

# Does Internet Voting Bias Election Results? Evidence from Estonia

Kristjan Vassil\*

May 20, 2014

## Abstract

Aggregate election results from the past elections in Estonia demonstrate that some parties consistently gain more electronic votes than others. This observable difference in the allocation of e-votes and normal votes has sparked a fierce political debate on whether e-voting produces a political bias and, more importantly, whether it affects aggregate election outcomes. In this paper, we empirically assess whether the use of e-voting in Estonian elections induces political bias by structurally favoring some political parties over the others. We conceptualize political bias both at the aggregate as well as at the individual level, and offer three individual level mechanisms for empirical testing that can potentially produce a bias in the election results. Based on a wide range of analytical techniques, across various datasets and elections no evidence can be found that e-voting biases election results in any discernible fashion. The implications of these findings are relevant for policy makers, research community and political parties.

**This is an unpublished manuscript made available ahead of the publication in an academic journal due to the increasing interest toward the Estonian e-voting system and its likely political consequences.  
Please cite with author's permission.**

---

\*Institute of Government and Politics, University of Tartu. E-mail: kristjan.vassil@ut.ee; Web: [www.ut.ee/kristjan.vassil](http://www.ut.ee/kristjan.vassil). This research was supported by the Software Technology and Applications Competence Center and the Archimedes Foundation grant nr 3.2.1201.13-0018 "Verifiable Internet Voting – Event Analysis and Social Impact".

# 1 Introduction

Aggregate election results from the past elections in Estonia demonstrate that some parties consistently gain more electronic votes than others. For example, in the last two national elections the largest share of e-votes was won by the current prime-ministerial party, the Reform Party. The second-largest share of e-votes was gained by the Pro Patria and Res Publica Union, followed by the Social Democrats and the Greens. The smallest share of e-votes was gained by the center/left opposition party, the Center Party, which at the same time consistently wins the bulk of traditional votes (refer to Table 1). The observable difference in the allocation of e-votes and normal votes has sparked a fierce political debate on whether e-voting produces a mode-specific political bias and, more importantly, whether it affects election outcomes. Those against e-voting rely on aggregate evidence of the unequal vote share and posit that e-voting should be abolished because it structurally favors some parties over the others and therefore violates the assumption of political neutrality of a voting mode. Conversely, those winning larger shares of e-votes advocate for its convenience and potential to mobilize young voters suggesting that the continuation of e-voting is pivotal for democratic participation. In this paper we attempt to go beyond this normative debate and offer an evidence based empirical account to assess the potential of e-voting to bias aggregate election results.

## What is political bias?

Before doing so, however, we wish to conceptually clarify what is meant by *political bias* and how it relates to the vote shares gained by parties overall and by specific modes of voting.

In this paper, we define political bias as the difference between the observed election outcome and the outcome that we had observed in the absence of remote internet voting. The fundamental problem of causality in this case is that we only observe one of the two counterfactual situations at any given election. We cannot turn back time, look at two identical elections in which one election has and the other election has not used e-voting and then compare the two election results. We also cannot directly compare elections before the 2005 local elections without e-voting to those in or after 2005 with e-voting, because any discernible difference might occur due to time.

A way out of this inferential deadlock is to define an additional potential for political bias that may occur within the modes of voting: traditional paper-based voting and internet voting. Here, the bias is defined as the difference between the mode-specific vote share that parties gain and the aggregate election results. It follows, that vote shares by different modes of voting may produce political biases within these modes, but it does not follow automatically, that these mode-specific biases translate into biases in aggregate election results.

Most importantly, such a conceptualisation allows us to empirically validate whether political biases within the modes of voting occur and thus, have the potential to skew aggregate election results. A way to approach this research problem is to specify individual level mechanisms that can be explicitly tested and verified. This is precisely the research logic in the following analysis.

We proceed by first looking at aggregate distribution of vote share by parties and by the two modes of voting - e-voting and conventional election day voting. We then identify the three mechanisms that can potentially induce political bias and empirically test whether we

find supporting evidence to any of these.

## 2 Vote shares by modes of voting

We begin by looking at aggregate elections results in order to understand how much the distribution of e-votes deviates from the distribution of traditional on-paper votes for each of the parties in Estonia. To achieve this, we rely on results from the elections of 2007 and 2011, the two national elections where e-voting has been used. In order to evaluate whether the share of e-votes is representative of the traditional votes we calculate for each party  $i$  the difference  $d$  between the traditional vote share  $t$  and the e-vote share  $e$  that this party gained ( $d_i = t_i - e_i$ ) (cf. Haljasmets (2013)). If both vote shares are equal,  $d = 0$ . If the traditional vote share exceeds the share of e-votes for a particular party  $d > 0$  and conversely, if the share of e-votes is larger than the share of traditional e-votes  $d < 0$ . Table 1 displays the findings.

