muni_id):year, data=munis_dropped)
> round(coeftest(mod1, vcov=vcovCluster(mod1, cluster=munis_dropped$csd_code)))[1:5,]
3)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	35.053	0.187	187.912	0.000
intvoting	0.036	0.019	1.895	0.059
as.factor(muni_id)m002	-51.933	0.849	-61.180	0.000
as.factor(muni_id)m005	-22.887	0.849	-26.962	0.000
as.factor(muni_id)m014	-6.737	0.187	-35.975	0.000

1. Was the replication successful?

- Yes, we were able to replicate the data from the paper.

Replicating the data: Table 1, column 3

TABLE 1 *Internet Voting's Effect on Turnout*

Internet voting (Treatment)	0.035*	0.036†	0.036*
	(0.016)	(0.019)	(0.016)
Population (log)			-0.195
			(0.178)
Population density (log)			0.362*
			(0.136)
Unemployment rate			0.002
			(0.003)
Median income (log)			0.013
			(0.088)
Population with university degree			0.131
			(0.204)
Population aged 65+ (%)			0.114
			(0.665)
Immigrant Population (% , log)			-0.019
			(0.037)
N	96	96	96
Fixed effects	Y	Y	Y
Unit-specific linear time trends	N	Y	N

Note: robust standard errors, clustered at municipal level. Intercepts are not reported.
Significant at † $p < 0.10$; * $p < 0.05$

```
> ### Table 1 column 3
> mod1 <- lm(turnout ~ intvoting + log(pop) + log(pop_den) + unemploy + log(median_
_inc) + p_65_plus + p_uni + log(p_imm) + as.factor(csd_code) + as.factor(year), da
ta=munis_dropped)
> round(coeftest(mod1, vcov=vcovCluster(mod1, cluster=munis_dropped$csd_code))[1:1
0,], 3)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.924	1.760	0.525	0.600
intvoting	0.036	0.016	2.211	0.028
log(pop)	-0.195	0.178	-1.097	0.273
log(pop_den)	0.362	0.136	2.663	0.008
unemploy	0.002	0.003	0.695	0.487
log(median_inc)	0.013	0.088	0.146	0.884
p_65_plus	0.114	0.665	0.172	0.864
p_uni	0.131	0.204	0.639	0.523
log(p_imm)	-0.019	0.037	-0.509	0.611
as.factor(csd_code)1011	-0.078	0.045	-1.720	0.086

1. Was the replication successful?

- Yes, we were able to replicate the data from the paper.

Important note about Table 1, column 3

TABLE 1 *Internet Voting's Effect on Turnout*

Internet voting (Treatment)	0.035*	0.036†	0.036*
	(0.016)	(0.019)	(0.016)
Population (log)			-0.195
			(0.178)
Population density (log)			0.362*
			(0.136)
Unemployment rate			0.002
			(0.003)
Median income (log)			0.013
			(0.088)
Population with university degree			0.131
			(0.204)
Population aged 65+ (%)			0.114
			(0.665)
Immigrant Population (% , log)			-0.019
			(0.037)
N	96	96	96
Fixed effects	Y	Y	Y
Unit-specific linear time trends	N	Y	N

Note: robust standard errors, clustered at municipal level. Intercepts are not reported. Significant at [†] $p < 0.10$; * $p < 0.05$

1. The paper states that coefficient for “population density (log)” is significant at $\alpha = 0.05$.
2. When the data is replicated, the p-value for this coefficient is equal to 0.008.
3. This means that it is in fact significant at $\alpha = 0.01$.

Replicating the data: Table 2, column 1

TABLE 2 *Internet Voting's Effect on Turnout, Controlling for other Electoral Reforms*

Internet voting (IV)	0.072*** (0.018)	0.072** (0.023)	0.074*** (0.018)
Vote by mail (VBM)	0.095*** (0.018)	0.096*** (0.028)	0.096*** (0.018)
Paper ballots eliminated (PBE)	0.008 (0.016)	0.007 (0.026)	0.007 (0.016)
<i>N</i>	96	96	96
Fixed effects	Y	Y	Y
Unit-specific linear time Trends	N	Y	N
Census covariates	N	N	Y

Note: robust standard errors, clustered at municipal level. Intercepts are not reported. Significant at ** $p < 0.01$; *** $p < 0.001$

