

Title of the Research Paper:

Viral Voting: A study in the information diffusion effects of different social pressure treatments on voter turnout.

Author Name:

Damanveer Singh Dhaliwal

Department of Economics
University of Toronto

December 16, 2025

1 Introduction

The most important aspect of a democratic society is the ability of its citizens to vote and select their representatives in the government. However, voter turnout has been declining consistently across all developed countries (Solijonov (2016)). In the 2024 US Presidential Election, the voter turnout was 63.9% percent, which was 2.7% percent lower than the previous election. Similar trends can be observed in Canada, the UK, and Europe.

In order to ensure that representation across government levels is fair and stable, it is very important to understand what influences voter turnout. In their 2006 field experiment, Gerber, Green, and Larimer tackled this question and tried to isolate how voter turnout might be influenced by social pressures such as civic duty, peer pressure, etc. (Gerber et al. (2008)). Their research question studied the direct effects that the social pressure had on a voter turnout of a household. In this paper, we are going to use the data from their field experiment and isolate how information diffuses through a social network, especially concerning different types of social pressure. We estimate whether social pressure treatments generate information diffusion effects that influence untreated voters and whether the magnitude of these diffusion effects differ by treatment type.

Understanding how social pressure shapes information diffusion reveals to us the dynamics that neither phenomenon produces alone. And these dynamics have now provided some key results across public health, election results, and the spread of both knowledge and misinformation across billions of people. Across disciplines, researchers have found that information simply just does not diffuse through a network like a disease. It creates new societal rules and reshapes the very networks through which it spreads.

To study the information diffusion effects of social pressure on voter turnout, we hypothesize that social pressure treatments like the one used by Gerber et al. (2008) will have positive diffusion effects on untreated individuals in the same neighborhood. We also hypothesize that the magnitude of these diffusion effects will differ by treatment type, with more intrusive treatments (like the Neighbors treatment) generating stronger diffusion effects compared to less intrusive treatments (like the Civic Duty treatment). We will test these hypotheses using a combination of econometric models and causal machine learning methods to analyze the data from the original field experiment.

Hong (2023) showed how positive social pressure from the public authority led to lower skepticism about the COVID-19 vaccine and increased vaccination rates. However, the study missed the effect of the information transfer among the citizens on the country and how that led to a positive outcome. Bond et al. (2012) found results similar to Gerber et al. (2008) where people who received social pressure treatments with familiar faces were 0.39% more likely to vote than people who received a treatment with a stranger's face. Bond et al. (2012) also found that the social contagion/diffusion effect resulted in an additional 280,000 voters compared to the direct effects of approximately 60,000 voters. This is similar to the research question we are trying to answer in this study. However, we are trying to isolate the diffusion effects of different types of social pressure treatments.

Atienza-Barthelemy et al. (2025) created a model to study the diffusion of information through social media. They found that social media allows for rapid dissemination of information but the

longevity and the reach of the content is severely limited due to the availability of an overwhelming amount of information on these platforms. Given our data is from a field experiment in 2006, we are unable to study similar metrics, however, this can easily be a future extension of this project.

In a similar project, Vosoughi et al. (2018) studied the spread of true and false news stories on Twitter. They found that false news had deeper and more rapid information diffusion effects. They also discovered that the diffusion effects of false news were human led and not influenced by bots. This is an interesting result and is the motivation behind our research question. We want to isolate which types of social pressure diffuse more effectively to combat misinformation as well as to allow for positive social pressure to be more effective.

Haenschen (2016) studied the effect of pride and shame social pressure treatments on voter turnout and found a treatment effect of 15.8% to 24.3%. However, they did not find any significant effect from indirect social pressure or diffusion. Their empirical strategy was to use a linear probability model with OLS estimation. We will be using a similar empirical strategy but we will extend the empirical work to include spatial autoregressive models and causal machine learning methods to better isolate the diffusion effects.

The rest of the paper is structured as follows. Section 2 describes the original experiment and the data. Section 3 describes the theoretical framework and the empirical strategy. Section 4 presents the results and section 5 concludes. We do include a future extensions section after the conclusion to discuss how this project can be extended further.

2 Data and Summary Statistics

The data used in this project is from a field experiment conducted by Gerber et al. (2008) prior to the August 2006 primary elections in the state of Michigan, United States. The August 2006 primary election was a statewide election with offices across the political spectrum being contested. The data was collected from 180,002 households and every individual residing at the same address was considered to be a part of the household. Individuals with missing records, and individuals living in apartment complexes were not included. Individuals living on streets with fewer than 10 neighbours were also excluded from the experiment.

Prior to the random assignment of the treatment, individuals who were more likely to vote by an absentee ballot were excluded to control for pre-treatment voting decisions. Individuals who had not voted in the extremely high turnout 2004 general election were considered to have moved, died or double registered and were also excluded from the study.

The remaining 180,002 households were randomly assigned to one of the four treatment groups or the control group. The control group received no mailings. Each household in one of the treatment groups received one of the four mailings 11 days prior to the election. Each mailing included "DO YOUR CIVIC DUTY - VOTE!" along with a varying degree of social pressure. The four treatment groups were as follows:

The **Civic Duty** treatment was the least intrusive and included the message: "Remember your rights and responsibilities as a citizen. Remember to vote." The **Hawthorne** treatment told

households "YOU ARE BEING STUDIED" and that their voting behaviour was being monitored for research purposes. The **Self** treatment included a message that informed the household that their voting record was public information and included their public record for the previous elections. Finally, the **Neighbors** treatment was the most intrusive and included a message that informed the household that their voting record was public information and included their public record for the previous elections. Additionally, it also included the voting records of their neighbours. Both the Self and the Neighbors treatments included the fact that an updated voting record would be mailed to them after the election.

The outcome variable of interest is whether an individual voted in the August 2006 primary election. The data also collected information from the previous elections in 2000, 2002 and 2004 that was supplemented from the Qualified Voter File. Finally, the data included demographic information such as age, gender and party affiliation coupled with census data of the household's neighbourhood.

Table 1: Summary Statistics

Treatment	N	Female	Avg Age	Voted 2004 Prop	Avg HH Income	Outcome
Control	191243	0.499	49.814	0.400	60176.515	0.297
Neighbors	38201	0.500	49.853	0.407	60132.334	0.378
Hawthorne	38204	0.499	49.705	0.403	60145.460	0.322
Civic Duty	38218	0.500	49.659	0.399	60123.159	0.315
Self	38218	0.500	49.793	0.402	60136.146	0.345

The summary statistics in Table 1 show that the average age of individuals in the sample is around 50 years old in each treatment group. Furthermore, we can also see that the voting proportion for the 2004 primary elections is almost identical at around 40%. We can also see that the median income for each group is around \$60,000 overall. The summary statistics show that the randomization was effective and the treatment groups are balanced across key demographic variables.

Table 2: Relationship between 2004 and 2006 Voting

treatment	No Vote 2004	Vote 2004	Lift Non Voters	Lift Voters
Control	0.237	0.386	0.000	0.000
Civic Duty	0.255	0.403	0.018	0.018
Hawthorne	0.260	0.414	0.023	0.028
Self	0.278	0.445	0.041	0.059
Neighbors	0.306	0.482	0.069	0.097

Table 2 shows the relationship between voting in the 2004 primary election and the 2006 primary election. The lift columns show the proportion of individuals who voted in 2006 given their voting history in 2004. We can see that there is a positive correlation and individuals who voted in 2004 are more likely to vote in 2006. This is an important control variable that we will include in our regression models to account for prior voting behaviour.

The data was downloaded from a Stanford GSB repository. Given the field experiment's thorough design and randomization, we are confident that the data is of high quality and can be used for our analysis. The data was already cleaned and there are no major issues with missing data or outliers.

3 Theoretical Model & Empirical Framework

To model the information diffusion effects of social pressure on voting, we will borrow a network diffusion model from Myers et al. (2012). The original model is designed to capture the spread of information through a social network. In our case, we will adapt the model to capture it based on geographical proximity. Effects are bound to be different since social media connections allow for rapid and far reaching dissemination of information.

3.1 Model Description

The original model by Myers et al. (2012) has three main components: the contagion, internal exposure and external exposure. With respect to the research question at hand, we define these components as follows:

We define **contagion** as the information about social pressure and the public disclosure of voting behaviour. A household becomes **infected** when they become aware that (1) voting behavior is being monitored, (2) their voting records may be publicized to neighbors, and (3) voting is subject to social surveillance. This information can spread through both the experimental mailings and through social diffusion within neighborhoods. Note that contagion is only the information itself, while infection is the state of being aware of this information.