Table 1: The vote share by mode of voting in the 2007 and 2011 national elections

| Party                      | Vote share 2007 |        |      | Vote share 2011 |        |       |
|----------------------------|-----------------|--------|------|-----------------|--------|-------|
|                            | Paper           | E-vote | $d$  | Paper           | E-vote | $d$   |
| Reform Party               | 27.4            | 34.5   | -7.1 | 25.8            | 36.9   | -11.1 |
| Center Party               | 27.1            | 9.1    | 18   | 27.7            | 10     | 17.7  |
| Pro Patria and Res Publica | 17.4            | 26.7   | -9.3 | 18.9            | 25.4   | -6.5  |
| Social Democrats           | 10.5            | 13.3   | -2.8 | 16.8            | 18     | -1.2  |
| Peoples Party              | 7.3             | 3.6    | 3.7  | 2.4             | 1.3    | 1.1   |
| Greens                     | 6.9             | 10.7   | -3.8 | 3.6             | 4.3    | -0.7  |
| Others                     | 3.4             | 2.1    | 1.3  | 4.7             | 4.1    | 0.6   |

Raw data available at: <http://www.vvk.ee>

First, we see that the largest discrepancy between the vote shares of the two voting modes occurs for the Center party reaching a difference of close to 18 percentage points in both elections. Following our operationalization it tells us that Center Party’s paper-vote share is 18 percentage points larger than its e-vote share confirming the popular understanding that Center Party voters are more likely to prefer traditional means of casting a ballot than voting online. The second largest discrepancy can be observed for the Pro Patria and Res Publica Union and the Reform Party. Both these parties win larger shares of e-votes than on-paper votes. For other parties, the differences are negligible and remain below a few percentage points.

Based on the aggregate findings we clearly see that e-votes are not representative of the election outcome. Quite the contrary, the within voting-mode bias appears to be quite substantial in favor to Reform Party in 2011, Social Democrats in 2011, Pro Patria and Res Publica Union in 2007 and 2011 and Greens in 2007 and 2011. Yet, it is imperative that one does not confuse mode-specific bias with that at the level of aggregate election results. Imagine that some voters simply switch from normal voting to e-voting and that this switching is more probable among those who vote for the Reform Party. Although this non-random switching produces a mode specific bias, the aggregate election result still remains unaffected because these voters have only switched from one mode to another. They

have not switched their vote choice. It follows that making inferences about the 'biasness' on the basis of aggregate election results (as shown in Table 1) is potentially spurious. However, bias can still occur due to a number of mechanisms associated with the non-random usage of e-voting. Our next task is to explicate behavioral mechanisms that could potentially translate mode-specific bias into differences in election outcomes.

### 3 Mechanisms inducing bias in elections outcomes

We explicate three mechanisms that could potentially skew aggregate election results due to availability of internet voting. In so doing we lay out the building blocks of each mechanism so that they can be later subjected to empirical testing in order find evidence whether any of the mechanisms are actually at play in Estonian elections.

#### Mobilization

The first mechanism is based on the idea that if internet voting can potentially bias elections results it may be due to the fact that it makes some voters to vote who would not have voted otherwise. We call this process as e-voting's capacity to mobilize new voters. The basic logic of this process is that the electorate with e-voting is larger than it would have been in the absence of e-voting.

Recent evidence has shown that about one tenth of e-voters claim that they would not have voted if they had no availability to vote online (Trechsel and Vassil, 2011). Next, suppose that mobilization is non-random and that e-voting mobilizes supporters of certain political preferences structurally giving more voice to some parties while undermining the others. Provided that these conditions hold, i.e., non-random mobilization based on political preferences happens, elections with e-voting would produce different election outcomes than those without it. Although we cannot compare elections with and without e-voting at the same time, because the counterfactual situation is never observed, we can test whether those who are mobilized by e-voting structurally represent some parties and fail to represent others. We thus specify the following hypothesis:

H1: E-voting mobilizes new voters who structurally prefer some parties over the others

#### Vote switching

The second mechanism is based on the idea of vote switching. Researchers have found that some voters change their vote choice between elections, preferring some parties in some elections and other parties in other elections (Solvak, 2011). It has been showed that vote switching is particularly dominant between elections at different levels, i.e., national, European or local with many voters using European Parliament elections as a means to signal their closest parties their discontent and warn them against what would happen in the next national elections. However, vote switching also occurs between the same-level elections. Particularly the young, with less stable party preferences are more likely to change their vote choice between two parliamentary elections. It may also be possible that the sheer availability of e-voting induces vote switching. Although not particularly plausible in reality, this is a logically possible outcome that can be tested empirically.

To be sure, e-voting may bias election results if e-voters start voting differently than they would have voted in the absence of e-voting and that this change, again, is not-random. It follows, that if e-voting induces switching, some parties may gain the competitive advantage over the others (provided that switching is not random). Thus, we propose the following hypothesis.

H2: The availability of e-voting induces non-random vote switching for some parties

### **Mode-specific bias**

Finally, we seek to test whether there is a mode-specific political bias in the first place. As we have explained and showed above at the aggregate level, political skew can occur within specific modes of voting. For example, this was found to be true within the mode of e-voting where some parties consistently gained more votes than others. Thus, if e-voting is indeed politically biased, it should be predictive of partisan choice at the individual level. I.e., participation in online voting should be associated with higher propensities to vote for certain political parties. However, when testing for such associations we need to look beyond bi-variate relationships and include relevant controls accounting for different propensities to use internet voting in the first place. Why?

