Who is mobilized to vote by short text messages? Evidence from a nationwide field experiment with young voters

Salomo Hirvonen, Maarit Lassander, Lauri Sääksvuori, Janne Tukiainen*

Abstract

We conduct a large-scale randomized controlled trial to evaluate the effectiveness of short text messages (SMS) as a tool to mobilize young voters, and thus, ameliorate the stubborn gap in political participation between younger and older citizens. We find that receiving an SMS reminder before the Finnish county elections in 2022 increases the probability of voting among 18-29-year-old voters by 0.9 percentage points. Additionally, we observe that the most simplified message is more effective than messages appealing to expressive or instrumental motivations to vote. Using comprehensive administrative data, we examine treatment effect heterogeneity and spillover effects. We document that SMS based mobilization of voters does not only reduce existing social inequalities in voting between the age cohorts but also among the young citizens. Moreover, we remarkably find that over 100 percent of the direct treatment effect spilled over to non-treated household members. Our results indicate that SMS can be a fairly low-cost tool for reducing gaps in political participation. Moreover, our results exemplify the importance of understanding spillover effects and treatment effect heterogeneities in the evaluation of voter mobilization interventions. In 2023 we conducted similar RCT (results available shortly) during the parliamentary elections in order to study persistence and dynamic effects from the 2022 experiment.

JEL: C93, D72

Keywords: Get-out-the-vote, Field experiments, Spillover effects, Voter turnout

*Hirvonen: Department of Economics, University of Turku, Rehtorinpellonkatu 3, FI-20014 Turku, Finland. Lassander: Prime Minister's Office, Snellmaninkatu 1, FI-00023 Helsinki, Finland. Sääksvuori: Finnish Institute for Health and Welfare. Mannerheimintie 166, FI-00271, Helsinki, Finland. Tukiainen (corresponding author): Department of Economics, University of Turku, Rehtorinpellonkatu 3, FI-20014 Turku, Finland, janne.tukiainen@utu.fi, +358503083620. We thank the seminar participants at FEAAM, the University of Helsinki, Laurea, and SABE/IAREP for helpful comments. We are grateful to the Ministry of Justice (Finland) and the Prime Minister's Office (Finland) for funding and help. This RCT was registered in the American Economic Association Registry for randomized control trials (AEARCTR-0008790). The Ethics Committee for Human Sciences at the University of Turku approved this study (48/2021).

1 Introduction

Political participation is a central feature of democratic governance and voter turnout a key indicator of how citizens participate in the governance. Despite increasingly educated electorates and reduced institutional barriers to vote, voter turnout has been declining across the globe since the early 1990s (Solijonov, 2016; Hooghe and Kern, 2017). Consequently, low voter turnout is often identified as a major challenge for the sustainability and legitimacy of public policies.

Low voter turnout is systematically associated with unequal turnout (Lijphart, 1997). As a result, uneven participation in voting typically leads to unequal descriptive and substantive representation that is biased against underprivileged citizens (Fowler, 2013; Harjunen et al., 2023).[1] Across the electorates, one of the largest demographic gaps in voting is by age (Mo et al., 2022). The low turnout rates of young voters can be expected to have large effects on election outcomes and steer public policy towards the preferences of older citizens (McClean, 2021). The sources of low young voter turnout and other demographic gaps in voting are heavily studied and debated in the literature (Holbein and Hillygus, 2020). However, there still exists limited knowledge on how to effectively address low young voter turnout and demographic gaps in political participation in practice.

This paper evaluates the promises and pitfalls of short text message (SMS) reminders as a tool to mobilize young voters and ameliorate the stubborn gap in political participation between younger and older citizens. First, we conduct a large randomized controlled trial to evaluate the effectiveness of three different types of non-partisan SMS reminders on voter turnout in nationwide county elections in Finland. The target population of our experiment are young adults aged between 18 and 29 years, a population group with high levels of human capital but persistently low turnout rates.[2] Second, we merge electronic voter turnout records with rich individual-level administrative data on eligible voters to investigate the average treatment effect of this large non-partisan text message-based get-out-the-vote (GOTV) campaign on social inequalities in voting. Third, using demographic data and past voting records, we examine the potential heterogeneity of treatment effects among eligible voters using pre-registered heterogeneity tests and data-driven machine learning methods. Finally, using unique household IDs, we investigate how turnout decisions transmit between household members.

[1] Alternatively, high turnout may also lead to adverse political and policy outcomes if it implies more uninformed vote choices (Hodler et al., 2015; Lo Prete and Revelli, 2020).

[2] A systematic assessment of expected human capital formation for children born in 195 different countries ranks Finland as the country with the highest level of expected human capital in the world (Lim et al., 2018). Despite the high levels of human capital among young adults, Finland has one of the largest age gaps in turnout between older (aged 60 and above) and younger (aged from 18 to 29 years) voters (Mo et al., 2022).

Our paper contributes to the get-out-the-vote literature by testing the effectiveness of different message contents, tailored to appeal to different motivations to vote, and utilizing new data-driven approaches to identify voters who are the most susceptible to be mobilized using text message reminders. By reporting new evidence on the effectiveness of SMS reminders on young voter turnout and conducting heterogeneity analyses that utilize exceptionally rich individual-level information on eligible voters, we provide new evidence to assess the efforts to increase young voter turnout and identify the characteristics of voters who are the most and least responsive to non-partisan political campaigning.

Our results are largely consistent with the existing literature that has systemically, although with varying magnitudes, documented the effectiveness of SMS reminders on voter mobilization (Bhatti et al., 2017a,b; Bergh et al., 2021; Bergh and Christensen, 2022; Naess, 2022). We find a statistically significant, about 0.9 percentage point, direct average treatment effect in the probability of voting. The effect size of 0.9 percentage points equals 3% increase compared to the control group average turnout of 30.9%. The effect is larger for a neutral than expressive or instrumental messages. Our study provides a nuanced picture about the prospects of using SMS reminders as a mobilization tool to increase turnout and narrow the enduring gap in political participation between younger and older citizens. Importantly, we document heterogeneous treatment effects showing that SMS-based mobilization strategies are more likely to diminish than exacerbate existing social inequalities in voting also within the young voters. Moreover, we observe that over 100 percent of the direct effect spilled over to other household members - above all to older household members of young voters. Overall, our results suggest that SMS reminders are effective at mobilizing young low-propensity voters and their household members who are typically less well represented among the voters. More generally, our results demonstrate that RCTs with a limited focus on the analysis of individuals in the treatment and control groups alone may substantially underestimate the net effect of get-out-the-vote interventions.

Our paper relates to several strands of literature. First, our study builds on and contributes to the literature that has investigated the effectiveness of numerous voter mobilization strategies and different mediums of communication as well as different contents of campaign messages on voter turnout and choice (Green et al., 2013; Green and Gerber, 2019). Our experimental design and the use of text message reminders resembles the original field experiments that established the potential usefulness of text messages as mobilization tools and led to the formulation of the Noticeable Reminder Theory (Dale and Strauss, 2009; Malhotra et al., 2011). To date, there is a modest but growing body of experimental literature that has extended the study of text messages as mobilization tools to different cultural, geographical, and electoral contexts. Most closely related to our study are the experiments conducted in Denmark

(Bhatti et al., 2017a,b) and Norway (Bergh et al., 2021; Bergh and Christensen, 2022; Naess, 2022) where the registration of all eligible voters is automatic and based on nationwide population registers, allowing experimental testing of impersonal voter mobilization strategies in large and representative population groups.

Second, our paper relates to the very few experimental studies on voter mobilization with an explicit objective to measure spillover effects. Prior to our work, Nickerson (2008), Sinclair et al. (2012) and Bhatti et al. (2017a) have investigated how the effects of different get-out-the-vote appeals may transmit from treated to untreated individuals and reported within household spillover effects varying from 30% to 60% of the direct effect. Our findings complement the existing literature on the measurement of spillover effects by further stressing that if spillovers are not carefully analyzed, the overall impact of the intervention is likely to be severely understated. At the same time, the exceptionally large spillover effects from younger to older voters suggest that there are large heterogeneities in transmission of voting decisions across different types of social relationships.

Finally, and more generally, our study relates to the literature on social inequalities in political participation. A recent literature on the compositional effects of get-out-the-vote mobilization strategies suggests that many current mobilization strategies may widen existing social disparities in voting by predominantly mobilizing high-propensity voters instead of under-represented low-propensity voters (Arceneaux and Nickerson, 2009; Enos et al., 2014). Our paper complements the existing literature on the compositional effects of GOTV mobilization strategies in three ways. First, we assess the compositional effects of GOTV mobilization in an electoral context where all eligible citizens are automatically registered to vote. Second, to date, there is very little evidence on the compositional effects of text message-based mobilization strategies. Third, we assess the robustness of the prevailing empirical strategy in the relevant literature that estimates baseline voting propensities using within-sample covariates and interacts the predicted propensities to vote with the GOTV treatment indicator. To address the concern that the within-sample estimates of voting propensities may not predict turnout out-of-sample, we estimate citizens' propensities to vote in the absence of treatment using a machine learning technique that separates the choice of covariates and fitting of the prediction model.

The paper proceeds as follows. Section 2 describes the relevant electoral context in Finland. In section 3, we describe the experimental design and sample. Section 4 presents our empirical methods. Section 5 presents our main results and findings from several auxiliary analyses aiming to explore potential heterogeneities and spillovers in treatment effects. Section 6 concludes.