An **internal exposure** ($\sum T_j$) occurs when a household learns about the social pressure treatments from their neighbors who received the mailings. This includes both direct social interactions (neighbors discussing the mailings) and indirect observations (seeing neighbors vote, observing physical mailings, or experiencing increased social pressure within the neighborhood). Internal exposures capture all information transmission along the edges of our observed network structure.

External exposures (T_i) represent the instances when information reaches a household from sources outside their immediate neighbor network. In our experiment, external exposures correspond to the direct receipt of a social pressure mailing.

3.2 Model Specification and Linear Approximation

We assume that the probability of a household i voting is a function of its infection state. Let K_i denote the total number of exposures household i receives. Mathematically, we can write this as:

$$K_i = T_i + \lambda \sum_j T_j$$

Where:

- $T_i \in \{0, 1\}$ indicates if household i received a social pressure mailing (external exposure).
- $\sum_j T_j$ is the count of treated neighbors of household i .
- $\lambda \geq 0$ is the transmission rate that captures how much of a neighbor's treatment effect spills over to household i (internal exposure).

The probability that household becomes infected (i.e., votes) after receiving K_i exposures is given by the probability that at least one exposure leads to infection. In Myers et al. (2012), this is modelled non-linearly as:

$$P(\text{Infected}_i | K_i) = 1 - (1 - \eta)^{K_i}$$

where η is the probability that a single exposure leads to infection.

3.2.1 Linear Model

Directly estimating η is challenging because information diffusion is unobserved. To allow for estimation using observed data, we linearize the model using a first-order Taylor expansion around $K_i = 0$. This yields:

$$P(\text{Infected}_i | K_i) \approx \eta K_i = \eta(T_i + \lambda \sum_j T_j)$$

The key assumption here is that the probability of infection increases linearly with the number of exposures. Given that we use a Zip+4 based neighborhood definition, we expect most households to have a small number of treated neighbors, making the linear approximation reasonable.

Under this linear model, the probability that household i votes can be expressed as:

$$P(\text{Vote}_i) = \alpha + P(\text{Infected}_i | K_i) + \epsilon_i = \alpha + \eta T_i + \eta \lambda \sum_j T_j + \epsilon_i \quad (1)$$

This parameterization allows us to estimate the direct effect of receiving a social pressure mailing (η) and the diffusion effect from having more neighbors treated ($\eta \lambda$) using econometric methods.

We can estimate the reduced form of the model using OLS regression as follows:

$$\text{Vote}_i = \alpha + \beta_1 T_i + \beta_2 \sum_j T_j + \epsilon_i \quad (2)$$

The structural parameters of interest can be recovered as $\eta = \beta_1$ and $\lambda = \beta_2 / \beta_1$.

A key limitation of this linear approximation is that it does not capture non-linearities in the diffusion process such as a decrease in η as more exposures are received. We assume that the machine learning methods employed later will help capture some of these non-linearities.

4 Empirical Results

4.1 Identification Strategy

In the original study by Gerber et al. (2008), the authors used a randomized controlled trial (RCT) design to identify the causal effects of different social pressure treatments on voter turnout. The experiment and the randomization process have already been described in Section 2.

This paper seeks to study the causal relationship between social pressure and its information diffusion effect (are neighbours of treated individuals acting differently than neighbours of untreated individuals?). We will rely on the randomization of the treatment assignment as well to identify the information diffusion effects of social pressure on voter turnout. Social pressure cannot be directly tracked, but the treatment by Gerber et al. (2008) presents a unique opportunity to study how information about social pressure diffuses through a social network.

For the linear model, randomization ensures that the treatment variables and the neighbors treated count are both exogenous. Therefore, we can estimate the parameters of interest using OLS regression. β_1 and β_2 can be interpreted as the average treatment effect of receiving a social pressure mailing and the average diffusion effect from having more neighbors treated, respectively.

$$\text{ATE}_{T_i} = E[Y_i(1, n) - Y_i(0, n)] = \beta_1$$

$$\text{ATE}_{S_i} = E[Y_i(t, n+1) - Y_i(t, n)] = \beta_2$$

Where $Y_i(t, n)$ is the potential outcome for household i when they receive treatment t and have n neighbors treated. Specifically, under random assignment, we are trying to estimate:

$$P(\text{vote}|T_i, S_i, X_i) = P(\text{vote}|\text{do}(T_i), S_i, X_i)$$

We assume that due to randomization, the treatment assignment is independent of any unobserved confounders that may affect voter turnout. The unexplained portion of the outcome variable is captured in the error term ϵ_i which is assumed to be orthogonal to the treatment variables.

4.2 Directed Acyclic Graphs

Here's a DAG to illustrate the relationships between the variables in our study:

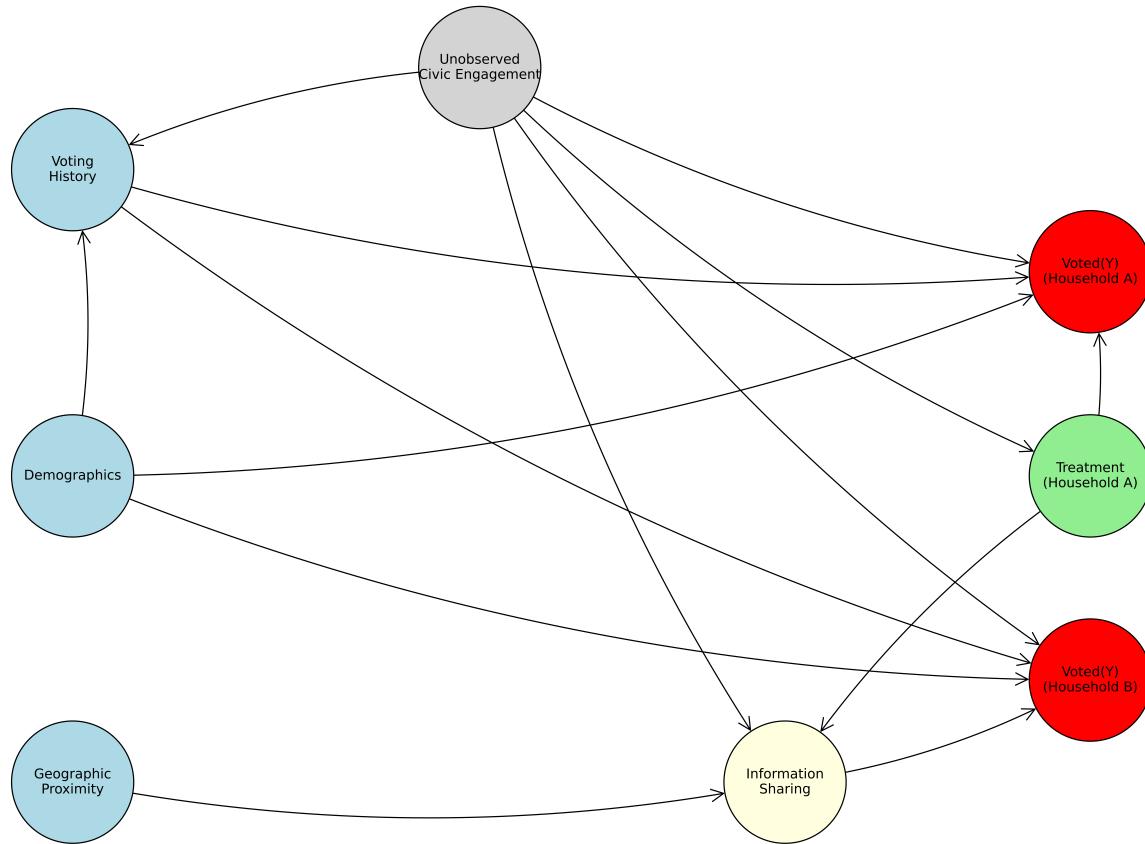


Figure 1: Directed Acyclic Graph (DAG)

In this DAG, the treatment groups (Civic Duty, Hawthorne, Self, Neighbors) directly influence the outcome variable (Voted in 2006). The control variables as well geographic proximity. (clustering) affects the information sharing node which affects the outcome variable for the control group.