Not all people are equally likely to vote online, because they lack technical skills, physical access, time, experience and other prerequisites for online voting. For example, young voters with higher socio-economic status may be more likely to vote online than others. Moreover, the same set of characteristics that predicts internet voting at the individual level, may be predictive of partisan choice. In other words, young people with higher socio-economic status are more likely to lean to right of the political spectrum and support the Reform Party; and they are also more likely to vote online. Therefore, analyzing the impact of e-voting on party choice contains two mechanisms: 1) the propensity to vote online as a function of party choice and 2) the propensity to vote online as a function of a set of socio-demographic characteristics that, at the same time, is predictive of partisan choice.

The key analytical objective for us is to disentangle the two mechanisms and demonstrate whether voting online is associated with political choices in either of the two mechanisms. The common strategy for separating the two is to control for relevant characteristics such as age, education and most importantly, the position on the left-right political spectrum. In so doing we are able to address the question to which extent political preference matters with respect to voting online and, more importantly, to which extent baseline socio-demographic characteristics determine the choice to vote online. Thus, we specify the third hypothesis to test for this relationship.

H3: The usage of e-voting is predictive of party choice even when controlling for a number of socio-demographic and attitudinal characteristics

Notice that we have formulated our hypotheses so that if confirmed we find evidence to support the view that election outcomes with and without internet voting probably would yield different outcomes. Naturally, depending on the size of the effects found, the question remains whether the difference would be discernible. In the following section we provide empirical tests for all three hypotheses.

## 4 Data and context

We draw inferences from Estonia, the only country in the world that has offered remote internet voting to its citizens since 2005. After regaining independence in 1991, Estonia has held six parliamentary elections to its 101-seat parliament. Using an open list proportional electoral system with multimember districts, the country is a consolidated democracy with turnout levels and party system stability comparable to those of European polities. Since 2004 Estonia is a full member of European Union and NATO; in 2011 Euro was adopted as country's currency. Estonia is widely recognized as a global leader in adopting information and communication technologies both in public as well as in private sector (for further details refer to [Trechsel and Vassil \(2011\)](#); [Vassil and Weber \(2011\)](#)).

We primarily make use of two datasets stemming from different research projects: Estonian E-Voting Surveys (Trechsel & Vassil, 2011) and the Estonian module of the Comparative Study of Electoral Systems of 2011 (CSES, 2011).

### **Estonian internet voting survey**

We employ data from the Estonian Internet Voting Survey of 2011 [Trechsel and Vassil \(2011\)](#). It is a CATI-based post-election survey conducted continuously since the 2005 local elections.<sup>1</sup> The survey takes place in the immediate aftermath of Elections, typically starting the following Monday after the Election Day. The survey is representative with regard to basic demographic characteristics, such as age, place of residence and language. However it uses a quota sample for interviewing internet voters, traditional voters and those abstaining from elections in approximately three equal groups. The reasons for using a quota sample stems from 2005 when the number of e-voters in the population was just under 1 per cent so that the random probability sample would not have been feasible in order to examine the profile of internet voters. To remedy the situation a quota sample was used to select respondents by the mode of voting, i.e., voters who fill in the paper ballot in the ballot station, voters who vote over the internet and eligible citizens who abstain from elections. In order to preserve methodological comparability, quota sampling was used in all of surveys, including that of 2011. Other than setting quotas by voting mode, random sampling was used to interview respondents from all groups.

---

<sup>1</sup>Estonian Internet Voting Survey has surveyed e-voters after the 2005 local elections, 2007 national elections, 2009 elections to the European Parliament, 2009 local elections and 2011 national elections.

Table 2: Proportion of voter types in the sample

| Voter type   | 2005         | 2007         | 2009         | 2009         | 2011           | Total          |
|--------------|--------------|--------------|--------------|--------------|----------------|----------------|
|              | Local        | National     | EP           | Local        | National       |                |
| Normal voter | 318<br>33.9  | 365<br>37.3  | 278<br>28.0  | 337<br>33.7  | 333<br>33.1    | 1,631<br>33.2  |
| E-voter      | 315<br>33.6  | 367<br>37.5  | 400<br>40.3  | 328<br>32.8  | 335<br>33.3    | 1,745<br>35.5  |
| Non-voter    | 305<br>32.5  | 247<br>25.2  | 315<br>31.7  | 334<br>33.4  | 339<br>33.7    | 1,540<br>31.3  |
| Total        | 938<br>100.0 | 979<br>100.0 | 993<br>100.0 | 999<br>100.0 | 1,007<br>100.0 | 4,916<br>100.0 |

## Comparative Study of Electoral Systems – Estonian module of 2011

CSES 2011 was conducted as a post-election survey in Estonia in the immediate aftermath of the parliamentary elections held on March 6, 2011. The survey uses a random probability sample and involves 1000 respondents. Interviews were conducted using face-to-face interviews. The sample is representative with respect to Estonian voting eligible citizens. Response rate was 61%. CSES is a large international consortium making use of national elections studies and making them available for comparative research purposes. For further details refer to [CSES \(2011\)](#).