2 Background and Context

We conducted our RCT in the context of Finnish nationwide county elections held on January 23, 2022. Counties are the mid-tier level of decision-making in Finland between the municipalities and the central government. They resulted from a recent large social and healthcare reform. Thus, the elections were the first of their kind in Finland. The elections were expected to be of low salience and interest. This expectation also turned out to be true as turnout in the elections was 47.5%, which is lower than in any parliamentary elections in the Finnish history.

The allocation of seats in county elections is proportional to the votes following d'Hondt system of open party list proportional representation and identical to the Finnish parliamentary and municipal elections. Finland uses a very pure form of open-lists in the sense that personal vote is obligatory: each voter gives exactly one vote to one candidate. Parties are assigned seats based on the sum of its candidates' personal votes and seats within the party are assigned purely based on the personal votes. Overall, the open list electoral system in Finland may increase incentives for individual campaigning compared to several democracies with closed list or mixed electoral systems.[3]

Voters are automatically registered in all elections in Finland. An electronic register of all eligible voters (voting register) is established based on the Population Information System on the 46th day before the election day (Jääskeläinen, 2020). All voters listed in the voting register receive a notice of their right to vote (polling card) no later than 24 days before the election day. The polling card indicates the date of the election, the period for advance voting, the locations of advance polling stations within the voter's electoral district, the address of the voter's election day polling station, and contact information of the electoral authorities. The polling stations have only an administrative role as the elections are held at-large in the whole county. A typical characteristic of the Finnish elections is that a relatively large share of voters cast their ballots at polling stations during the period for advance voting that begins 11 days before the election day and ends five days before the actual election day. In the 2022 county elections, 57% of individuals who voted used the advance period to cast their vote.

Prior to our study, text message-based mobilization experiments have been conducted in the US, Denmark and Norway. The Finnish electoral system and voter mobilization environment closely resembles the other Nordic countries. Turnout in Finnish local and regional elections is typically markedly higher than in the local US elections, but has been in many recent elections noticeably lower than in comparable

[3] In contrast, in the other Nordic countries, parties have a larger role in the electoral system. Sweden nominally uses a flexible list where it is possible to give personal votes. However, a large number of those are needed to change the otherwise closed list. In Norway, municipal elections use open list, but parties can give large personal vote bonuses to their preferred candidates.

Danish and Norwegian elections (Bhatti et al., 2017b; Bergh et al., 2021). In the 2021 municipal elections, the turnout of eligible voters was 55.1%. There are notable demographic inequalities in voting. Young adults aged from 18 to 29 years are markedly less likely to vote than the older age cohorts. Their turnout in the 2021 municipal elections was 36.6%. The gender gap among young voters in the 2021 municipal elections was 8 percentage points as women had a turnout of 40.7% and men had a turnout of 32.7%.

Voters' access to information on party platforms and individual candidates is supported through wide-ranging public information campaigns and strong public media presence. Political campaigning and advertising is regulated by the Election and Data Protection Acts that restrict the use of personalized advertising using direct mailings, phone calls and text messages. To our knowledge, prior to this study, there has not been politically motivated or government sponsored non-partisan text-message campaigns to mobilize voters in Finland.

3 Experimental Design and Data

3.1 Sample

To conduct the experiment, we accessed the electronic register of eligible voters maintained by the Finnish Digital and Population Data Services Agency. This electronic register contains information on voters (e.g., name, personal identity code, electoral district, and the municipality of residence) as recorded in the Population Information System. Importantly, the electronic voting register enables us to link different treatment arms to individual-level electronic records on turnout. Our sample includes municipalities where voting districts having an electronic voting register cover at least 80% of the eligible voters in the municipality. This leads to a sample with 99 municipalities with full electronic voting registry coverage and 19 municipalities with more than 80% coverage out of 309 municipalities. Table 1 shows that 56% of all eligible voters aged 29 years and under live in these municipalities.

After extracting relevant personal information of all eligible voters aged from 18 to 29 years and residing in the voting districts covered by the electronic voting register, we contracted with an IT-company that conducted a search to provide the cell phone numbers of individuals included in the electronic voting register. The matching of eligible voters' personal information to valid cell phone numbers led to an analysis sample of 51101 individuals aged from 18 to 29 years of age.[4]

[4]The company was able to find cell phone numbers for 18.2 percent of individuals included in the electronic voting register.

Table 1: Summary statistics: Sample compared to population

	Analysis sample Full Sample (1)	Analysis sample Aged 19 to 29 (2)	Analysis Municipalities Aged 19 to 29 (3)	Full population Aged 19 to 29 (4)
Female	0.40	0.41	0.48	0.49
	(0.49)	(0.49)	(0.50)	(0.50)
Age	24.62	24.65	24.19	24.28
	(3.15)	(3.13)	(3.15)	(3.12)
High School Degree	0.44	0.44	0.44	0.45
	(0.5)	(0.5)	(0.5)	(0.5)
Taxable Income	158781	15808	13539	13972
	(13163)	(13160)	(12399)	(12553)
Immigrant	0.04	0.04	0.07	0.07
	(0.20)	(0.20)	(0.26)	(0.25)
Observations	51,101	50,899	280,925	496,042

Notes: Standard deviations in parentheses. Covariates are measured in year 2019 with the exception of age which is for year 2022. Number of observations for taxable income are 47,503 (Column 1), 47,416 (Column 2), 258,065 (Column 3) and 458,604 (Column 4).

Table 1 shows descriptive statistics for various samples. Column (1) shows the analysis sample that was used to randomize individuals into treatment and control groups. Column (2) drops from this analysis sample the 18 year old eligible voters to facilitate more accurate comparison between Columns (2) - (4). Column (3) describes all 19-29-year-old individuals contained in the electronic voting register. Column (4) contains the full population belonging to same age cohorts. As we have information only on the year of birth, and not on the date of birth that would be necessary for identifying 18-year-old eligible voters from the full population, Columns from (2) to (4) do not include any 18-year-old individuals. By comparing Columns (3) and (4), we find that the demographics in municipalities used to draw our sample due to the availability of electronic voting register closely resemble the demographics of full equally aged population in Finland. By comparing Columns (2) and (3), we find that the final analysis sample closely reminds the same aged population living in the same municipality with a somewhat lower share of females and immigrants. By contrast, the taxable income is somewhat higher in our analysis sample than in the same aged population at large. Overall, the comparison of our analysis sample to the full population sample suggests that the restriction to municipalities with an electronic voting register and loss of individuals because of not observing their phone numbers does not substantially affect the representativeness of our results.

3.2 Experimental design

To estimate the direct causal effect and potential spillover effects of alternative text message reminders on voter mobilization, we randomized all individuals in our analysis sample into control and treatment groups. There were three different treatment groups that varied the wording of text messages. We used an allocation ratio that assigned 40 percent of individuals into a control group and 60 percent of individuals into three equally sized treatment groups (Figure 1). We stratified the randomization by municipality to guarantee that 60% of all eligible 18 to 29-year-old voters received a reminder in each municipality. The stratification by municipality is expected to increase the precision of estimated treatment effects (Duflo et al., 2007) and enables us to provide more reliable estimates for local level analyses. At the time of randomization, we did not possess data on other covariates suitable for stratified randomization. The objectives of our RCT and a study protocol was pre-registered in the American Economic Association Registry for randomized control trials as AEARCTR-0008790. The Ethics Committee for Human Sciences at the University of Turku, Finland, approved this study (decision number: 48/2021).

Following the timing of polling opportunities in the Finnish elections, we sent two text messages for all individuals in treatment groups. The first message was sent a day before the beginning of the advance voting period. The second message was sent a day before the election day. There was no variation in the intraday timing of text messages. All messages were simultaneously sent at 4 pm using a mass text messaging service.

We measure the effect of SMS reminders on voter turnout using individual-level data recorded in an electronic register of turnout. The electronic voting record contains a unique identifier for each citizen and a variable indicating whether the person voted in the election. Using unique personal identifiers and household IDs, we merge the voting records with the treatment assignment, comprehensive socio-economic data and pre-existing turnout data that covers citizen's participation in all nationwide elections since 2015. Crucial to the treatment heterogeneity analyses, we are able to merge the voting records with individual-level data on prior voting histories and rich personal information including, among other information, data on voter's labor income, capital income, social transfers, education, ethnicity and employment records. The resulting data are proprietary and accessible only through Statistics Finland's remote access system, where all analysis is conducted. Thus we are not able to share the data, but our results can be replicated with the code provided by us and purchasing the mentioned data sets together with acquiring access to Statistics Finland's remote access system.