```
> ##Table 2, column 1
> mod1 <- lm(turnout ~ intvoting + VBM + paper_ballots_eliminated + as.factor(csd_code) + as.factor(year), data=munis_dropped)
> round(coeftest(mod1, vcov=vcovCluster(mod1, cluster=munis_dropped$csd_code)), 3)
```

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.382	0.015	2.5769e+01	<2e-16 ***
intvoting	0.072	0.018	4.0700e+00	<2e-16 ***
VBM	0.095	0.018	5.3190e+00	<2e-16 ***
paper_ballots_eliminated	0.008	0.016	4.7800e-01	0.633
as.factor(csd_code)1011	0.028	0.006	4.4080e+00	<2e-16 ***
as.factor(csd_code)1020	0.127	0.004	3.5434e+01	<2e-16 ***
as.factor(csd_code)1030	0.032	0.004	9.0320e+00	<2e-16 ***

1. Was the replication successful?

- Yes, we were able to replicate the data from the paper.

Replicating the data: Table 2, column 2

TABLE 2 *Internet Voting's Effect on Turnout, Controlling for other Electoral Reforms*

Internet voting (IV)	0.072*** (0.018)	0.072** (0.023)	0.074*** (0.018)
Vote by mail (VBM)	0.095*** (0.018)	0.096*** (0.028)	0.096*** (0.018)
Paper ballots eliminated (PBE)	0.008 (0.016)	0.007 (0.026)	0.007 (0.016)
N	96	96	96
Fixed effects	Y	Y	Y
Unit-specific linear time Trends	N	Y	N
Census covariates	N	N	Y

Note: robust standard errors, clustered at municipal level. Intercepts are not reported. Significant at ** $p < 0.01$; *** $p < 0.001$

```
> ### Table 2, column 2
```

```
> mod1 <- lm(turnout ~ intvoting + VBM + paper_ballots_eliminated + as.factor(muni_id) + as.factor(year) + as.factor(muni_id):year, data=munis_dropped)
> round(coeftest(mod1, vcov=vcovCluster(mod1, cluster=munis_dropped$csd_code)))[1:5,] 3)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	19.730	4.579	4.309	0.000
intvoting	0.072	0.023	3.107	0.002
VBM	0.096	0.028	3.485	0.001
paper_ballots_eliminated	0.007	0.026	0.261	0.794
as.factor(muni_id)m002	-36.815	3.657	-10.067	0.000

1. Was the replication successful?

- Yes, we were able to replicate the data from the paper.

Replicating the data: Table 2, column 3

TABLE 2 *Internet Voting's Effect on Turnout, Controlling for other Electoral Reforms*

Internet voting (IV)	0.072*** (0.018)	0.072** (0.023)	0.074*** (0.018)
Vote by mail (VBM)	0.095*** (0.018)	0.096*** (0.028)	0.096*** (0.018)
Paper ballots eliminated (PBE)	0.008 (0.016)	0.007 (0.026)	0.007 (0.016)
N	96	96	96
Fixed effects	Y	Y	Y
Unit-specific linear time Trends	N	Y	N
Census covariates	N	N	Y

Note: robust standard errors, clustered at municipal level. Intercepts are not reported. Significant at ** $p < 0.01$; *** $p < 0.001$

```
> ### Table 2 column 3
> mod1 <- lm(turnout ~ intvoting + VBM + paper_ballots_eliminated + log(pop) + log
(pop_den) + unemploy + log(median_inc) + p_65_plus + p_uni + log(p_imm) + as.facto
r(csd_code) + as.factor(year), data=munis_dropped)
> round(coeftest(mod1, vcov=vcovCluster(mod1, cluster=munis_dropped$csd_code))[1:1
0,], 3)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.516	1.624	0.318	0.751
intvoting	0.074	0.018	4.120	0.000
VBM	0.096	0.018	5.189	0.000
paper_ballots_eliminated	0.007	0.016	0.422	0.673
log(pop)	-0.167	0.184	-0.908	0.364
log(pop_den)	0.333	0.156	2.133	0.034
unemploy	0.001	0.003	0.232	0.817
log(median_inc)	0.035	0.079	0.449	0.654
p_65_plus	0.287	0.565	0.508	0.612
p_uni	-0.019	0.199	-0.093	0.926

1. Was the replication successful?

- Yes, we were able to replicate the data from the paper.

Replicating the data: Table 3

TABLE 3 *Variation in Electoral Rules Estimated Effects on Internet Voting Use*

No Registration Barrier	0.383*** (0.056)
Paper Ballots Eliminated	0.113*** (0.027)
Advance Internet Voting Only	-0.101 (0.066)
N	96
Fixed Effects	Y
Census Covariates	Y

Note: outcome variable is percentage of voters using the internet to cast their vote. Robust standard errors, clustered at municipal level. Intercepts are not reported. Significant at *** $p < 0.001$

```
> #### ELECTORAL RULES - DIFF IN DIFF 2 YEAR ANALYSIS ####
> mod1 <- lm(p_int_votes ~ no_reg_barrier + paper_ballots_eliminated + adv_only +
  log(pop) + log(pop_den) + unemploy + log(median_inc) + p_65_plus + p_uni + log(p_
  imm) + as.factor(csd_code) + as.factor(year), data=munis)
> round(coeftest(mod1, vcov=vcovCluster(mod1, cluster=munis$csd_code)))[1:12,], 3)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.040	2.066	-0.987	0.326
no_reg_barrier	0.383	0.056	6.878	0.000
paper_ballots_eliminated	0.113	0.027	4.257	0.000
adv_only	-0.101	0.066	-1.532	0.129
log(pop)	0.201	0.271	0.742	0.460
log(pop_den)	-0.066	0.269	-0.246	0.807
unemploy	-0.003	0.004	-0.675	0.502
log(median_inc)	0.058	0.110	0.533	0.595
p_65_plus	-0.596	0.579	-1.029	0.306
p_uni	-0.565	0.262	-2.153	0.034
log(p_imm)	0.024	0.056	0.438	0.663
as.factor(csd_code)1011	-0.069	0.082	-0.842	0.402

1. Was the replication successful?

- Yes, we we were able to replicate the data from the paper.