Table 3: Proportion of voter types in the CSES sample compared to the population

| Voter type   | CSES sample | Population |
|--------------|-------------|------------|
|              | 2011        | 2011       |
| Normal voter | 603         | 439500     |
| %            | 61.3        | 48.1       |
| E-voter      | 177         | 140764     |
| %            | 18.0        | 15.4       |
| Non-voter    | 204         | 333082     |
| %            | 20.7        | 36.5       |
| Total        | 984         | 913346     |
| %            | 100.0       | 100.0      |

## 5 Empirical tests

### Mobilization

Mode specific political bias can also occur when a particular mode of voting (e.g., remote internet voting) mobilizes new voters with certain political preferences. For example, if e-voting is more popular among those with higher socio-economic status and if it indeed mobilizes new voters from that particular group, these newly mobilized voters may lean to

the right of the political spectrum, thus increasing the vote share of liberal parties which would not have happened in the absence of e-voting. To make it explicit, we are not arguing that this is the case. Whether this is the case remains, again, an empirical question that we can investigate with the help of data of the Estonian e-voting project (see section 4).

In order to estimate the voting mode specific potential for political bias we first need a variable that measures mobilization. In our surveys we have used the following survey item: “Would you have voted if you did not have had the chance to vote online?” Response categories include “I would have voted anyway”, “I would have rather voted”, “I would have rather not voted”, “I would have not voted for sure”. Based on this question, we recode a dummy-variable taking the value 1 if a voter “would not have voted for sure” or “would rather not have voted” and 0 otherwise. Following this operationalization voters who are coded 1 are those who would not have voted in the absence of e-voting and thus, we can assume that they were mobilized to vote (refer to Table 4).

Table 4: Frequency of e-voters mobilized to vote

| Mobilization  | Frequency | Percent |
|---------------|-----------|---------|
| Not mobilized | 281       | 85.4%   |
| Mobilized     | 48        | 14.6%   |
| Total         | 329       | 100.0%  |

However, we would like to render caution in taking responses to this question at face value and interpreting the subsequent variable as mobilization in the strongest meaning of the word. Because people are usually bad judges of their own behavior, it is likely that more often than not our respondents have given ‘yes’ or ‘rather yes’ answers to this question not because they were mobilized, but because of other unobserved reasons. Therefore, it is more appropriate to treat this variable as a proxy to measure mobilization and interpret it as a self-reflection of potential mobilization. In any event, because true mobilization is lower than our survey measure, our analysis will only overestimate the potential political bias among those being mobilized and therefore offers a conservative estimate on the potential bias.

Our next task is to specify a model that would reveal whether voter mobilization predicts partisan choice. If e-voting mobilizes voters with certain political preferences, then the mobilization-dummy should be associated with partisan choice. Conversely, if mobilized e-voters come from the entire political spectrum, its effect will be heterogeneous lending support for the argument that e-voting is not a politically biased mode of voting. If so, we have all the necessary building blocks of a model that can be specified in order to estimate the potential bias.

In order to assess the relationship between the mobilization and the party choice we use data from the 2011 national election and estimate four logit models with the dependent variables specified as discrete choices for each of the parties currently in the parliament ( $y_i$ ) – respondents who voted for the Center Party are coded 1 and 0 otherwise for the first dummy; respondents who voted for the Reform Party are coded 1 and 0 otherwise for the second dummy; respondents who voted for Social Democrats are coded 1 and 0 otherwise for the third dummy; and finally, respondents who voted for Pro Patria and Res Publica Union are coded 1 and 0 otherwise for the fourth dummy. Additionally to our mobilization-dummy,

we also include socio-demographic, behavioral and attitudinal controls, such as age (running from 18 to 96), gender (1 male, 0 females), higher education (1 higher education, 0 otherwise), income (continuous ranging from 1 through 11), PC-literacy (5 categories ranging from no PC literacy (1) to excellent PC literacy (5), treated as continuous), political activity (1 if the respondent has never participated in previous elections, 2 if in some, 3 if in most and 4 if in all elections where she has been eligible to vote), trust toward the system of e-voting (four categories from no trust to full trust, treated as continuous), left-right position (continuous from 0 to 10 where 0 means left and 10 right). The generic form of our model takes the following form:

$$y_i = \beta_0 + \beta_1 \text{mobilization} + \gamma + \epsilon \quad (1)$$

The parameter of interest is  $\beta_1$  which, if significant at least at 0.05 level, will tell us that new voters are structurally mobilized from the supporters of these political parties for which the parameter is significant. Findings are presented in Table 5.