3.3 Message contents

Since the popularization of the nudge theory (Thaler and Sunstein, 2009), there has been a large influx of studies testing the effectiveness of varying message contents for multiple purposes in numerous different contexts. While there are some broadly heralded examples of cases in which small variations in message

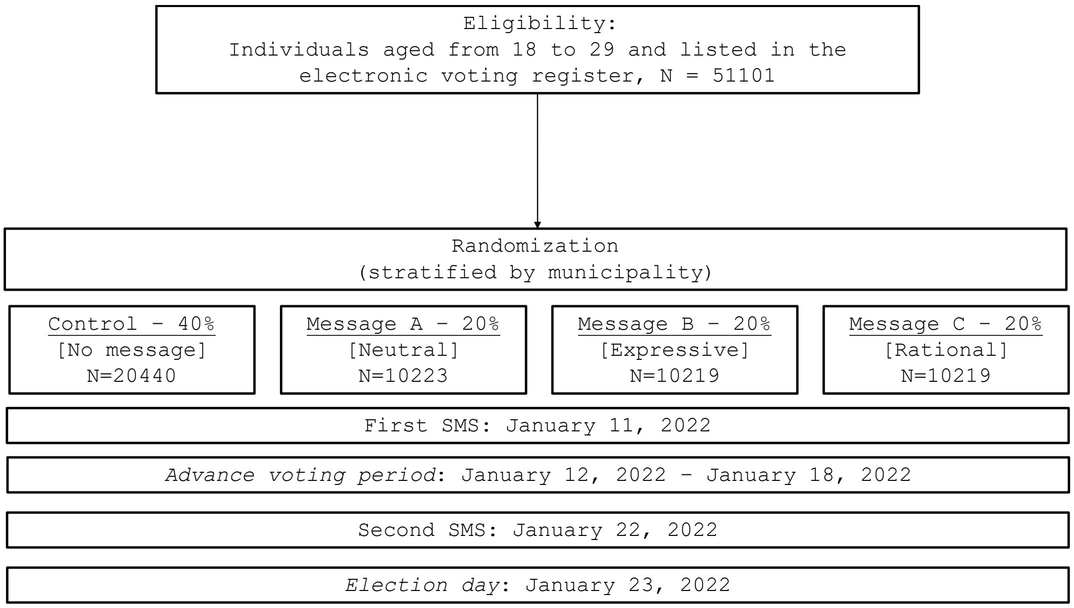


Figure 1: Eligibility, Randomization and Treatment.

contents have led to meaningful differences in behavioral outcomes, in the context of voter mobilization the noticeable reminder theory (Dale and Strauss, 2009) implies that the content of text messages should not affect turnout. However, there is still little empirical research testing how text message reminders with different types of appeals impact the likelihood of being mobilized to vote. In addition to examining the overall causal effect of text message reminders on voting, we tested the effectiveness of different message contents. For this purpose, we developed three different types of messages that appeal to different motivations to vote. The first type of message was a neutral message that just briefly informed recipients about the forthcoming elections and abstained from expressive and instrumental motivations to vote. The second type of messages was developed to appeal to the expressive motive of voting (Brennan and Hamlin, 1998; Brennan and Brooks, 2013) and highlighted voters' right to express their voice by voting. The third type of messages was developed to appeal to a more instrumental or rational motive of voting (Downs, 1957; Lyytikäinen and Tukiainen, 2019) and emphasized recipients' chance to influence the direction of policies and provision of public services through voting. The exact wording of different types of messages is available in Table 2.

All message contents were developed by the authors in collaboration with the electoral authority (Ministry of Justice, Finland) to ensure that the contents conformed with the existing electoral code of conduct. All messages included a hyperlink to a homepage www.vaalit.fi [www.elections.fi] maintained by

Table 2: Overview of message contents by treatment

Treatment	Message #	Message text
Neutral	#1	"Hi, please remember that county elections will be held on January 23. Domestic advance voting period is from January 12 until January 18. More information at vaalit.fi. Regards, Ministry of Justice."
Neutral	#2	"Hi, please remember that county elections will be held on January 23. More information at vaalit.fi. Regards, Ministry of Justice."
Expressive	#1	"Hi, please remember to use your right to vote in country elections on January 23. Domestic advance voting period is from January 12 until January 18. More information at vaalit.fi. Regards, Ministry of Justice."
Expressive	#2	"Hi, please remember that county elections will be held on January 23. Democracy needs your voice, please use your right to vote. More information at vaalit.fi. Regards, Ministry of Justice."
Instrumental	#1	"Hi, have your say on community services in county elections on January 23. Domestic advance voting period is from January 12 until January 18. More information at vaalit.fi. Regards, Ministry of Justice."
Instrumental	#2	"Hi, please remember county elections on January 23. By voting, you can have a say on the organization of health and social care services, and fire and rescue care. More information at vaalit.fi. Regards, Ministry of Justice."
Control	-	[None]

the electoral authority to provide reliable and unbiased information about the organization of elections in Finland. The electoral authority served as the sender of the messages which is likely to have increased the credibility of messages and set the notifications apart from standard promotional messages that individuals may receive to their phones.

4 Estimation methods

Following the randomization procedure, access to administrative data containing unique personal and household IDs, and our focus on understanding potential spillover effects and treatment effect heterogeneities, we provide results from four different types of empirical analyses that we now describe.

4.1 Direct effects

To assess the direct impact of SMS reminders on turnout at large, we estimate the pooled average treatment effect of receiving any type of reminder in contrast to the counterfactual of receiving no reminder. Moreover, to investigate the direct effect of different contents of reminders on turnout, we estimate the average treatment effects by treatment. As pre-registered, we estimate the direct treatment effects using a linear probability model and progressively add control variables to the model:

$$Y_i = \beta_0 + \beta_1 Treatment_i + \boldsymbol{X}_i'\boldsymbol{\beta} + \epsilon_i,$$

where $Treatment_i$ indicates treatment assignment and $X_i'\boldsymbol{\beta}$ individual level demographic controls.[5] Our demographic controls include educational background, which is defined as the mother of the individual having a high school degree or using individual's own high school degree status if our data does not allow us to identify the mother of the individual (29% of our sample) based on the household data going back to year 2011.[6] In addition to the educational background, we use logarithm of individual's mother's taxable income and occupation as controls for the socio-economic background. As our sample consists of young voters, we believe that mothers' characteristics are more accurate in describing individuals circumstances and predicting voting than their own characteristics. In addition to educational and socio-economic background, we include individuals' ethnicity, which takes value 1 if person's both parents are born outside of Finland. We also include age, gender and an indicator variable documenting if the individual was eligible to vote in the 2022 elections for the first time. Adding control variables to the estimations of average treatment effects in a randomized experiment is not expected to affect the point estimates, but can reduce residual variance and increase the precision of estimates. We cluster standard errors at the municipal level.[7]

4.2 Spillover effects

Unique household IDs included in our data enable us to investigate the spillover effects of our get-out-the-vote intervention within the households.[8] To study the intra-household transmission of treatment effects after receiving an SMS reminder, we restrict our sample to households where there was either exactly one young voter who was part of the treatment group or there was exactly one young voter who was part of the control group, leading to a sample size of 51.4% of the total sample as a high proportion of individuals in our sample are living alone. Therefore, the treatment group for spillover effects includes all individuals living within the same household in the end of year 2020 (as this is the most recent data

[5]Following our pre-analysis plan, we conduct supplementary analyses using Logit models to study the robustness of our linear probability model estimates. Results from these estimations are reported in the Online Appendix (Tables A3 and A4) and show that our results are robust to the choice of the estimation method.

[6]Online Appendix (Table A5) shows sample means of covariates by whether the mother is identified. Individuals whose mother is not identified are more likely to be older, have higher income and have foreign background. Result do not qualitatively change if we use only individual's own covariates.

[7]From a design-based perspective, clustering may not be necessary as our treatment is assigned at the individual level (Abadie et al., 2022). However, in order to generalize our results to the whole population of young voters clustering accounts for municipality-level sampling variance as we observe only a subset of Finnish municipalities.

[8]As the number of treated individuals living together with control group individuals is small (5% of the control group individuals) even very large spillovers of over 100% would not affect our direct effect estimates at any meaningful decimal level. Thus, we do not examine potential spillovers from treatments groups to control groups, but focus on the intra-household transmission of treatment effects from our target sample (voters aged 18 to 29 years) to other eligible voters.

point available to us) with an individual who received an SMS reminder and the control group consists of all individuals who were cohabiting with a young voter who was part of the control group. On average there are 1.52 voting aged individuals in addition to the SMS receiver (or control group member) in these households. We estimate the same set of models for the spillover sample as we do for the direct effects sample.

4.3 Inequality analysis using propensities to vote

The estimation of direct and spillover effects enables us to assess the effects of SMS reminders on turnout at large. However, these effects may not be evenly distributed in the electorate and may either exacerbate or ameliorate existing disparities in political participation. Building upon the work by Arceneaux and Nickerson (2009) and Enos et al. (2014), we analyze the effect of text message-based mobilization on the composition of the electorate. Our estimation procedure involves the following steps. First, we predict a propensity to vote for every individual using the available administrative data and the following logistic regression model:

$$Pr(Y_i = 1|\boldsymbol{X}_i) = \frac{exp(\boldsymbol{X}_b)}{1 + exp(\boldsymbol{X}_b)}$$

where $Pr(Y_i = 1|\boldsymbol{X}_i)$ is the predicted probability of voting based on individuals' gender, age, logarithm of (mother's) taxable income, ethnicity, education, SES background, eligibility to vote for the first time and municipality fixed effects. It is noteworthy that we are able to estimate these individual propensities to vote using a much richer set of personal information than what has been been available in previous studies.

To estimate individual voting propensities in the absence of treatment, we conduct the propensity score estimation in a sample that is restricted to individuals assigned to the control group. The random assignment of individuals into the treatment and control groups guarantees that the propensity estimates in the control group are equally representative of the treatment group. Consequently, we compute for every individual in the sample their predicted probability to vote in the Finnish 2022 county elections in the absence of the SMS mobilization campaign. Second, we group the voting propensities by 25th, 25-75th, and top 25th percentiles.[9] This grouping is done to detect possible non-linear effects by voting propensity (Arceneaux and Nickerson, 2009; Fowler, 2015). Splitting the sample into three groups is a

[9]This grouping splits the sample into half between the marginal group, where we would theoretically expect the largest effect, and the others. To study the robustness of our results based on this grouping, we use an alternative grouping that splits the sample into three equally sized group. The results using this alternative grouping are reported in the Online Appendix (Tables A1 and A2).

more flexible approach compared to imposing a functional form for voting propensity by adding it into an OLS specification, while it retains statistical power for doing group comparisons compared to finer groupings. Finally, we estimate the effect of receiving an SMS reminder in these groups using the linear probability model to test whether the treatment systematically interacts with the existing disparities between high-propensity voters, marginal voters, and under-represented low-propensity voters.