Results of the paper:

1. Table 1: Adoption of internet voting increases turnout by around 3.5% in each of the three models.
2. Table 2: if VBM (vote by mail) is already in use, the effect of internet voting is reduced.
3. Table 3: Increasing barriers to internet voting decrease its use, eliminating paper ballots increases its use.
4. Registration barriers (having to register online to vote) decrease use of internet voting.



Our extension: testing an interaction model between “Internet voting” and “Population Density”

TABLE 1 *Internet Voting’s Effect on Turnout*

Internet voting (Treatment)	0.035*	0.036†	0.036*
	(0.016)	(0.019)	(0.016)
Population (log)			-0.195
			(0.178)
Population density (log)			0.362*
			(0.136)
Unemployment rate			0.002
			(0.003)
Median income (log)			0.013
			(0.088)
Population with university degree			0.131
			(0.204)
Population aged 65+ (%)			0.114
			(0.665)
Immigrant Population (% , log)			-0.019
			(0.037)
N	96	96	96
Fixed effects	Y	Y	Y
Unit-specific linear time trends	N	Y	N

Note: robust standard errors, clustered at municipal level. Intercepts are not reported. Significant at † $p < 0.10$; * $p < 0.05$

- There are two significant effects in column 3 of table 1: internet voting and population density (log)
- We want to add a “fourth column” to this table, in which the interaction between internet voting and population density is modelled.

Replication extension: interaction between internet voting and population density.

```
> # MÖF twist.
> # Testing for an interaction effect in the data.
> mod1 <- lm(turnout ~ intvoting*log(pop_den) + log(pop) + unemploy + log(median_inc) + p_65_plus + p_uni + log(p_imm) + as.factor(csd_code) + as.factor(year), data=munis_dropped)
> round(coefest(mod1, vcov=vcovCluster(mod1, cluster=munis_dropped$csd_code))[1:109,], 3)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.866	1.745	0.496	0.620
intvoting	0.045	0.028	1.575	0.116
log(pop_den)	0.360	0.132	2.718	0.007
log(pop)	-0.183	0.179	-1.021	0.308
unemploy	0.002	0.003	0.702	0.483
log(median_inc)	0.009	0.090	0.095	0.925
p_65_plus	0.090	0.662	0.135	0.893
p_uni	0.136	0.208	0.657	0.512
log(p_imm)	-0.018	0.038	-0.462	0.644
as.factor(csd_code)1011	-0.076	0.045	-1.690	0.092
as.factor(csd_code)1020	0.130	0.029	4.452	0.000
as.factor(csd_code)1030	0.017	0.038	0.451	0.652
as.factor(csd_code)1042	0.099	0.092	1.071	0.285

as.factor(csd_code)58075	0.907	0.229	3.953	0.000
as.factor(csd_code)60010	-0.248	0.143	-1.729	0.085
as.factor(year)2003	-0.019	0.017	-1.082	0.280
as.factor(year)2006	-0.030	0.027	-1.137	0.256
as.factor(year)2010	-0.026	0.043	-0.600	0.549
as.factor(year)2014	-0.050	0.060	-0.827	0.409
intvoting:log(pop_den)	-0.002	0.005	-0.455	0.650

1. Does the interaction between internet voting and population have an effect?
 - a. The p-value of the interaction is equal to 0.65.
 - b. With a significance level of 0.05, there is not enough evidence to reject the null hypothesis that the interaction is equal to zero.

Table 1, with “Column 4” of interactive model

Internet voting (treatment)	0.035* (0.016)	0.036† (0.019)	0.036* (0.016)	0.045 (0.028)
Population (log)			-0.195 (0.178)	-0.183 (0.179)
Population density (log)			0.362** (0.136)	0.360** (0.132)
Unemployment Rate			0.002 (0.003)	0.002 (0.003)
Median Income (log)			0.013 (0.088)	0.009 (0.090)
Population with university degree			0.131 (0.204)	0.136 (0.208)
Population aged 65+			0.114 (0.665)	0.090 (0.662)
Immigrant Population (% , log)			-0.019 (0.037)	-0.018 (0.038)
Internet voting × population density (log)				-0.002 (0.005)
N	96	96	96	96
Fixed effects	Y	Y	Y	Y
Unit-specific linear time trends	N	Y	N	N

** - $\alpha = 0.01$

* - $\alpha = 0.05$

† - $\alpha = 0.1$

Interpretation of our interactive model

1. There is not enough evidence to suggest that the interaction term we have added is necessary within this model, as we cannot conclude that the interaction coefficient is not statistically significant from zero.
2. The data is therefore adequately explained by an additive model.



Thank you for your attention!