Table 5: Party choice and mobilization

|                     | Model 1<br>Center Party | Model 2<br>Reform Party | Model 3<br>Social Democrats | Model 4<br>Pro Patria |
|---------------------|-------------------------|-------------------------|-----------------------------|-----------------------|
| Mobilization        | -0.02<br>(0.04)         | 0.02<br>(0.14)          | 0.09<br>(0.08)              | -0.07<br>(0.10)       |
| Age                 | -0.00<br>(0.00)         | -0.01 * **<br>(0.00)    | 0.00<br>(0.00)              | 0.00 * *<br>(0.00)    |
| Male                | 0.01<br>(0.02)          | 0.07<br>(0.08)          | -0.08<br>(0.06)             | 0.01<br>(0.06)        |
| Higher education    | -0.01<br>(0.02)         | -0.15*<br>(0.08)        | -0.01<br>(0.06)             | 0.17 * *<br>(0.07)    |
| Income              | -0.01<br>(0.01)         | 0.02<br>(0.02)          | -0.01<br>(0.02)             | -0.00<br>(0.02)       |
| PC literacy         | 0.00<br>(0.01)          | -0.11 * *<br>(0.05)     | 0.03<br>(0.04)              | 0.04<br>(0.04)        |
| Political activity  | -0.01<br>(0.02)         | -0.06<br>(0.09)         | 0.08<br>(0.06)              | 0.04<br>(0.06)        |
| Trust               | -0.03 * *<br>(0.01)     | 0.12*<br>(0.07)         | 0.05<br>(0.05)              | -0.03<br>(0.04)       |
| Left-right position | -0.01*<br>(0.00)        | 0.08 * **<br>(0.02)     | -0.05 * **<br>(0.01)        | 0.02*<br>(0.01)       |
| Observations        | 217                     | 217                     | 217                         | 217                   |
| Pseudo R-squared    | 0.19                    | 0.22                    | 0.15                        | 0.13                  |

Marginal effects at means; Standard errors in parentheses

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Our first and foremost interest with Table 5 lies with the first row that contains the variable *mobilization*. If the effect of mobilization is significant (designated with an asterisk)

it would tell us that those who are mobilized to vote due to the availability of e-voting are likely to vote for or against (depending on the sign of the effect) a political party that is used as dependent variable in a given model.

The results from these models are immediately apparent. We find that mobilization is not predictive of any party choice in all four models. It shows that there is no statistically significant relationship between the mobilization of new voters and their political choices. In other words, mobilization that occurs due to the presence of e-voting is random with respect to partisan preferences when controlling for all relevant covariates. We argue that this is the key to understand the mechanism by which internet voting functions with regard to partisan choice. As shown in this model left-right self position is significant predictor of party choice with corresponding signs in expected directions. The effect is negative for Center party and Social Democrats showing that their supporters lean to the left of the political spectrum; positive for the Reform Party and Pro Patria and Res Publica Union showing that their supporters lean to the right of the political spectrum (refer to previous section on how left-right position was measured and operationalized). Effectively it shows that one's political preference is the first and foremost driver of party choice and not the mobilization mechanism of e-voting.

We thus infer from this exercise that our first hypothesis of the mode specific political bias due to mobilization does not hold, irrespective of the fact that as many as 14.6 percent of e-voters felt that the availability of e-voting induced them to vote. Our model shows that these voters, when controlling for their political preferences, are mobilized on a random basis with respect to party choice. They are a heterogeneous group of voters who support all four political parties with no structurally identifiable pattern.

## Switching

Political bias occurring from the mode of voting can occur, because some voters non-randomly choose to switch their partisan preferences. In other words, if Reform Party voters choose to start voting for other parties because of the availability of e-voting, we might say, that this mode of voting induces people to switch their vote choice for another party. Although this is not very plausible in reality, we know that vote switching appears as a feature inherently linked to Estonian elections for a number of individual-level reasons (Solvak, 2011). If vote switching happens at random, no bias can be produced because switchers randomly change their vote choice to any of the political parties available. If, by contrast, switching has a pattern we might find that switching to some parties is more likely than others, potentially leading to a different election outcome.

Using the CSES data (refer to section 4) we code a dummy variable measuring whether the voter voted differently in 2011 national elections as compared to 2007 national elections. We call this variable switching and it takes the value 1 if a voter changed her vote choice between the two last national elections and 0 if she did not. Next, we compare the frequency of switchers by the mode of voting, i.e., among e-voters and among traditional voters in order to verify whether there is a statistically significant difference between the two groups. Data show that 35 percent of traditional voters and 28 percent of e-voters switch from one party choice to another between the two elections. The difference is 7 percentage points and it is not statistically significant. On the contrary, findings show that vote switching is by 7 percentage points less frequent among e-voters than among traditional voters. Thus, even if switching mechanism would be plausible hypothetically, we find no evidence for it

empirically.

Notice, that in the Estonian bi-lingual society switching is predominantly a feature of native Estonians, while Russian speaking minority is considerably less likely to change their party choice between elections (Solvak, 2011). In order to control for the potential confounding effect of language, we also tested the relationship for the two subsamples separately (Estonian versus non-Estonian speaking). No significant effect was found even when controlling for the language.