We note that the estimation of voting propensities through logistic regression may pose a risk of overfitting the data by fitting random variation and using outlier observations in demographic variables that could lead to biased comparison of treatment heterogeneities between high-propensity voters and under-represented voters. To address this concern, we complement the initial analysis by estimating voting probabilities through the Electic Net (Zou and Hastie, 2005; Hastie et al., 2015). In contrast to the propensity score estimation using logistic regression models, the Elastic Net chooses an optimal combination of predictors using two penalty terms: one from LASSO (based on absolute value of the estimated coefficient, enabling elimination of predictors) and one from ridge regression methods (based on the square of the estimated coefficient, not enabling elimination of predictors). Thus, the Elastic Net overcomes, first, the tendency of LASSO to select only one predictor among highly correlated covariates. Second, the method allows dropping out predictors, which is not done by ridge regression alone. The procedure employs sample folding to separate the choice of parameters for penalty terms and fitting the model. Taken together, the Elastic Net trades bias for less variance by using penalty terms, reducing the risk of over-fitting the data.[10]

4.4 Heterogeneity analysis using honest causal forest

Finally, we employ a more data-driven machine learning approach for the estimation of potential heterogeneous treatment effects. The honest causal forest approach by Wager and Athey (2018) explores the heterogeneity of treatment effect using a multi-step procedure to avoid over-fitting the data. The honest causal forest method partitions sample according to splits by covariates into leafs and estimates conditional average treatment effects in each of these leafs. This splitting procedure is repeated many times to find which splittings lead to consistently larger differences in the treatment effects giving conditional treatment effect prediction for each observation. Observations are ranked by their conditional treatment effect prediction and quantiles are formed to compare covariate means across the predicted treatment

[10] We report in the Online Appendix (Figure A1) the Receiving Operating Characteristic (ROC) curves for in-sample and out-of-sample predictions using the logit and elastic net models. We find that the Logit model is slightly better in terms of Area Under Curve (AUC) for in-sample prediction, whereas the elastic net model has higher AUC for the out-of-sample prediction. Online Appendix (Tables A6 and A7) shows the covariates by voting propensity groups for the logit and the elastic net models. The latter has a steeper gradient in terms of gender and educational background.

effects. The honest causal forest algorithm separates the splitting and estimation of the conditional average treatment effect by using part of the sample for the former task and another part for the latter. The advantage of this method is that we do not need to assume at which dimensions the treatment effect heterogeneity takes place, which could be difficult to do based on theory ex-ante. The drawback is that a smaller sample is used for the actual estimation leading to noisy estimates. For this reason and given our limited sample size, we do not consider this analysis to be as revealing or important as the two approaches above.

5 Results

5.1 Direct effects

We begin by estimating the effect of SMS reminders on turnout at large and report the direct Average Treatment Effect (ATE) in Table 3. We observe that receiving an SMS reminder leads to a 0.9 percentage point (p.p.) increase in turnout. This effect is statistically significant at the conventional 1% significance level. As expected, the ATE estimate remains stable around 0.9 p.p. after progressively adding demographic control variables. To put the effect size into perspective, we note that the turnout in the control group is 30.8 percent. Thus, the effect size of 0.9 p.p. equals around 3% increase compared to the turnout in the control group. Moreover, we observe that receiving an SMS reminder bridges the gap between the 18-29 year-old voters and all other voters with an average turnout of 47.0% by 5.6%. Analogously, an SMS reminder bridges the gap between the 18-29-year-old voters and 30-39-year-old voters with an average turnout of 36.6% by 16%. Overall, the positive direct effect is consistent with the findings from previous studies that have examined the effectiveness of text message reminders in the Nordic counties.

In the following, we estimate direct treatment effects across the different treatment arms. Table 4 shows point estimates by treatment using the same set of control variables as in Column (3) in Table 3. We find that the estimated treatment effect for the Neutral treatment is 1.6 p.p. and statistically significant at 1% significance level. This effect size is almost twice as large as the effect size for the Expressive treatment (0.9 p.p). However, the difference between the two estimates is not statistically significant at conventional significance levels. Moreover, we find that the point estimate for the Neutral treatment is eight times larger than the point estimate for the Instrumental treatment (0.2 p.p.). This difference between these two coefficients is statistically significantly different at 5% significance level. Overall, these observations suggest that the most simplified message not appealing to any particular motivation to vote may have been the most effective at getting the young voters to turn out their vote.

Table 3: Average Treatment Effect

	Voted			
	(1)	(2)	(3)	(4)
Treatment (pooled)	0.009***	0.009**	0.009***	0.009***
	(0.003)	(0.003)	(0.003)	(0.003)
Controls				
Gender	✗	✓	✓	✓
Age	✗	✓	✓	✓
Ethnicity	✗	✓	✓	✓
Ln income	✗	✓	✓	✓
SES	✗	✗	✓	✓
Education	✗	✗	✓	✓
First-time voter	✗	✗	✓	✓
Municipality FE	✗	✗	✗	✓
Control group $\bar{Y}$	0.307	0.308	0.308	0.308
Observations	50.140	49.679	49.679	49.679

Notes: *** p < 0.01, ** p < 0.05 ,* p < 0.1, standard errors clustered at the municipal level in parentheses.

Table 4: Different Treatments

	Voted			
	Pooled	Neutral	Expressive	Instrumental
	(1)	(2)	(3)	(4)
Treatment	0.009***	0.016***	0.009*	0.002
	(0.003)	(0.005)	(0.005)	(0.004)
Controls	✓	✓	✓	✓
Control group $\bar{Y}$	0.308	0.308	0.308	0.308
Observations	49.679	29.799	29.806	29.832
Differences		Neutral - Expressive	Expressive - Instrumental	Instrumental - Neutral
		0.007	0.007	-0.015**
		(0.007)	(0.006)	(0.007)

Notes: *** p < 0.01, ** p < 0.05 ,* p < 0.1, standard errors clustered at the municipal level in parentheses. Controls include gender, age, ethnicity, ln taxable income, SES background groups, education (high school completion) and an indicator variable whether individual was eligible to vote for the first time.

5.2 Spillover effects

In this section, we conduct similar analyses as in the previous section but apply the estimation procedure to measure within household spillover effects on non-treated individuals. Table 5 shows that the ATE for the intra-household spillovers is around 1.4 p.p., suggesting that over 100 percent of the direct treatment effect spilled over to non-treated household members. The effect size of 1.4 p.p. equals around 2.8% increase compared to the baseline turnout of 49.4% in the control group. The observed spillover effect leads to two important implications. First, in the presence of sizable spillover effects, impact evaluation analyses not able to detect spillovers among social ties may lead to a substantial underestimation of the net causal effect. Second, spillovers from the target populations (e.g., young voters) to other population groups (e.g., older voters) could mean that the gap in turnout between the targeted population group and the other population groups does not shrink as much as suggested by simplistic comparisons based on estimated direct treatment effects. At the same time, interventions with large spillovers may influence social inequalities within the spillover group.

Table 5: Spillovers - Average Treatment Effect

	Voted			
	(1)	(2)	(3)	(4)
Treated in HH	0.014***	0.014***	0.013***	0.011***
	(0.006)	(0.005)	(0.006)	(0.005)
Controls				
Gender	✗	✓	✓	✓
Age	✗	✓	✓	✓
Ethnicity	✗	✓	✓	✓
Ln income	✗	✓	✓	✓
SES	✗	✗	✓	✓
Education	✗	✗	✓	✓
First-time voter	✗	✗	✓	✓
Municipality FE	✗	✗	✗	✓
Control group $\bar{Y}$	0.494	0.496	0.496	0.496
Observations	37.207	36.876	36.876	36.876

Notes: *** $p < 0.01$, ** $p < 0.05$,* $p < 0.1$, standard errors clustered at the municipal level in parentheses.

Table 6 presents estimated spillover effects by treatment type. As for the direct treatment effects, the estimated spillover effect for the Neutral treatment is higher than for the two other treatments. The estimated effect size of 2.1 p.p. is statistically significant at 1% significance level. Spillover estimates for the Expressive and Instrumental treatments are 0.6 p.p. and 1.2 p.p., respectively. However, unlike in the case of direct effects, we do not detect statistically significant differences in spillover effects between the different treatments.

Table 6: Spillovers - Different Treatments

	Voted			
	Pooled	Neutral	Expressive	Instrumental
	(1)	(2)	(3)	(4)
Treated in HH	0.013**	0.021***	0.006	0.012
	(0.006)	(0.008)	(0.008)	(0.008)
Controls	✓	✓	✓	✓
Control group $\bar{Y}$	0.496	0.496	0.496	0.496
Observations	36.876	22.028	22.113	22.059
Differences		Neutral - Expressive	Expressive - Instrumental	Instrumental Neutral
		0.016	-0.006	-0.010
		(0.011)	(0.011)	(0.011)

Notes: *** p < 0.01, ** p < 0.05 ,* p < 0.1, standard errors clustered at the municipal level in parentheses. Controls include gender, age, ethnicity, ln taxable income, SES background groups, education (high school completion) and an indicator variable whether individual was eligible to vote for the first time.