4.3 Identification Assumptions

The identification strategy relies on the assumption that the treatment assignment is random and uncorrelated with the error term. Given the randomization in the original experiment, this assumption is likely to hold. The balance table below shows the average characteristics of individuals across the treatment groups and the control group. For brevity, we only show a subset of the variables:

Table 3: Balance Table

	Control	Civic Duty	Hawthorne	Self	Neighbors
Household Size	2.18	2.19	2.18	2.18	2.19
Nov 2002	0.81	0.81	0.81	0.81	0.81
Nov 2000	0.84	0.84	0.84	0.84	0.84
Aug 2004	0.40	0.40	0.40	0.40	0.41
Aug 2002	0.39	0.39	0.39	0.39	0.39
Aug 2000	0.25	0.25	0.25	0.25	0.25
Female	0.50	0.50	0.50	0.50	0.50
Age (in years)	49.81	49.66	49.70	49.79	49.85
N	191243.00	38218.00	38204.00	38218.00	38201.00

This shows that the randomization was effective. However, there may be some concerns about spillover effects from the Neighbors treatment to the control group.

To estimate the effect of these diffusion effects, consider the following plot showing the voter turnout in the control group against the neighbourhood treatment intensity:

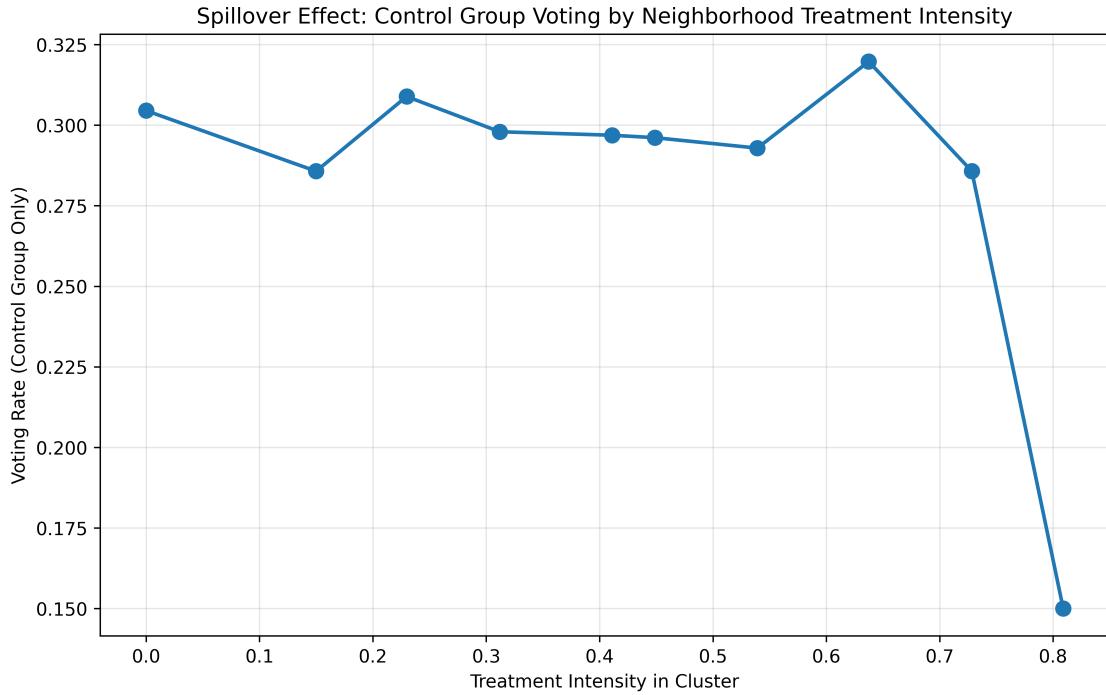


Figure 2: Voter Turnout in Control Group by Neighbourhood Treatment Intensity

Voter turnout in the control group is pretty flat with respect to the neighbourhood treatment intensity but drops off at the higher end steeply. This can be accounted for by the randomization process which prevented any neighborhoods having a high intensity of treated individuals. This

suggests that there are probably no major diffusion effects from the treatments to the control group and that λ is probably small or zero and that η should be the main driver of the voter turnout. Given the linear approximation of the reduced form, we expect that there be a bias towards zero in the estimate of β_2 .

4.4 OLS

To estimate the direct effects of the different social pressure treatments on voter turnout, we will explore all the widely used econometric methods in the literature. We will start with a simple OLS regression of the form:

$$Y_i = \beta_0 + \beta_1 CivicDuty_i + \beta_2 Hawthorne_i + \beta_3 Self_i + \beta_4 Neighbors_i + \beta_5 X_i + \epsilon_i \quad (3)$$

Where Y_i is a binary variable that takes the value of 1 if the individual voted in the 2006 primary election and 0 otherwise. $CivicDuty_i$, $Hawthorne_i$, $Self_i$ and $Neighbors_i$ are binary variables that take the value of 1 if the individual was assigned to that treatment group and 0 otherwise. The control group is the omitted category. ϵ_i is the error term. X_i is a vector of control variables that includes sex, year of birth and voting history in the 2004 primary elections. We will also cluster the standard errors at the household level to account for any correlation in the error terms within households.

Now to estimate the diffusion effects, we will include the neighborhood treatment intensity variable in the regression. The neighborhood treatment intensity is defined as the proportion of treated households in the same block as the individual. The regression model is specified as:

$$Y_i = \beta_0 + \beta_1 CivicDuty_i + \beta_2 Hawthorne_i + \beta_3 Self_i + \beta_4 Neighbors_i + \beta_5 Intensity_i + \beta_6 X_i + \epsilon_i \quad (4)$$

Where $Intensity_i$ is the neighborhood treatment intensity variable.

It is intuitive that treated individuals will also face certain diffusion effects, therefore we will include an interaction term between the treatment variables and the neighborhood treatment intensity variable. The regression model is specified as:

$$\begin{aligned} Y_i = & \beta_0 + \beta_1 CivicDuty_i + \beta_2 Hawthorne_i + \beta_3 Self_i + \beta_4 Neighbors_i + \beta_5 Intensity_i \\ & + \beta_6 (CivicDuty_i \times Intensity_i) + \beta_7 (Hawthorne_i \times Intensity_i) \\ & + \beta_8 (Self_i \times Intensity_i) + \beta_9 (Neighbors_i \times Intensity_i) \\ & + \beta_{10} X_i + \epsilon_i \end{aligned}$$

Results from all three regression models are presented in Table 2.

Table 4: OLS Regression Results with Diffusion Effects

	Baseline	Diffusion	Interaction
Constant	7.9765*** (0.1193)	7.9890*** (0.1195)	7.9924*** (0.1195)
Civic Duty	0.0187*** (0.0032)	0.0187*** (0.0032)	0.0162 (0.0140)
Hawthorne	0.0257*** (0.0032)	0.0258*** (0.0032)	-0.0030 (0.0139)
Neighbors	0.0802*** (0.0033)	0.0803*** (0.0033)	0.0677*** (0.0146)
Self	0.0483*** (0.0032)	0.0484*** (0.0032)	0.0581*** (0.0143)
Female	-0.0080*** (0.0011)	-0.0080*** (0.0011)	-0.0080*** (0.0011)
Year of Birth	-0.0040*** (0.0001)	-0.0040*** (0.0001)	-0.0040*** (0.0001)
Voted Primary 2004	0.1482*** (0.0019)	0.1482*** (0.0019)	0.1482*** (0.0019)
Treatment Intensity		-0.0167* (0.0099)	-0.0256* (0.0132)
Civic Duty_x_intensity			0.0061 (0.0320)
Hawthorne_x_intensity			0.0677** (0.0317)
Neighbors_x_intensity			0.0296 (0.0334)
Self_x_intensity			-0.0227 (0.0327)
R-squared	0.0450	0.0450	0.0451
R-squared Adj.	0.0450	0.0450	0.0450
N	344,084	344,084	344,084

Standard errors in parentheses.

* p<.1, ** p<.05, ***p<.01

As a robustness check, we also estimated the models with the data clustered at the block level. The results were qualitatively similar and can be found in the appendix.

4.4.1 Results Discussion

The OLS results are consistent with the findings of Gerber et al. (2008). All the treatment groups have a positive and statistically significant effect on voter turnout compared to the control group. The Neighbors treatment has the largest effect, increasing the probability of voting by approximately 6.77 to 8.03 percentage points. The Self treatment increases the probability of voting by 4.8 to 5.8 percentage points, the Hawthorne treatment by 2.5 percentage points and the Civic Duty treatment by 1.8 percentage points.