From this brief exercise we find little support for the mechanism by which a particular voting channel biases elections results because voters who use this channel switch their vote choice and therefore we refute our second hypothesis.

### Mode-specific bias

As explicated above our empirical strategy needs to account for two mechanisms that may potentially induce political bias due to the availability of e-voting. Let us start with a simple proposition. If e-voting is indeed politically biased, we should be able to observe, at the individual level, the relationship between the usage of e-voting and party choice. Refer to association  $a$  between e-voting and party choice in Figure 1. In simple terms it simply states that one's likelihood to use internet voting is associated with her party choice. For example, people who are likely to vote online are likely to vote for the Reform Party.

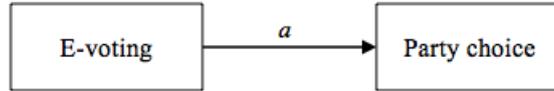


Figure 1: Relationship between e-voting and party choice

At the same time we know that party choice is first and foremost determined by one's socio-demographic profile. As explained above, young people with higher income and Estonian ethnicity support Reform Party more frequently than the Center Party. However, the same set of characteristics induce the probability of e-voting for reasons explained earlier. In order to account for this we expand our simple proposition by demographic determinants and link them to both - e-voting and party choice. Figure 2 links the usage of e-voting to party choice (link  $a$ ), but it also links demographic profile to party choice (link  $b$ ) and the probability of e-voting (link  $c$ ). Our empirical goal is to verify whether  $a$  prevails irrespective of controlling for  $b$  and  $c$ . Consequently, we can extend the model by one's political preferences and argue that while they affect party choice, they might, through demographics, also induce the probability of e-voting. These confounding patterns are of particular interest for us, because e-voting can only produce the bias in election results when indeed it is predictive of party choice irrespective of relevant confounders.

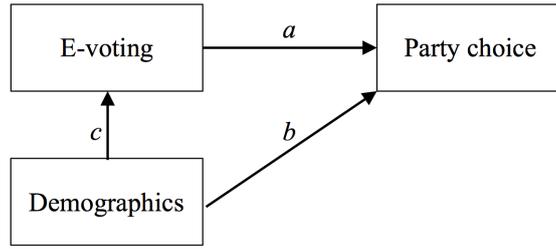


Figure 2: Relationship between e-voting and party choice

Using the CSES data (refer to section 4) we propose the following modeling strategy: we specify a logit model where the dependent variable is operationalized as a party choice (e.g., a binary variable 1 for respondents who voted for the Reform Party and 0 otherwise) and the independent variables is operationalized as dichotomy between e-voters (coded 1) and others (coded 0). Because we are interested in the effect of e-voting while controlling for demographics and other related covariates, we include a set of traditional determinants of party choice (such as age, gender, education, ethnicity, place of residence, income, left-right-placement and internet usage). Subsequently we estimate the chances of voting for each of the four parliamentary parties. A generic form for our model takes the following form:

$$y_i = \beta_0 + \beta_1 \text{evoting} + \gamma + \epsilon \quad (2)$$

where the primary parameter of interest is  $\beta_1$  while controlling for the vector of socio-demographic, attitudinal and behavioral covariates ( $\gamma$ ). If our primary parameter of interest is statistically significant, e-voting is associated with voting for this particular party. Results are displayed in Table 6, where columns contain four separate models each estimated for each of the parties.

Table 6: Relationship between party choice and e-voting

|                     | Model 1<br>Center Party | Model 2<br>Reform Party | Model 3<br>Social Democrats | Model 4<br>Pro Patria |
|---------------------|-------------------------|-------------------------|-----------------------------|-----------------------|
| E-voter             | -0.07<br>(0.06)         | 0.07*<br>(0.03)         | -0.04<br>(0.04)             | -0.04<br>(0.04)       |
| Age                 | 0.00<br>(0.00)          | -0.00<br>(0.00)         | -0.00<br>(0.00)             | -0.00<br>(0.00)       |
| Male                | 0.06<br>(0.04)          | -0.05*<br>(0.03)        | -0.03<br>(0.03)             | 0.01<br>(0.03)        |
| Higher education    | 0.01<br>(0.05)          | -0.07*<br>(0.03)        | 0.05<br>(0.04)              | 0.06<br>(0.04)        |
| Estonian speaker    | -0.49***<br>(0.06)      | 0.31***<br>(0.07)       | 0.34***<br>(0.06)           | 0.29***<br>(0.07)     |
| Urban               | 0.02<br>(0.04)          | 0.03<br>(0.03)          | -0.06*<br>(0.04)            | 0.03<br>(0.04)        |
| Income              | -0.02*<br>(0.01)        | 0.02***<br>(0.01)       | -0.01*<br>(0.01)            | -0.00<br>(0.01)       |
| Left-right position | -0.05***<br>(0.01)      | 0.04***<br>(0.01)       | -0.04***<br>(0.01)          | 0.04***<br>(0.01)     |
| Internet users      | -0.07<br>(0.05)         | -0.03<br>(0.04)         | 0.09*<br>(0.05)             | 0.00<br>(0.05)        |
| Observations        | 530                     | 530                     | 530                         | 530                   |
| Pseudo R2           | 0.59                    | 0.27                    | 0.17                        | 0.17                  |

Marginal effects at means; Standard errors in parentheses

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

The findings show that e-voting has no statistically significant effect in explaining the party choice for Center Party, Social Democrats and Pro Patria and Res Publica Union after controlling for a set of relevant characteristics of voting behavior. The only exception is the Reform Party in which case we see that e-voters are about 7 percentage points more likely to vote for the Reform Party than traditional voters even when controlling for relevant covariates.