5.3 Heterogeneous effects by voting propensities

To assess the impact of our intervention on turnout inequality, we estimate in this section heterogeneous treatments effects by voting propensity. Table 7 (Panel A) presents direct treatment effects for voters divided into three voting propensity groups - Low Propensity Voters, Marginal Voters and High Propensity Voters - based on a logit model estimating predicted individual voting probabilities.[11] Table 7 (Panel B) reiterates the same analysis for within household spillover estimates. We find that the direct effect estimate for the low propensity voters is 2.0 p.p.. The direct effect for the marginal voters is 1.2 p.p.. The former coefficient is statistically significant at 1% level, while the latter coefficient is statistically significant at 5% level. The point estimate for the high propensity voters is -0.8 p.p., albeit not statistically significantly different from zero. The estimates of the first two voting propensity groups are significantly different from the high propensity voter's estimate at 1% significance level for the low propensity group and at 5% significance level for the marginal propensity group. Given that the baseline turnout rate for the low propensity voters is only around a half of that of the marginal voters and less than third compared to the high propensity voters, the relative effect size for the low propensity voters is remarkably larger than for the two other groups. Overall, our intervention seems to have reduced existing social inequalities in voting among the young voters - or at least it did not exacerbate existing inequalities in voting.

[11] Online Appendix (Table A3 and Table A4) shows results by three equal percentile splits. The results from these estimations are not qualitatively different.

Table 7: Heterogeneity by Vote Propensity

	Voted			
	All	Low Propensity {Bottom 25%}	Marginal Voters {25-75%}	High Propensity {Top 25%}
	(1)	(2)	(3)	(4)
Panel A: Direct Effects				
Treated	0.009***	0.020***	0.012**	-0.008
	(0.003)	(0.007)	(0.005)	(0.008)
Controls	✓	✓	✓	✓
Control group $\bar{Y}$	0.309	0.151	0.299	0.485
Observations	49.458	12.363	24.727	12.368
Differences		Marginal - Low	Marginal - High	High - Low
		-0.008	0.020**	-0.028***
		(0.008)	(0.009)	(0.010)
Panel B: Spillover Effects				
Treated in HH	0.013**	0.021**	0.016**	-0.006
	(0.006)	(0.010)	(0.008)	(0.011)
Controls	✓	✓	✓	✓
Control group $\bar{Y}$	0.497	0.242	0.495	0.761
Observations	36.723	9.180	18.362	9.181
Differences		Marginal - Low	Marginal - High	High - Low
		-0.005	0.022*	-0.027*
		(0.013)	(0.013)	(0.015)

Notes: *** $p < 0.01$, ** $p < 0.05$,* $p < 0.1$, standard errors clustered at the municipal level in parentheses. Controls include gender, age, ethnicity, ln taxable income, SES background groups, education (high school completion) and an indicator variable whether individual was eligible to vote for the first time.

Table 7 (Panel B) shows heterogeneous treatment effects by voting propensities for the spillover sample, in which each individual living in a same household together with a treated or non-treated young voter has an estimated individual voting propensity and is assigned into the three groups based on that prediction. We observe that the low propensity group has a point estimate of 2.1 p.p. and the marginal voters group has a point estimate of 1.6 p.p. - both coefficients are statistically different from zero at 5% significance level. The estimate for high propensity voters is -0.6 p.p. and not statistically significant. In the same manner as the estimated direct effects, spillover effects seem to reduce turnout inequality among the non-treated individuals.

To alleviate the concern of over-fitting the data while estimating the predicted probabilities to vote, we reproduce the analysis reported in Table 7 using predictions estimated by Elastic Net (Zou and Hastie, 2005; Hastie et al., 2015). Table 8 shows results using this alternative estimation procedure. We observe in Table 8 (Panel A) that the group of Marginal Voters now has the highest point estimate of 1.5 p.p., which is statistically significantly different from zero at 1% significance level. This group of marginal voters is followed by the low propensity voters with an estimate of 0.6 p.p. and the high propensity voters with an estimate of -0.2 p.p. T-test for difference between marginal propensity voters and high propensity voters is statistically significant at 5% significance level. In the spillover sample (Panel B), marginal voters have the highest point estimate of 1.5 p.p., which is statistically different from zero at 10% significance level. Estimated coefficients for the low propensity and the high propensity groups are 1.3 p.p. and 0.9 p.p., respectively. T-tests for differences between these three groups do not yield any statistically significant p-values. We interpret these results as evidence against the conjecture that an SMS based mobilization strategy would have widened disparities in participation by mainly mobilizing high-propensity individuals rather than under-represented population groups. These results complement existing heterogeneity results that have reported the largest treatments effects among high propensity voters and highlight the importance of studying the generalizability and robustness of estimated heterogeneous effects.

5.4 Heterogeneous effects by honest causal forest

In this section, we employ a machine learning method, honest causal forest (Wager and Athey, 2018), to further assess the potential heterogeneity of treatment effects and their consequences for social inequalities in voting. As detailed in section 4.4, the advantage of using honest causal forest algorithm is that we do not need to ex-ante impose the dimensions of potential treatment effect heterogeneities, but can let the machine learning method flexibly estimate the conditional treatment effects. Consequently, it is possible to assess which unique covariates are correlated with low and high conditional treatment effect estimates.

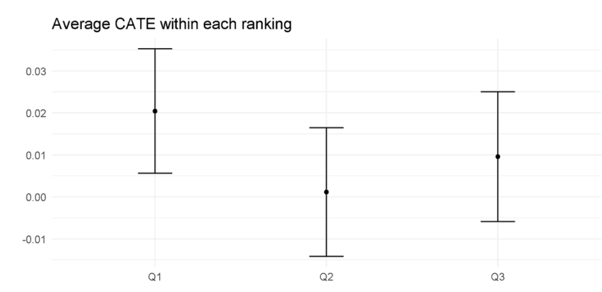
Table 8: Heterogeneity by Vote Propensity - Elastic Net

	Voted			
	All	Low Propensity {Bottom 25%}	Marginal Voters {25-75%}	High Propensity {Top 25%}
	(1)	(2)	(3)	(4)
Panel A: Direct Effects				
Treated	0.009***	0.006	0.015***	-0.002
	(0.003)	(0.006)	(0.005)	(0.007)
Controls	✓	✓	✓	✓
Control group $\bar{Y}$	0.308	0.161	0.294	0.481
Observations	49.679	12.361	24.806	12.512
Differences		Marginal - Low	Marginal - High	High - Low
		0.009	0.017**	-0.008***
		(0.008)	(0.008)	(0.009)
Panel B: Spillover Effects				
Treated in HH	0.013**	0.013**	0.015*	0.009
	(0.006)	(0.011)	(0.009)	(0.012)
Controls	✓	✓	✓	✓
Control group $\bar{Y}$	0.496	0.247	0.491	0.753
Observations	36.876	9.219	18.438	9.219
Differences		Marginal - Low	Marginal - High	High - Low
		0.002	0.005	-0.004
		(0.014)	(0.015)	(0.016)

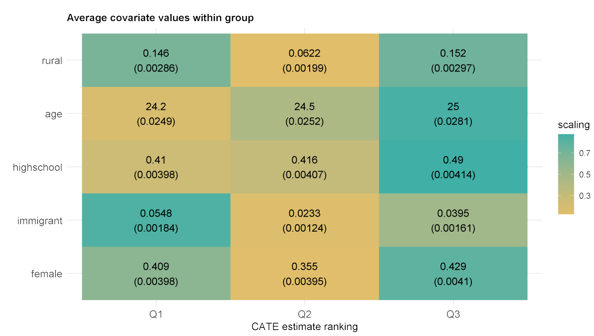
Notes: *** $p < 0.01$, ** $p < 0.05$,* $p < 0.1$, standard errors clustered at the municipal level in parentheses. Controls include gender, age, ethnicity, ln taxable income, SES background groups, education (high school completion) and an indicator variable whether individual was eligible to vote for the first time.

Figure 2 (Panel A) shows that there is no evidence for treatment effect heterogeneity for direct treatment effect estimates. In fact, the first group, which is predicted to have the lowest conditional average treatment effect (CATE) using the data in the splitting sub-sample, has the highest point estimate in the estimating sub-sample and none of the estimates are statistically different from each other.

Figure 2 (Panel B) shows that the highest CATE ranking group (colors scaled as 0.5 being the mean) has the highest mean of educational background and the highest proportion of females. Individuals in this group are also on average slightly older compared to the individuals in the two other groups. However, observed demographic differences between the groups are not large, which is not surprising as we do not observe differences in CATE estimates between the groups. We interpret these observations as further evidence that the intervention did not have an exacerbating effect on turnout inequality among the young voters. However, the lack of heterogeneous effects in this analysis is likely a result of low statistical power that results from dividing the sample by folding, and thus, may not serve as good evidence for contradicting the previous more precise heterogeneity analyses.