There is also a very significant positive effect of having voted in the 2004 primary election, increasing the probability of voting by 14.82 percentage points. Age and gender are statistically significant but the effects are very small. The treatment intensity of the neighborhood has a negative and statistically significant effect on voter turnout. This suggests that there are no positive diffusion effects from the treatments to the control group. In fact, it seems that having more treated neighbors slightly decreases the probability of voting. This could be due to free-riding behaviour where individuals rely on their neighbors to vote instead of voting themselves.

However, upon adding interaction terms with each of the treatment variables, we observe that the civic duty and hawthorn treatments no longer have a statistically significant effect on voter turnout. We also observe that the interaction term for the Hawthorne treatment is positive at 6.77 percentage points and statistically significant. This suggests that the Hawthorne treatment has very large and positive diffusion effects, where having more treated neighbors increases the probability of voting for treated individuals. The interaction terms for the Self, Civic Duty and Neighbors treatments are not statistically significant. This suggests that there are no diffusion effects for these treatments.

Returning to our structural framework, we seek to recover the transmission rate $\lambda = \beta_2/\beta_1$ and calculate its standard error using the Delta Method approximation. However, the estimation proves to be structurally unstable from the observed parameters.

For the Civic Duty, Self and Neighbors treatments, the estimates of β_2 , the coefficient of the interaction term, are statistically indistinguishable from zero, implying no diffusion effects. However, for the Hawthorne treatment, while β_2 is positive and significant, β_1 , the coefficient for the treatment directly, becomes statistically insignificant when interaction terms are included. This instability prevents us from reliably estimating λ for any treatment group.

Therefore, we conclude that the linear structural approximation is insufficient to capture the Hawthorne dynamic, where the effect is estimated to be entirely driven by social interaction rather than direct exposure. This motivates our use of non-linear machine learning methods in the following sections.

4.5 Spatial Autoregressive Model

To further check for robustness of the lack of diffusion effects and to account for potential spatial correlation in the error terms, we will use a spatial autoregressive model (SAR) (Lord et al. (2021)). The SAR model can be specified as:

$$y_i = \rho \sum_{j=1}^N w_{ij} y_j + X_i \beta + \epsilon_i \quad (5)$$

The spatial weights matrix W is constructed using a K-Nearest Neighbors (KNN) algorithm based on geographical coordinates derived from zip codes. Specifically, for every individual i , we assign a weight of $w_{ij} > 0$ to the $k = 8$ spatially closest neighbors based on Euclidean distance, and $w_{ij} = 0$ otherwise. The weights matrix is row-standardized such that $\sum_j w_{ij} = 1$, ensuring that the spatial lag term represents the average outcome of an individual's neighbors.

y_i is the binary outcome variable indicating whether individual i voted in the 2006 primary election. ρ is the spatial autoregressive coefficient that captures the degree of spatial dependence in voter turnout. X_i is a vector of covariates including the treatment indicators (Civic Duty, Hawthorne, Self, Neighbors), neighborhood treatment intensity, and individual characteristics. The results from the SAR model are presented in Table 4 below:

Table 5: Spatial Autoregressive Model
Results

Variable	Coefficient
Constant	7.9589*** (0.1060)
Civic Duty	0.0159 (0.0111)
Hawthorne	-0.0030 (0.0112)
Self	0.0578*** (0.0112)
Neighbors	0.0678*** (0.0112)
Treatment Intensity	-0.0274*** (0.0105)
Civic Duty x Intensity	0.0068 (0.0255)
Hawthorne x Intensity	0.0677*** (0.0256)
Self x Intensity	-0.0220 (0.0255)
Neighbors x Intensity	0.0294 (0.0256)
Female	-0.0080*** (0.0015)
Year of Birth	-0.0039*** (0.0001)
Voted Primary 2004	0.1477*** (0.0016)
Spatial Lag (ρ)	0.0352** (0.0144)
Observations	344,084
Neighbors (k)	8

Standard errors in parentheses.

* p<.1, ** p<.05, *** p<.01

This table shows that the spatial autoregressive coefficient ρ is negative and statistically insignificant, indicating that there is no spatial correlation in voter turnout. The results are similar to the OLS results, with all treatment groups having similar estimates as the OLS model with the

interaction terms. Once again, without reliable estimates for both β_1 and β_2 , we cannot compute the transmission rate λ . Therefore, the SAR model corroborates the OLS findings of no significant diffusion effects from social pressure treatments on voter turnout.

4.6 Regularization

To further support our findings and to regularize our estimates, we will use LASSO and Ridge regressions. The regression model is specified as:

$$\begin{aligned} Y_i = & \beta_0 + \beta_1 \text{CivicDuty}_i + \beta_2 \text{Hawthorne}_i + \beta_3 \text{Self}_i + \beta_4 \text{Neighbors}_i + \beta_5 \text{Intensity}_i \\ & + \beta_6 (\text{CivicDuty}_i \times \text{Intensity}_i) + \beta_7 (\text{Hawthorne}_i \times \text{Intensity}_i) \\ & + \beta_8 (\text{Self}_i \times \text{Intensity}_i) + \beta_9 (\text{Neighbors}_i \times \text{Intensity}_i) \\ & + \beta_{10} X_i + \epsilon_i \end{aligned}$$

Where X_i is a vector of covariates that includes sex, year of birth, previous voting history for the years 2000, 2002 and 2004 and census-level neighbourhood data such as education, race, etc. Results from the LASSO and Ridge regressions are presented in Table 4.

Both LASSO and Ridge regression arrived at consistent coefficient estimates. The Neighbors treatment has the most effect with a LASSO coefficient of 0.1055, followed by Self (0.0734) and Civic Duty (0.0220). Consistent with the OLS, the coefficient of the Hawthorne treatment was reduced to 0 and the effect was entirely captured by the interaction term (0.0420). The coefficient of the neighborhood intensity is negative (-0.0183) and consistent.

Furthermore, the regularization explicitly selected interaction terms, revealing heterogeneity in diffusion: the interaction between Hawthorne and Intensity is the strongest (0.0420), followed by Neighbors (0.0183) and Civic Duty (0.0063). In contrast, the interaction for the Self treatment is effectively zero (-0.0005), suggesting that intrinsic motivation does not benefit from social reinforcement in the same way as external pressure. The LASSO and Ridge regression coefficient paths are plotted below:

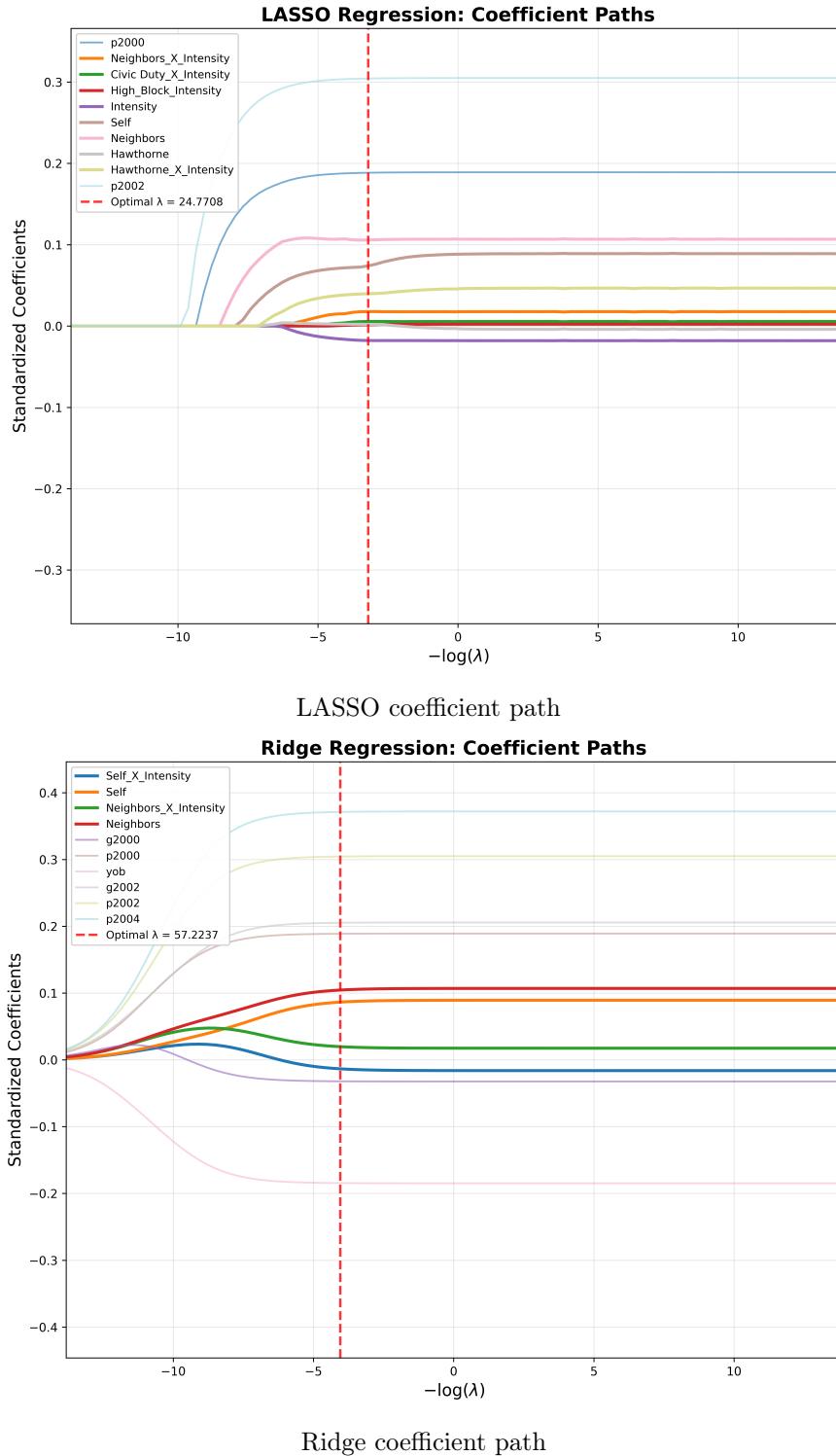


Figure 3: Regularization coefficient paths: LASSO (left) and Ridge (right)

4.7 Machine Learning Methods

4.7.1 Regression Trees and Random Forests

The results from the random tree are presented on the next page:

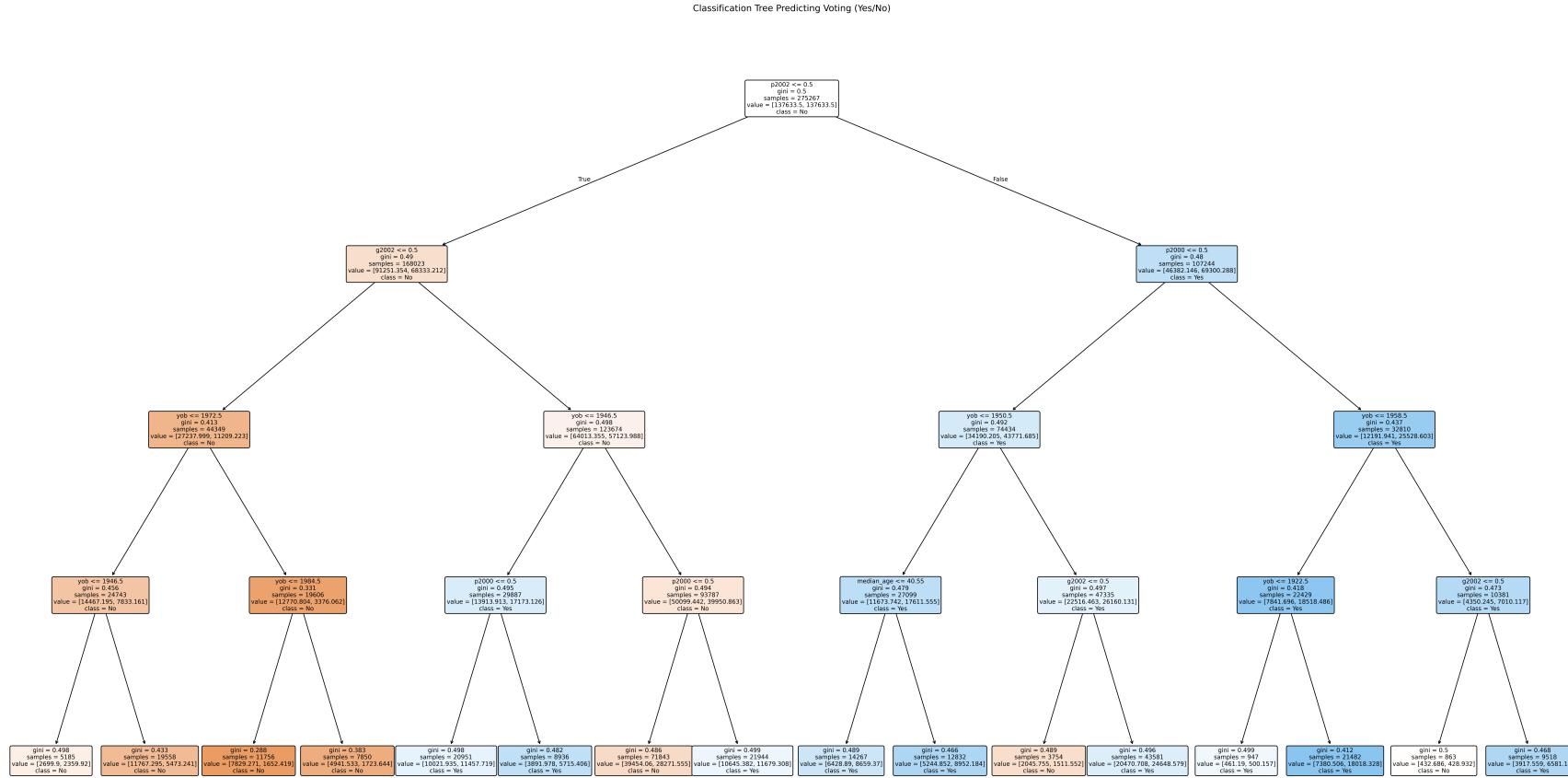


Figure 4: Regression Tree

The regression tree shows that the age, previous voting history and the Neighbors treatment are the most important variables in prediction. The regression tree also does not show that intensity of the treatment in the neighborhood is an important variable to be a leading predictor for the outcome variable. Another surprising result was that the median age in the neighborhood was an important predictor. This suggests either that older neighborhoods are more civically engaged or that older individuals influence younger individuals to vote.

The random forest importance matrix presented similar variables as the most important. The variable importance plot is presented below:

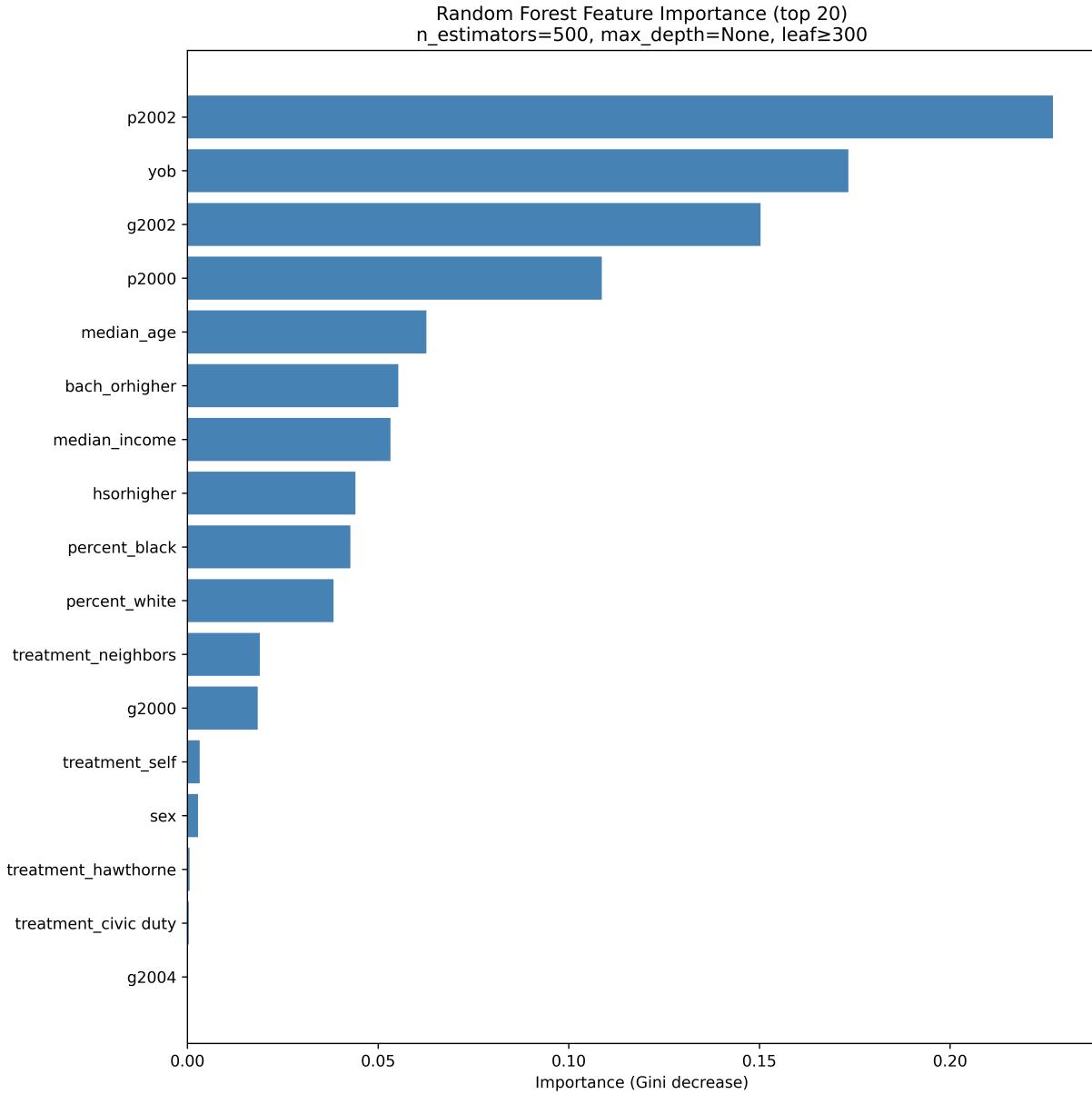


Figure 5: Random Forest Variable Importance Plot

The random forest importance matrix supports the findings from the regression tree. Age and previous voting history are important predictors, however, neighbourhood demographics are also important predictors. The treatment variables are below the neighbourhood demographics in importance with the neighbors and the self treatment being more important than hawthorne and civic duty which is in line with the regression results. The treatment intensity variable is also not an important predictor of the outcome variable.

4.7.2 Bagging and Boosting

Bagging and boosting models were also trained on the same data. For bagging, 300 trees were used with no limitation on the max depth to allow for maximum flexibility. For boosting, 100 trees were grown sequentially with a learning rate of 0.1 and a max depth of 3 to prevent overfitting.

Bagging and an accuracy of 0.617 with a log loss of 0.639 and MSE of 0.225. Boosting achieved an accuracy of 0.696 with a log loss of 0.583 and MSE of 0.199. The sequential nature of boosting allows it to capture complex patterns in the data better than the other ensemble methods.

The results from the ensemble methods are presented below:

Table 6: Model Comparison: Tree, Forest, Bagging, Boosting

Model	Accuracy	Precision	Recall	F1	ROC AUC	Log Loss	MSE
Tree	0.561	0.391	0.697	0.501	0.636	0.662	0.235
Forest	0.607	0.424	0.676	0.521	0.677	0.644	0.227
Bagging	0.617	0.432	0.668	0.524	0.681	0.639	0.225
Boosting	0.696	0.591	0.119	0.198	0.671	0.583	0.199

4.7.3 Inverse Probability Weighting

Average Treatment Effect (ATE) was estimated using Propensity Score Matching (PSM) and Inverse Probability Weighting (IPW) to account for confounding variables. Results are presented below:

Table 7: Comparison of IPW and PSM Estimates

Treatment Group	IPW ATE	PSM ATE	Difference (PSM-IPW)
Civic Duty	0.0175	0.0450	0.0276
Hawthorne	0.0247	0.0771	0.0524
Self	0.0478	0.0286	-0.0192
Neighbors	0.0814	0.0731	-0.0083

Note: Both IPW and PSM use multinomial propensity scores estimated via multinomial logistic regression. PSM uses 1-nearest neighbor matching with caliper = 0.2.

The IPW estimates are closely aligned with the original results that were presented in the paper. The propensity score matching ATE is vastly overstated. IPW weights observations by the inverse

of their treatment probability. So even when propensity scores are constant across individuals, as they are here, IPW remains valid because these constant probabilities still contain information.

In contrast, PSM relies on variation in propensity scores to distinguish good matches from poor matches. When propensity scores are constant, as they are, under perfect randomization, the matching distance between any treated unit and any control unit is zero. Therefore, PSM degenerates into random matching rather than matching on covariate similarity, introducing bias and adding spurious correlations.

For our research question with treatment intensity, the following results were obtained:

Table 8: Treatment Intensity Effects: IPW vs. PSM Estimates

Method	ATE	SE	95% CI Lower	95% CI Upper
IPW	0.0019	0.0024	-0.0029	0.0066
PSM	-0.0497	0.0024	-0.0543	-0.0451

A positivity check was performed and the distribution of propensity scores across treatment groups showed sufficient overlap, satisfying the common support assumption. The positivity check plot is presented in the appendix.

IPW presents a null treatment intensity effect that provides no evidence of information diffusion. If social pressure spread through word-of-mouth or observation, we would expect treated individuals in high-intensity neighborhoods to show higher turnout due to spillover effects. The absence of this effect suggests social pressure operates primarily through direct treatment exposure rather than neighborhood contagion.

4.7.4 S-Learner, T-Learner, X-Learner

The ATE estimates using the S learners, T learners, and X learners are presented below:

Table 9: Summary of Treatment Effect Estimates

Treatment	S-Learner	T-Learner	X-Learner	S vs T	S vs X	T vs X
Civic Duty	0.0185	0.0186	0.0186	0.0309	0.0262	0.0134
Hawthorne	0.0252	0.0251	0.0251	0.0305	0.0263	0.0133
Self	0.0474	0.0477	0.0476	0.0322	0.0283	0.0138
Neighbors	0.0801	0.0798	0.0798	0.0311	0.0269	0.0133

Notes: Pairwise model different is the Mean Absolute Error (MAE) between the CATE estimates of the two models. All models use XGBoost as the base regressor.

In order to understand the heterogeneity of treatment effects and the information diffusion process, we will analyze the average CATE estimates from the X-Learner model across 3 subgroups of treatment intensity. Results are presented below:

Table 10: ATE By Treatment Intensity (X-Learner)

Treatment Arm	Low	Mid	High
Civic Duty	0.0146	0.0207	0.0206
Hawthorne	0.0222	0.0289	0.0243
Self	0.0441	0.0483	0.0504
Neighbors	0.0743	0.0870	0.0780

The findings from the three meta learners and the analysis of the treatment heterogeneity show that there is a stability of the average treatment effect across different modeling approaches. The choice of meta learner does not alter the conclusion that social pressure works. Where the models do disagree are on the individual level prediction shown by the pairwise CATE differences on the right side of Table 7. This is probably because the S learner pools the treated and control into a single machine learning model, potentially missing out on heterogeneity, while the T learner and X learner do not. Using X learner, we have also tried to analyze how the effect changes based on treatment density in different groups.

4.7.5 Doubly Robust Models

The results from a doubly robust estimator for the treatment variables are presented below:

Table 11: Treatment Effects on Voter Turnout: Doubly Robust Estimates vs. Original Paper

Treatment	DR ATE	Interaction Effect
Civic Duty	0.0217 (0.0108, 0.0293)	0.0083 (0.0001, 0.0168)
Hawthorne	0.0193 (0.0098, 0.0281)	-0.0018 (-0.0108, 0.0094)
Self	0.0581 (0.0500, 0.0665)	0.0025 (-0.0084, 0.0150)
Neighbors	0.0844 (0.0741, 0.0924)	-0.0026 (-0.0106, 0.0082)

Note: Our DR estimates use logistic regression for propensity scores and linear regression for outcome models. 95% confidence intervals shown in parentheses, computed via bootstrap with 1000 iterations. The Interaction Effect column reflects the interaction between treatment assignment and neighborhood treatment intensity.