Albeit its low statistical significance this finding merits further attention. Is this one significant difference a cause of concern when evaluating the individual level bias of internet voting? Think of our problem set in experimental research terms where e-voting can be conceptualized as treatment and traditional voting as control. In an experiment we would inquire whether a causal link between e-voting and party choice exists. In an experimental set-up randomization of treatment assignment would ensure that the groups compared are similar to each other in all relevant covariates except of treatment status (usage of e-voting vs non-usage of e-voting). As we make use of observational data, people have self-selected themselves into modes of voting on the basis of their socio-demographic properties and other qualities relevant for making this decision (including political preference). Thus, our current sample is structurally skewed in a number of relevant variables that, if different across two groups, would still show the effect of e-voting on party choice. That this is the case becomes apparent when consulting Table 7 first three columns (referred to as Raw data). In this table we observe that e-voters are younger with higher education, of urban residence

and have a higher income, and most notably, are frequent internet users. If the treatment and control samples are not balanced, differences in voting behavior could be reflection this unbalancedness, rather than the treatment itself (Lassen, 2005).

A conventional way in alleviating the problem of diverse subsamples is to use matching methods (Stuart and Rubin, 2008). Essentially, matching tries to recover a random research design from observational, or imperfectly randomized, data to provide a basis for causal interpretation of the estimates (Lassen, 2005, p. 109). The idea of matching is to match to every respondent in the treatment condition (e-voters) and respondent in the control condition (traditional voting) on the basis of observed covariates. In so doing, matching methods emulate randomization by selecting a reference group that is statistically no different from the treatment group and thereby approximates randomization in experimental designs. Thus, the matching procedure ensures that inference is not based on treatment and control samples that are too different (refer to Table 6), which is not guaranteed by simply including these variables as controls in a regression in the standard fashion (Lassen, 2005, p. 109).

In the following we make use of genetic matching that uses a search algorithm to iteratively check and improve covariate balance, and it is a generalization of propensity score and Mahalanobis Distance matching (Diamond and Sekhon, 2005). In essence, it is a multivariate matching method that uses an evolutionary search algorithm developed by Sekhon and Mebane (1998) to maximize the balance of observed covariates across matched treated and control units (Diamond and Sekhon, 2012). Relevant observed variables used for matching are age, gender, education, urban residence, ethnicity, income, left-right self-position and internet usage. Sample balance and discarded observations are reported in Table 7 (refer to last two columns).

Table 7: Comparing balance before and after matching

|                     | Raw data |              | Matched data |              |
|---------------------|----------|--------------|--------------|--------------|
|                     | E-voter  | Normal voter | E-voter      | Normal voter |
| Age                 | 40.20    | 50.33        | 40.20        | 40.33        |
| Male                | 0.38     | 0.42         | 0.38         | 0.38         |
| Higher              | 0.49     | 0.25         | 0.49         | 0.49         |
| Urban               | 0.73     | 0.61         | 0.73         | 0.73         |
| Estonian speaker    | 0.86     | 0.78         | 0.86         | 0.86         |
| Income              | 8.00     | 5.88         | 8.00         | 7.97         |
| Left-right position | 5.95     | 5.53         | 5.95         | 5.95         |
| Internet user       | 0.92     | 0.57         | 0.92         | 0.92         |
| Observations        | 160      | 565          | 160          | 114          |

160 treated units matched to 114 control units, 451 control units discarded.

Entries are mean values for given variables.

After preprocessing the data we continue by specifying exactly the same model as that specified in Equation 2, only now we estimated the model on the balanced sample and thus, our results provide basis for causal interpretation. Findings are reported in Table 8. Because observed covariates were used as a basis of matching they are not directly interpretable with

respect to the outcome of interest. We have included them into the model as the convention prescribes, but refrained from reporting them in the results table.

Table 8: Causal effect of e-voting on party choice

|              | Model 1<br>Center Party | Model 2<br>Reform Party | Model 3<br>Social Democrats | Model 4<br>Pro Patria |
|--------------|-------------------------|-------------------------|-----------------------------|-----------------------|
| E-voter      | -0.81<br>(0.61)         | 0.42<br>(0.38)          | -0.37<br>(0.37)             | -0.12<br>(0.33)       |
| Covariates   | <i>Included</i>         | <i>Included</i>         | <i>Included</i>             | <i>Included</i>       |
| Intercept    | 2.53<br>(1.98)          | -5.80<br>(1.50)         | -1.20<br>(1.28)             | -4.01<br>(1.30)       |
| Observations | 530                     | 530                     | 530                         | 530                   |

Logit coefficients; Standard errors in parentheses

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Quite clearly, findings from Table 8 show that e-voting dummy has no statistically significant impact on any of the party choices including the one for the Reform Party. We believe that this exercise produces a reliable and unbiased estimate of the effect of e-voting and we therefore conclude that electronic voting at the individual level does not cause voters to structurally prefer some parties over the others. Instead, the differences observed in the aggregate levels and in naive estimation models are due to the inherent imbalances in terms of observed characteristics between those voting on- and offline.