(a) Conditional Average Treatment Effect Estimates

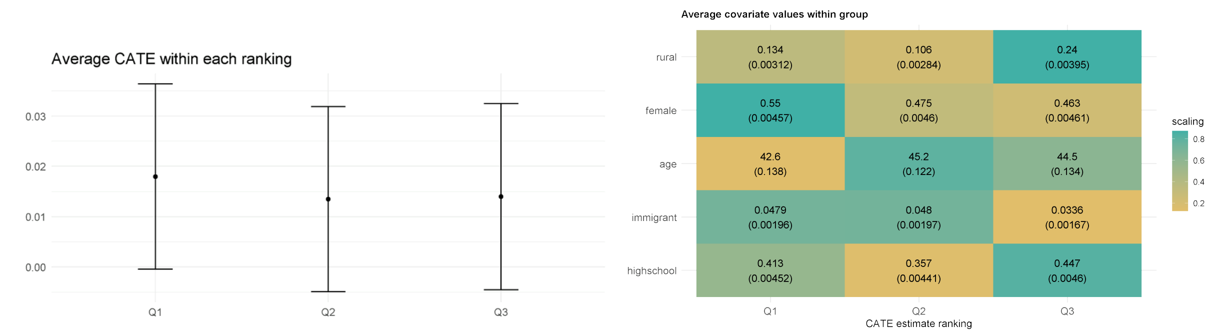


(b) Composition of CATE Ranking Groups

Notes: For Panel B colors scaled as 0.5 being the mean of the covariate in the CATE ranking group.

Figure 2: Honest Causal Random Forest Estimates - Direct Effects

Figure 3 (Panel A) shows that there are no statistically significant differences between the CATE ranking groups for spillover effect estimates. The highest CATE group has 24.0% individuals living in rural municipalities compared to 13.4% and 10.6% in the lowest and the middle groups, respectively. The proportion of women in the lowest CATE group is 55.0%, whereas there proportion of women in the middle group is 47.5% and the proportion of women in the highest group is 46.3%. For other covariates, the differences between group means are smaller. Overall, we find little evidence for large treatment effect heterogeneities in spillover effects to non-treated household members.



(a) Conditional Average Treatment Effect Estimates (b) Composition of CATE Ranking Groups

Notes: For Panel B colors scaled as 0.5 being the mean of the covariate in the CATE ranking group.

Figure 3: Honest Causal Forest Estimates - Spillover Effects

5.5 Heterogeneous effects by various subsamples

Finally, we examine treatment effect heterogeneity by splitting the sample according to single observed characteristic at a time, that is, educational background, ethnicity, voting in 2021 municipality elections and type of residential area (urban vs. rural).[12] By comparing Columns (1) and (2) in Table 9, we observe that the point estimates for the direct effects (Panel A) and for the spillover effects (Panel B) are higher for individuals whose mother has a high school degree than for individuals whose mother did not finish high school. However, these estimates are not statistically significantly different from each other. Turning into ethnicity, we observe that individuals born in Finland to Finnish parents have positive point estimates for the direct effects (Panel A) and for the spillover effects (Panel B), whereas immigrants have a negative direct effect estimate (-0.9 p.p.) and a negative spillover estimate (-1.7 p.p.). However, the sample size for individuals with an immigration status is small and the observed negative coefficients are not statistically different from zero. For spillover effects, the coefficient for the difference between native and non-native individuals is statistically significant at 10% significance level. Overall, these observations suggest that the intervention could have widened the turnout gap between the immigrants and the natives. Here, it is noteworthy that our reminders were sent in Finnish and Swedish, the two official languages in Finland, while all individuals aged 18 and above with a permanent residence in Finland are eligible to vote in the county elections. Thus, the eligibility to vote in the context of our study did not depend on the citizenship and associated language requirements, which may have contributed to the widening participation gap between the immigrants and the natives.

[12] The pre-analysis plan registered at the American Economic Association Registry for RCTs mentions age, geographical area, previous voting history, education and income as potential grouping variables for heterogeneous treatment effects. However, in the pre-analysis plan we did not present any specific hypotheses about the direction or magnitudes of potential effects.

Table 9: Heterogeneous Effects by Education and Ethnicity

	Voted			
	Educational background		Ethnicity	
	No High S.	High School	Native	Non-native
	(1)	(2)	(3)	(4)
Panel A: Direct Effects				
Treated	0.009	0.009*	0.010***	-0.009
	(0.006)	(0.005)	(0.003)	(0.013)
Controls	✓	✓	✓	✓
Control group $\bar{Y}$	0.230	0.405	0.316	0.120
Observations	27.659	22.020	47.696	1.983
Differences	-0.001		0.019	
	(0.007)		(0.014)	
Panel B: Spillover Effects				
Treated in HH	0.012	0.015**	0.014**	-0.017
	(0.008)	(0.007)	(0.006)	(0.018)
Controls	✓	✓	✓	✓
Control group $\bar{Y}$	0.410	0.620	0.510	0.175
Observations	21.811	15.065	35.324	1.552
Differences	-0.004		0.032*	
	(0.011)		(0.019)	

Notes: *** $p < 0.01$, ** $p < 0.05$,* $p < 0.1$, standard errors clustered at the municipal level in parentheses. Controls include gender, age, ethnicity, ln taxable income, SES background groups, education (high school completion) and an indicator variable whether individual was eligible to vote for the first time.

Table 10 presents results for heterogeneous treatment effects by voting in 2021 municipality elections and urbanity of the resident municipality. Panel A shows estimates for direct effects. Point estimate for individuals who voted in the 2021 elections is 2.8 p.p. and statistically significant at 1% significance level. The point estimate for those who did not vote in 2021 elections is 0.6 p.p. and not statistically different from zero. The difference of coefficients is statistically significant at 1% significance level. Panel B shows the results for spillover estimation. We find that those who voted in 2021 have a higher spillover effect with a point estimate of 2.0 p.p. compared to a coefficient of 0.8 p.p. for those who did not vote in 2021. This provides some evidence for the experiment having a widening effect on the participation gap. However, it should be noted that the effect sizes compared to untreated baseline are not too dissimilar from each other as the baseline for those who did not vote in 2021 is around five times smaller.

Table 10: Heterogeneous Effects by Voting in 2021 and Urbanity

	Voted			
	Voting in 2021		Urbanity	
	Voted	Not voted	Rural	Urban
	(1)	(2)	(3)	(4)
Panel A: Direct Effects				
Treated	0.028***	0.006	0.005	0.012***
	(0.007)	(0.004)	(0.014)	(0.004)
Controls	✓	✓	✓	✓
Control group $\bar{Y}$	0.593	0.114	0.313	0.306
Observations	17.643	27.800	5.335	38.791
Differences	0.022***		-0.007	
	(0.008)		(0.014)	
Panel B: Spillover Effects				
Treated in HH	0.020***	0.008	0.034*	0.012**
	(0.007)	(0.007)	(0.018)	(0.006)
Controls	✓	✓	✓	✓
Control group $\bar{Y}$	0.745	0.164	0.531	0.496
Observations	20.646	15.170	5.599	29.418
Differences	0.012		0.023	
	(0.010)		(0.019)	

Notes: *** $p < 0.01$, ** $p < 0.05$,* $p < 0.1$, standard errors clustered at the municipal level in parentheses. Controls include gender, age, ethnicity, ln taxable income, SES background groups, education (high school completion) and an indicator variable whether individual was eligible to vote for the first time.

Columns (3) and (4) in Table 10 split the sample into individuals living in rural and urban municipalities. In Panel A for the direct effects, the treatment estimate for residents in urban municipalities is higher (1.2 p.p., statistically significantly different from zero at 1% significance level) compared to young voters residing in rural municipalities (0.5 p.p., not statistically significantly different from zero). For the case of spillovers in Panel B, it is the other way around as individuals living in rural municipalities have a higher point estimate (3.4 p.p., statistically significantly different from zero at 10% significance level) compared to individuals living in urban municipalities (1.2 p.p., statistically significantly different from zero at 5% significance level). However, for both direct effect and spillover effect samples the differences between rural and urban municipality residents are not statistically significant.

Overall, this section demonstrates that interpreting effect heterogeneity with respect to whether SMS messages widen or diminish the participation gap among the youth may depend to some extent on whether the heterogeneity analysis is conducted using univariate or multivariate sample splits. In this case, the multivariate approach is arguably more rigorous, even if harder to interpret, and thus, we conclude based on those results.

6 Conclusions

This paper presents new evidence about the effects of short text message reminders on young voter mobilization. Using an RCT design and data-driven estimation techniques, we provide new insights to assess the utility of noticeable reminders to mobilize young voters and identify the characteristics of voters who are the most and least responsive to text message-based mobilization efforts. We obtain four findings. First, we find that receiving text message reminder before the Finnish country elections in 2022 increased the probability of voting by 0.9 percentage points in contrast to the counterfactual of receiving no reminder. Second, we find suggestive evidence that the most simplified phrasing of messages merely reminding recipients about the approaching elections was more effective than the messages appealing to expressive and instrumental motivations to vote. Third, we document remarkably large spillover effects in voting behavior, suggesting that the behavior of adult children with voting rights may influence their parents' turnout decisions. Fourth, we obtain comprehensive evidence to conclude that the employed get-out-the-vote strategy did not exacerbate existing social inequalities in voting within our target sample, 18 to 29-year-old voters.

Our paper is complimentary to studies that have previously examined the effectiveness of text messages as a tool to mobilize voters. Our main contribution is to advance the literature on the impact evaluation of get-out-the-vote interventions and expand the existing knowledge how voter turnout transmits within the households. Our study documents new findings that hold practical implications for non-partisan political mobilization and academic scholars of voter mobilization. First, we document that several customary methods of impact evaluation implicitly assuming zero or little spillovers among social ties may underestimate the true effectiveness of voter mobilization interventions. Second, our results suggest that the previously observed compositional effects of get-out-the-vote interventions that have widened the disparities in participation by mobilizing more effectively high-propensity individuals than under-represented low-propensity citizens do not readily generalize to text-message based interventions among young voters. In fact, we observe that SMS reminders are, in the context of our study, effective at mobilizing low-propensity voters and their household members. Overall, our results hold promise that impersonal but inclusive means of communication, like text messages, may not only successfully raise aggregate voter turnout but also encourage less likely voters to turn out their vote.