The treatment intensity has also been estimated using the doubly robust estimator, with results presented below:

Table 12: Treatment Effects by Neighborhood Intensity: Doubly Robust Estimates

Treatment Arm	Treatment Intensity		
	Low	Mid	High
Civic Duty	0.0148	0.0206	0.0193
Hawthorne	0.0254	0.0275	0.0241
Self	0.0435	0.0504	0.0511
Neighbors	0.0764	0.0890	0.0796

Intensity levels are: Low (0.00–0.44), Mid (0.44–0.45), High (0.45–1.00) of treatment intensity with a high concentration at 0.44–0.45. The results from the doubly robust estimator reveal a non-monotonic, inverted U relationship for three of the four treatments with effects peaking in the mid-intensity neighborhoods rather than the high-intensity ones. This contradicts simple information diffusion models which often predict monotonic increases with neighbourhood exposure. This suggests complex social dynamics where moderate intensity actually creates optimal conditions for social pressure.

The only exception is the self-treatment which shows household members their own voting records but does not publicize neighbor information. The monotonic relationship indicates that voter turnout increases as neighborhood treatment density rises, suggesting that household-level social pressure benefits from a general pro-voting atmosphere without suffering from the saturation effects that affect neighborhood-level treatments

4.7.6 Double Machine Learning & Causal Forest

The results from the non-linear Double ML model are presented below:

Table 13: Treatment Effects on Voter Turnout: Causal Forest Double ML Estimates

Treatment	CF DML ATE	SE	CATE Std	Paper ATE	Difference
Civic Duty	0.0180	0.0000	0.0063	0.0180	0.0000
Hawthorne	0.0253	0.0000	0.0084	0.0260	-0.0007
Self	0.0478	0.0000	0.0100	0.0490	-0.0012
Neighbors	0.0810	0.0000	0.0143	0.0810	0.0000

The treatment intensity ATE estimates from the non-linear Double ML model also closely align with those reported by the linear double ML model:

Table 14: Treatment Effects by Neighborhood Intensity: Causal Forest Double ML Estimates

Treatment	Treatment Intensity		
	Low	Mid	High
Civic Duty	0.0147	0.0210	0.0194
Hawthorne	0.0256	0.0260	0.0241
Self	0.0435	0.0502	0.0507
Neighbors	0.0748	0.0892	0.0789

The results from both the non-linear Double ML models indicate that the treatment effects on voter turnout are consistent across different modeling approaches. The estimates closely match those reported in the original study, suggesting robustness in the findings. The treatment intensity is also robust and aligns with the findings from the meta learners and the Doubly Robust estimator. This consistency across models strengthens the validity of the conclusions drawn regarding the impact of social pressure treatments on voter turnout.

A cumulative gain plot showing heterogeneity in treatment effects is presented in the appendix. The plot shows that the Civic Duty and Hawthorne treatments exhibit most heterogeneity, as indicated by the steeper cumulative gain curves. This implies that these treatments have varying effects on different subpopulations, potentially influenced by factors such as demographics or prior voting behavior.

4.7.7 Results Discussion

The causal machine learning methods yield a few principal findings that extend the original analysis of Gerber et al. (2008).

Across all estimation approaches, the direct treatment effects remain remarkably stable. This consistency across methods provides strong evidence that the treatment effect gradient reflects genuine causal effects rather than statistical artifacts.

Contrary to our initial hypothesis, we find no evidence of positive spillover effects from treated neighbors to untreated individuals. The coefficient on neighborhood treatment intensity is consistently negative across OLS (-0.0167), spatial autoregression (-0.0157), and LASSO (-0.0183), while IPW produces a null effect (0.0019, not significant).

Interpreting these findings through our structural framework reveals why the linear OLS and SAR models failed to capture λ reliably. The structural model assumed a constant transmission rate λ and a constant infection probability η . However, the machine learning results discovered a clear violation of this assumption, the inverted U relationship that indicated that η is not constant but a function of treatment density ($\eta(K)$).

The inverted-U relationship between neighborhood treatment intensity and treatment effects is the most striking finding. The Self treatment is the exception, showing monotonically increasing effects from 4.4 to 5.1 percentage points as intensity rises. This makes theoretical sense: the Self

treatment operates through private shame rather than public social pressure.

The inverted-U pattern for public pressure treatments suggests complex social dynamics. At low intensity, social pressure lacks any effect. At moderate intensity, treated neighbors create visible norms without overwhelming individual responsibility. At high intensity, the treatment becomes ubiquitous, potentially triggering the free-rider problem where individuals feel less compelled to vote themselves.

4.7.8 Limitations

The experiment does have several limitations that we would like to acknowledge here. Our measure of neighbors relies on geographic proximity via ZIP+4 codes, which may poorly capture actual social networks since individuals often interact more with coworkers, family, or online contacts than physical neighbors. The 11-day window between mailing and election may be insufficient for information to diffuse through neighborhood interactions. We cannot directly observe the infection process—whether neighbors discussed the mailings or observed each other’s voting behavior. Finally, the negative intensity coefficient, while consistent across methods, may partially reflect unobserved heterogeneity in neighborhood characteristics correlated with both sampling density and baseline civic engagement.

5 Conclusion

This paper investigates not just whether social pressure increases voter turnout, but how it diffuses through neighborhoods. Using a variety of econometric and machine learning methods, we analyze data from a large-scale field experiment on social pressure and voting behavior.

Our analysis yields three main findings that directly map to the structural parameters of our theoretical model. First, we find robust estimates for the direct infection probability consistent with the results presented by Gerber et al. (2008). Across all specifications, the direct mailing treatments significantly increase the probability of voting, with the Neighbors treatment generating the highest effect (or direct infection rate $\eta_{Neighbors} \approx 8\%$).

Second, regarding the transmission rate λ , we reject our initial hypothesis of a universally positive diffusion effect. For the Civic Duty, Self, and Neighbors treatments, the structural parameter is statistically indistinguishable from zero, indicating no meaningful spillover to untreated neighbors. This suggests that for most forms of social pressure, the infection is contained within the household and does not diffuse to untreated neighbors.

Third, we identify a crucial exception that validates the structural importance of social context. The Hawthorne treatment is almost entirely driven by the interaction term, implying a high positive indirect effect on the neighbors. Conversely, for the neighbors treatment, there is an inverted u shaped pattern discovered by machine learning estimates. The effect is higher at moderate intensities but lowers at low and high intensities. This highlights that social pressure acts as a contagion only up to a saturation point, beyond which free riding takes over.

The deployment of Causal Machine Learning methods (specifically X-Learners and Causal Forests) was instrumental in revealing the hidden non-linear structure of the diffusion process that traditional linear models could not capture. At low levels of neighborhodod treatment intensity, social pressure fails to establish strong norms, moderate intensity is revealed as optimal where community creates positive reinforcement without being overbearing. High intensity neighborhoods suggested a social saturation effect and a decline in the effect of the pressure treatment.

These findings have important implications for understanding how social pressure campaigns function. They suggest that while direct appeals can effectively mobilize voters, simply saturating a neighborhood with treatment may not yield additional benefits and could even backfire due to free-riding dynamics. Future campaigns should consider targeting moderate intensity levels to optimize turnout effects, creating clusters of social pressure large enough to spark conversation but small enough to avoid triggering community fatigue.

6 Future Steps

We can explore other models and supplementary data to further validate our findings. Analysis of similar field experiments with clearly identifiable social networks can help corroborate our results. Additional extensions can include exploring social media data (such as X (formerly Twitter)) to see how information diffusion effects spread in the modern age where spatial proximity is less relevant. Chetty et al. (2022) created a new measure of economic connectedness, social cohesion and civic engagement using Facebook data. The data is from 2022, however, and may not be directly comparable to the 2006 data used in this study. But it does provide a useful framework for future research on social pressure and information diffusion effects in the digital age.