## 6 Discussion and concluding remarks

In this paper we have tested three individual level mechanisms that, if found true, would bias aggregate election outcome. Through a series of empirical tests we found no evidence that any of these mechanisms would substantially entice voters with certain political preferences. Based on these evidence we have thus no reason to infer that remote internet voting has biased the election outcome in the period under scrutiny in a discernible fashion. Yet, how to reconcile these individual level null-findings with the aggregate vote share distribution by parties? Why do some parties consistently gain more e-votes than others? And how is it possible that irrespective of the observed skew in aggregate results, no marked bias occurs in the election outcome?

In order to respond to these concerns two outcomes need to be considered that are relevant for voting behavior. The first is a decision to vote for one of the parties competing in an election, i.e., vote choice. Although not exclusively, but often voters act based on their political preferences. Whether short- or long-term, whether more or less responsive to policy changes, media influence, party leadership, political events, etc., voters base their decision to vote for parties on their core political preferences. Even in the event of strategic voting, to some degree political preferences determine one's choices. To put it explicitly, political preferences determine people's vote choice. The second aspect of political behavior is the decision to participate in elections, i.e., decision to turn out. Decision to vote can be, and often is, a matter of principle - some voters choose to participate in elections at all

costs, irrespective of unsatisfactory party performance or a rainy election day. Similarly, some voters may abstain as a matter of principle. Others may participate in elections occasionally, depending on the available choices, past party performance, whether conditions or else.

We argue, that the availability of voting modes – advance voting in the ballot station, postal voting, election day voting, as well as internet voting during the advance voting period – is a feature of the electoral system that can first and foremost affect decision to participate in elections. Consequently, we find it difficult to argue that the mode of voting, and e-voting in particular, can affect one’s political preferences and therefore impact on one’s vote choices. The latter is logically possible, but based on the findings of this report, empirically implausible.

We suggest, that the mechanism behind the unequal share of e-votes by different parties is, in fact, unrelated to political preference *per se*. Instead participation in e-voting requires its users to have certain skills and resources. For example, internet voters are likely to have high computer literacy, they might be young citizens with relatively high socio-economic status. These characteristics determine their frequency and habits of internet usage, as they define a myriad of other aspects of their everyday behavior. At the same time, these characteristics co-vary with their political preferences and consequently are predictive of the choices they make in political elections. Should e-voting become available as a new means to cast a vote, these voters switch from traditional means of casting a ballot to online voting, because they have higher propensity to use new technologies, not because they prefer certain political parties. The cause for e-voting, thus, is their propensity to use new technologies and *not* their political preferences. However, the latter becomes accentuated when only looking at the aggregate election results by mode of voting by parties, whereas the individual level mechanism driving the aggregate view remains uncovered. In this paper we have approached and scrutinized these individual level mechanisms with a number of analytical techniques, across various datasets and elections, and found no evidence that e-voting as a mode of voting suffers from a political bias at the individual level. Therefore, we find it difficult to argue that it is biased at the aggregate level, too.

## References

- CSES (2011). Comparative study of electoral systems. Data available at: <http://www.cses.org>.
- Diamond, A. and Sekhon, J. S. (2005). Genetic matching for estimating causal effects: a general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics*.
- Haljasmets, K. (2013). E-valimiste poliitiline kallutatus. Bachelor thesis. Defended in the Institute of Government and Politics, University of Tartu.
- Lassen, D. D. (2005). The effect of information on voter turnout: Evidence from a natural experiment. *American Journal of Political Science*, 49(1):103–118.
- Sekhon, J. S. and Mebane, W. R. (1998). Genetic optimization using derivatives. *Political Analysis*, 7(1):187–210.
- Solvak, M. (2011). Parteivaliku vahetamine 2011. aasta valimistel. In Toomla, R., editor, *Riigikogu Valimised*. Tartu: Tartu University Press.
- Stuart, E. A. and Rubin, D. B. (2008). Best practices in quasi-experimental designs. *Best practices in quantitative methods*, pages 155–176.
- Trechsel, A. H. and Vassil, K. (2011). Internet Voting in Estonia. A Comparative Analysis of Five Elections since 2005. Report to the Estonian National Electoral Committee. Available at: <http://www.vvk.ee/valijale/e-haaletamine/raportid/>.
- Vassil, K. and Weber, T. (2011). A bottleneck model of e-voting. why technology fails to boost turnout. *New media & society*, Forthcoming.