More generally, our paper advances the literature that has begun to examine how different sub-populations respond to a given treatment and assess the potential of enhancing the effectiveness of behavioral interventions through selective targeting of existing interventions. The application of a causal forest machine learning algorithm to our empirical setting does not reveal significant heterogeneities in treatment effects, suggesting limited potential for increasing the effectiveness of text message-based get-out-the-vote interventions through individually targeted treatments. Overall, despite the humble success of detecting heterogeneous treatments effects through a machine learning tool in the context of this study, we believe that the blend of RCT designs, comprehensive individual-level administrative datasets and suitable high-resolution predictive methods like the causal forest constitute a promising approach to enhance the effectiveness of behavioral interventions aiming to motivate behavioral change.

Our results raise new questions and directions for future research. Our consistently positive effectiveness estimates among young low-propensity voters hold promise that text message-based interventions may successfully raise turnout in this population group. A natural step towards better understanding the promises and limits of get-out-the-vote interventions as a tool to ameliorate demographic gaps in political participation is to study the effectiveness of text messages in hard-to-reach populations who may be beyond the reach of conventional get-out-the-vote interventions but are accessible through their mobile phones. Attempts to address the minuscule political participation in certain hard-to-reach populations, like young immigrants, may also substantially benefit from the tailoring of treatment designs (e.g., use of their native language) to these specific subgroups. Overall, questions about the potential impact of different message contents remain still largely unanswered. Here, the discovery of superior treatments with the most effective message contents may benefit from the development of so-called megastudy designs that test a large set of different treatments synchronously in one large sample using a common outcome

(Milkman et al., 2021; Duckworth and Milkman, 2022). Likewise, recent developments in the design of adaptive experimental designs that dynamically allocate larger assignment probabilities to more promising treatments hold a promise to hasten the discovery of superior treatments to increase voter turnout (Offer-Westort et al., 2021).

In order to explore whether our results generalize to higher salience parliamentary elections, and to study persistence of the effect and dynamic effects (i.e. receiving reminder during before the county and parliamentary elections vs. only before the parliamentary elections), we conducted similar RCT during the 2023 parliamentary elections. Results from this study are available shortly.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge, "When Should You Adjust Standard Errors for Clustering?," *The Quarterly Journal of Economics*, 10 2022, *138* (1), 1–35.

Arceneaux, Kevin and David W Nickerson, "Who is mobilized to vote? A re-analysis of 11 field experiments," *American Journal of Political Science*, 2009, *53* (1), 1–16.

Bergh, Johannes and Dag Arne Christensen, "Getting out the vote in different electoral contexts: The effect of impersonal voter mobilization techniques in middle and high salience Norwegian elections," *Journal of Elections, Public Opinion and Parties*, 2022, pp. 1–17.

_ , _ , and Richard E Matland, "When is a reminder enough? Text message voter mobilization in a European context," *Political Behavior*, 2021, *43* (3), 1091–1111.

Bhatti, Yosef, Jens Olav Dahlgaard, Jonas Hedegaard Hansen, and Kasper M Hansen, "How voter mobilization from short text messages travels within households and families: Evidence from two nationwide field experiments," *Electoral Studies*, 2017, *50*, 39–49.

_ , _ , _ , and _ , "Moving the campaign from the front door to the front pocket: Field experimental evidence on the effect of phrasing and timing of text messages on voter turnout," *Journal of Elections, Public Opinion and Parties*, 2017, *27* (3), 291–310.

Brennan, Geoffrey and Alan Hamlin, "Expressive voting and electoral equilibrium," *Public choice*, 1998, *95* (1), 149–175.

_ and Michael Brooks, "Expressive voting," in "The Elgar Companion to Public Choice, Second Edition," Edward Elgar Publishing, 2013.

Dale, Allison and Aaron Strauss, "Don't forget to vote: Text message reminders as a mobilization tool," *American Journal of Political Science*, 2009, *53* (4), 787–804.

Downs, Anthony, *An Economic Theory of Democracy*, New York: Harper, 1957.

Duckworth, Angela L and Katherine L Milkman, "A guide to megastudies," *PNAS Nexus*, 2022, *1* (5), pgac214.

Duflo, Esther, Rachel Glennerster, and Michael Kremer, "Using randomization in development economics research: A toolkit," *Handbook of development economics*, 2007, *4*, 3895–3962.

Enos, Ryan D, Anthony Fowler, and Lynn Vavreck, "Increasing inequality: The effect of GOTV mobilization on the composition of the electorate," *The Journal of Politics*, 2014, *76* (1), 273–288.

Fowler, Anthony, "Electoral and policy consequences of voter turnout: Evidence from compulsory voting in Australia," *Quarterly Journal of Political Science*, 2013, *8* (2), 159–182.

_ , "Regular voters, marginal voters and the electoral effects of turnout," *Political Science Research and Methods*, 2015, *3* (2), 205–219.

Green, Donald P and Alan S Gerber, *Get out the vote: How to increase voter turnout*, Brookings Institution Press, 2019.

_ , Mary C McGrath, and Peter M Aronow, "Field experiments and the study of voter turnout," *Journal of Elections, Public Opinion and Parties*, 2013, *23* (1), 27–48.

Harjunen, Oskari, Tuukka Saarimaa, and Janne Tukiainen, "Love Thy (Elected) Neighbor? Residential Segregation, Political Representation, and Local Public Goods," *The Journal of Politics*, 2023, *Forthcoming.*

Hastie, Trevor, Robert Tibshirani, and Martin Wainwright, "Statistical learning with sparsity," *Monographs on statistics and applied probability*, 2015, *143*, 143.

Hodler, Roland, Simon Luechinger, and Alois Stutzer, "The Effects of Voting Costs on the Democratic Process and Public Finances," *American Economic Journal: Economic Policy*, February 2015, *7* (1), 141–71.

Holbein, John B and D Sunshine Hillygus, *Making young voters: Converting civic attitudes into civic action*, Cambridge University Press, 2020.

Hooghe, Marc and Anna Kern, "The tipping point between stability and decline: Trends in voter turnout, 1950–1980–2012," *European Political Science*, 2017, *16*, 535–552.

Jääskeläinen, Arto, "The Finnish Election System: Overwiew," 2020.

Lijphart, Arend, "Unequal participation: Democracy's unresolved dilemma - Presidential address, American Political Science Association, 1996," *American Political Science Review*, 1997, *91* (1), 1–14.

Lim, Stephen S, Rachel L Updike, Alexander S Kaldjian, Ryan M Barber, Krycia Cowling, Hunter York, Joseph Friedman, R Xu, Joanna L Whisnant, Heather J Taylor et al., "Measuring human capital: A systematic analysis of 195 countries and territories, 1990–2016," *The Lancet*, 2018, *392* (10154), 1217–1234.

Lyytikäinen, Teemu and Janne Tukiainen, "Are voters rational?," *European Journal of Political Economy*, 2019, *59*, 230–242.

Malhotra, Neil, Melissa R Michelson, Todd Rogers, and Ali Adam Valenzuela, "Text messages as mobilization tools: The conditional effect of habitual voting and election salience," *American Politics Research*, 2011, *39* (4), 664–681.

McClean, Charles T, "Does the Underrepresentation of Young People in Political Institutions Matter for Social Spending?," *Unpublished manuscript*, 2021.

Milkman, Katherine L, Dena Gromet, Hung Ho, Joseph S Kay, Timothy W Lee, Pepi Pandiloski, Yeji Park, Aneesh Rai, Max Bazerman, John Beshears et al., "Megastudies improve the impact of applied behavioural science," *Nature*, 2021, *600* (7889), 478–483.

Mo, Cecilia Hyunjung, John B Holbein, and Elizabeth Mitchell Elder, "Civilian national service programs can powerfully increase youth voter turnout," *Proceedings of the National Academy of Sciences*, 2022, *119* (29), e2122996119.

Naess, Ole-Andreas Elvik, "Increasing turnout with a text message: Evidence from a large campaign from the government," *Journal of Elections, Public Opinion and Parties*, 2022, pp. 1–19.

Nickerson, David W, "Is voting contagious? Evidence from two field experiments," *American Political Science Review*, 2008, *102* (1), 49–57.

Offer-Westort, Molly, Alexander Coppock, and Donald P Green, "Adaptive experimental design: Prospects and applications in political science," *American Journal of Political Science*, 2021, *65* (4), 826–844.

Prete, Anna Lo and Federico Revelli, "Voter Turnout and City Performance: Evidence from Italian Municipalities," *The Journal of Law, Economics, and Organization*, 09 2020, *37* (1), 168–197.

Sinclair, Betsy, Margaret McConnell, and Donald P Green, "Detecting spillover effects: Design and analysis of multilevel experiments," *American Journal of Political Science*, 2012, *56* (4), 1055–1069.

Solijonov, Abdurashid, *Voter turnout trends around the world*, IDEA, 2016.

Thaler, Richard H and Cass R Sunstein, *Nudge: Improving decisions about health, wealth, and happiness*, Penguin, 2009.

Wager, Stefan and Susan Athey, "Estimation and inference of heterogeneous treatment effects using random forests," *Journal of the American Statistical Association*, 2018, *113* (523), 1228–1242.

Zou, Hui and Trevor Hastie, "Regularization and variable selection via the elastic net," *Journal of the royal statistical society: series B (statistical methodology)*, 2005, *67* (2), 301–320.

Online Appendix for

Who is mobilized to vote by short text messages? Evidence from a nationwide field experiment with young voters

Salomo Hirvonen, Maarit Lassander, Lauri Sääksvuori, Janne Tukiainen*

Contents

*Hirvonen: Department of Economics, University of Turku, Rehtorinpellonkatu 3, FI-20014 Turku, Finland. Lassander: Prime Minister's Office, Snellmaninkatu 1, FI-00023 Helsinki, Finland. Sääksvuori: Finnish Institute for Health and Welfare. Mannerheimintie 166, FI-00271, Helsinki, Finland. Tukiainen (corresponding author): Department of Economics, University of Turku, Rehtorinpellonkatu 3, FI-20014 Turku, Finland, janne.tukiainen@utu.fi, +358503083620. We thank the seminar participants at the University of Helsinki and Laurea for helpful comments. We are grateful to the Ministry of Justice (Finland) and the Prime Minister's Office (Finland) for funding and help.

A Supplementary results

A.1 Supplementary Tables

Table A1: Heterogeneity by Thirds of Vote Propensity

	Voted			
	All	Low Propensity {Bottom 33%}	Marginal Voters {33-67%}	High Propensity {Top 33%}
	(1)	(2)	(3)	(4)
Panel A: Direct Effects				
Treated	0.009***	0.022***	0.010**	-0.006
	(0.003)	(0.007)	(0.005)	(0.007)
Controls	✓	✓	✓	✓
Control group $\bar{Y}$	0.309	0.166	0.298	0.483
Observations	49.458	16.321	16.814	16.323
Differences		Marginal - Low	Marginal - High	High - Low
		-0.012	0.016*	-0.027***
		(0.008)	(0.008)	(0.010)
Panel B: Spillover Effects				
Treated in HH	0.013**	0.025***	0.014**	-0.003
	(0.006)	(0.008)	(0.008)	(0.010)
Controls	✓	✓	✓	✓
Control group $\bar{Y}$	0.497	0.272	0.494	0.727
Observations	36.723	12.118	12.486	12.119
Differences		Marginal - Low	Marginal - High	High - Low
		-0.011	0.017	-0.028**
		(0.012)	(0.013)	(0.013)

Notes: *** $p < 0.01$, ** $p < 0.05$,* $p < 0.1$, standard errors clustered at the municipal level in parentheses. Controls include gender, age, ethnicity, ln taxable income, SES background groups, education (high school completion) and an indicator variable whether individual was eligible to vote for the first time.

Table A2: Heterogeneity by Thirds of Vote Propensity - Elastic Net

	Voted			
	All	Low Propensity {Bottom 33%}	Marginal Voters {33-67%}	High Propensity {Top 33%}
	(1)	(2)	(3)	(4)
Panel A: Direct Effects				
Treated	0.009***	0.009	0.009	0.009
	(0.003)	(0.007)	(0.006)	(0.006)
Controls	✓	✓	✓	✓
Control group $\bar{Y}$	0.308	0.176	0.294	0.451
Observations	49.679	16.266	16.718	16.695
Differences		Marginal - Low	Marginal - High	High - Low
		-0.000	0.001	0.000
		(0.009)	(0.008)	(0.009)
Panel B: Spillover Effects				
Treated in HH	0.013**	0.013	0.017**	0.007
	(0.006)	(0.009)	(0.008)	(0.010)
Controls	✓	✓	✓	✓
Control group $\bar{Y}$	0.496	0.277	0.494	0.720
Observations	36.876	12.169	12.537	12.170
Differences		Marginal - Low	Marginal - High	High - Low
		0.004	0.009	0.006
		(0.012)	(0.013)	(0.013)

Notes: *** $p < 0.01$, ** $p < 0.05$,* $p < 0.1$, standard errors clustered at the municipal level in parentheses. Controls include gender, age, ethnicity, ln taxable income, SES background groups, education (high school completion) and an indicator variable whether individual was eligible to vote for the first time.

Table A3: Average Treatment Effect - Logit Model

	Voted			
	(1)	(2)	(3)	(4)
Treatment (pooled)	0.041***	0.042**	0.045***	0.047***
	(0.015)	(0.016)	(0.016)	(0.017)
Controls				
Gender		✓	✓	✓
Age		✓	✓	✓
Ethnicity		✓	✓	✓
Ln income		✓	✓	✓
SES			✓	✓
Education			✓	✓
First-time voter			✓	✓
Municipality FE				✓
Control group $\bar{Y}$	0.307	0.308	0.308	0.309
Observations	50.140	49.679	49.679	49.599

Notes: *** p < 0.01, ** p < 0.05 ,* p < 0.1, standard errors clustered at the municipal level in parentheses. Controls include gender, age, ethnicity, ln taxable income, SES background groups, education (high school completion) and an indicator variable whether individual was eligible to vote for the first time.

Table A4: ATE for spillovers - Logit Model

	Voted			
	(1)	(2)	(3)	(4)
Treatment (pooled)	0.054**	0.060**	0.060**	0.052**
	(0.025)	(0.024)	(0.026)	(0.026)
Controls				
Gender		✓	✓	✓
Age		✓	✓	✓
Ethnicity		✓	✓	✓
Ln income		✓	✓	✓
SES			✓	✓
Education			✓	✓
First-time voter			✓	✓
Municipality FE				✓
Control group $\bar{Y}$	0.494	0.496	0.496	0.496
Observations	37.207	36.876	36.876	36.796

Notes: *** p < 0.01, ** p < 0.05 ,* p < 0.1, standard errors clustered at the municipal level in parentheses. Controls include gender, age, ethnicity, ln taxable income, SES background groups, education (high school completion) and an indicator variable whether individual was eligible to vote for the first time.

Table A5: Covariates by Mother Identified

	Mother Known	Mother Not Known
Female	0.39	0.43
	(0.49)	(0.49)
Age	23.15	27.47
	(2.57)	(2.00)
Highschool degree	0.33	0.39
	(0.47)	(0.49)
Taxable Income	12410.49	21910.64
	(11599.83)	(13618.41)
Immigrant	0.03	0.07
	(0.16)	(0.25)
Observations	33.509	17.592

Notes: Standard deviation in parenthesis.

Table A6: Covariates by Voting Propensity - Logit

	Low Propensity	Marginal Voters	High Propensity
Female	0.14	0.40	0.69
	(0.35)	(0.49)	(0.46)
Age	24.92	24.54	24.47
	(3.03)	(3.10)	(3.30)
High school Background	0.05	0.41	0.91
	(0.21)	(0.49)	(0.29)
Taxable Income	18837.28	15423.02	13387.01
	(14680.37)	(12725.48)	(11706.14)
Immigrant Background	0.15	0.01	0.00
	(0.35)	(0.08)	(0.02)
Observations	12.476	25.046	12.487

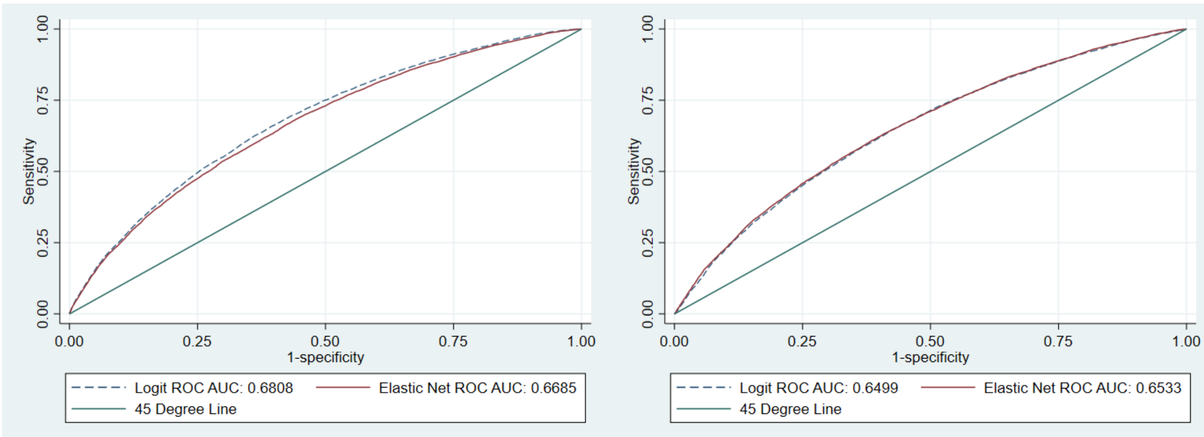
Note: Standard deviation in parenthesis.

Table A7: Covariates by Voting Propensity - Elastic Net Logit

	Low Propensity	Marginal Voters	High Propensity
Female	0.06	0.42	0.72
	(0.25)	(0.49)	(0.45)
Age	24.89	24.51	24.58
	(3.03)	(3.12)	(3.30)
High school Background	0.01	0.39	0.97
	(0.10)	(0.49)	(0.18)
Taxable Income	19306.74	15185.56	13388.72
	(14947.32)	(12577.50)	(11550.46)
Immigrant Background	0.15	0.01	0.00
	(0.35)	(0.09)	(0.02)
Observations	12768	25357	12673

Note: Standard deviation in parenthesis.

A.2 Supplementary Figures



(a) ROC Curves for In-Sample Prediction

(b) ROC Curves for Out-of-Sample Prediction

Notes: For out-of- sample prediction sample is randomly split in order to estimate the model in the other half of the sample and compute the prediction error in the other half of the sample.

Figure A1: Receiving Operating Characteristic (ROC) Curves