Appendix 1: OLS Block Clustered Results

Table 15: OLS Regression Results with Block Clustered Standard Errors

	Baseline	Diffusion	Interaction
Constant	7.9765*** (0.1623)	7.9890*** (0.1637)	7.9924*** (0.1643)
Civic Duty	0.0187*** (0.0029)	0.0187*** (0.0029)	0.0162*** (0.0049)
Hawthorne	0.0257*** (0.0033)	0.0258*** (0.0033)	-0.0030 (0.0092)
Neighbors	0.0802*** (0.0028)	0.0803*** (0.0028)	0.0677*** (0.0050)
Self	0.0483*** (0.0032)	0.0484*** (0.0032)	0.0581*** (0.0050)
Female	-0.0080*** (0.0012)	-0.0080*** (0.0012)	-0.0080*** (0.0012)
Year of Birth	-0.0040*** (0.0001)	-0.0040*** (0.0001)	-0.0040*** (0.0001)
Voted Primary 2004	0.1482*** (0.0029)	0.1482*** (0.0029)	0.1482*** (0.0029)
Treatment Intensity		-0.0167*** (0.0052)	-0.0256*** (0.0060)
Civic Duty_x_intensity			0.0061 (0.0130)
Hawthorne_x_intensity			0.0677*** (0.0217)
Neighbors_x_intensity			0.0296** (0.0128)
Self_x_intensity			-0.0227* (0.0136)
R-squared	0.0450	0.0450	0.0451
R-squared Adj.	0.0450	0.0450	0.0450
N	344,084	344,084	344,084

Standard errors in parentheses.

* p<.1, ** p<.05, ***p<.01

Appendix 2: Positivity Check Plot

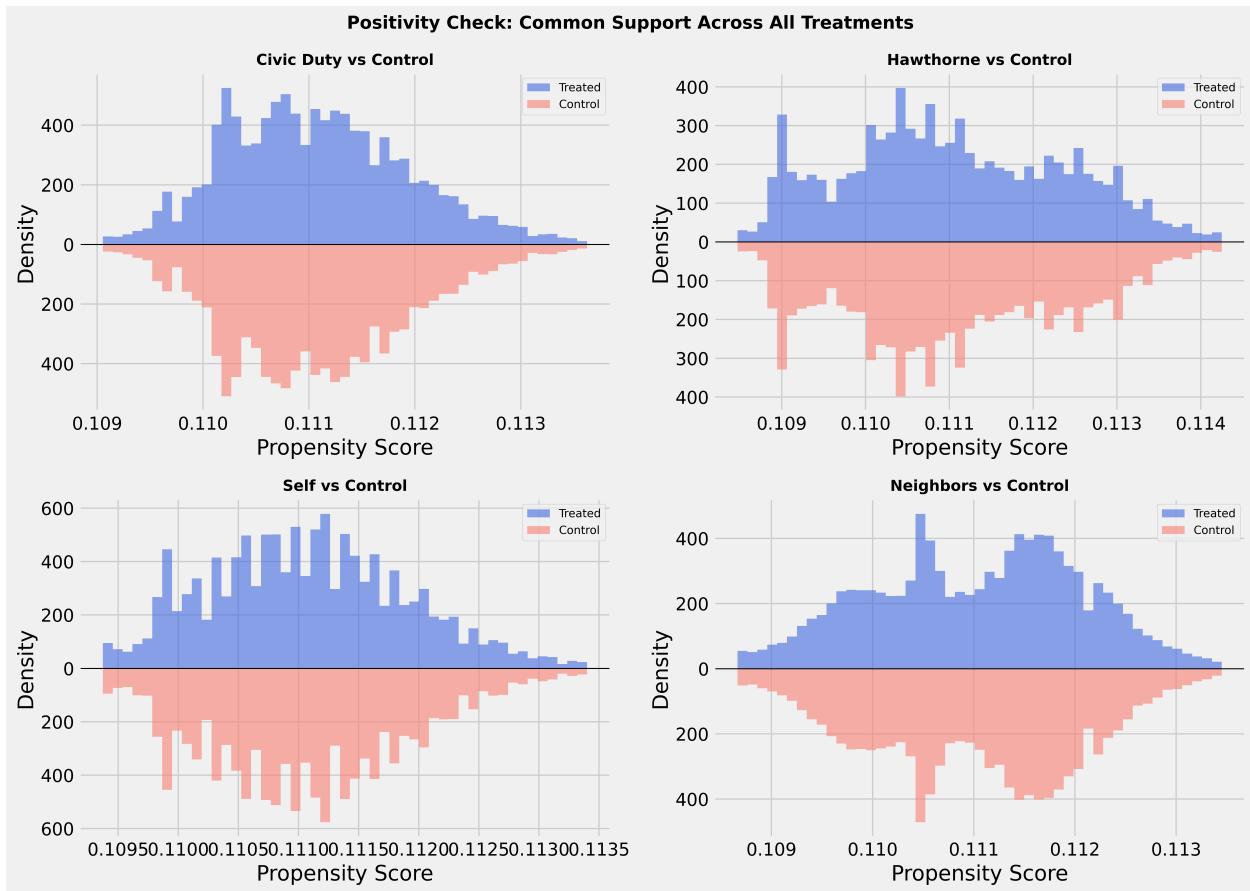


Figure 6: Positivity Check Plot

Appendix 3: Bootstrapped ATEs

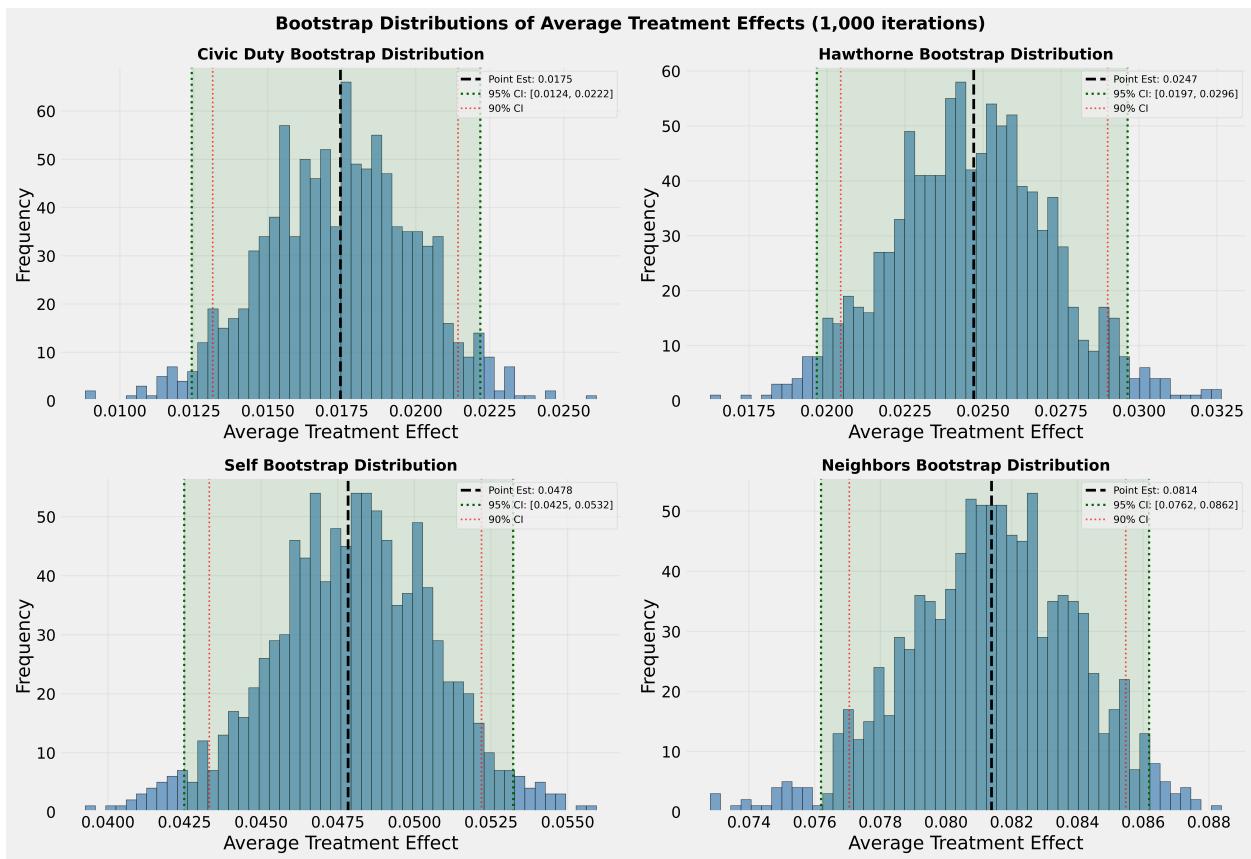


Figure 7: Bootstrapped ATEs for Treatment Effects

Appendix 4: Cumulative Gain Plots

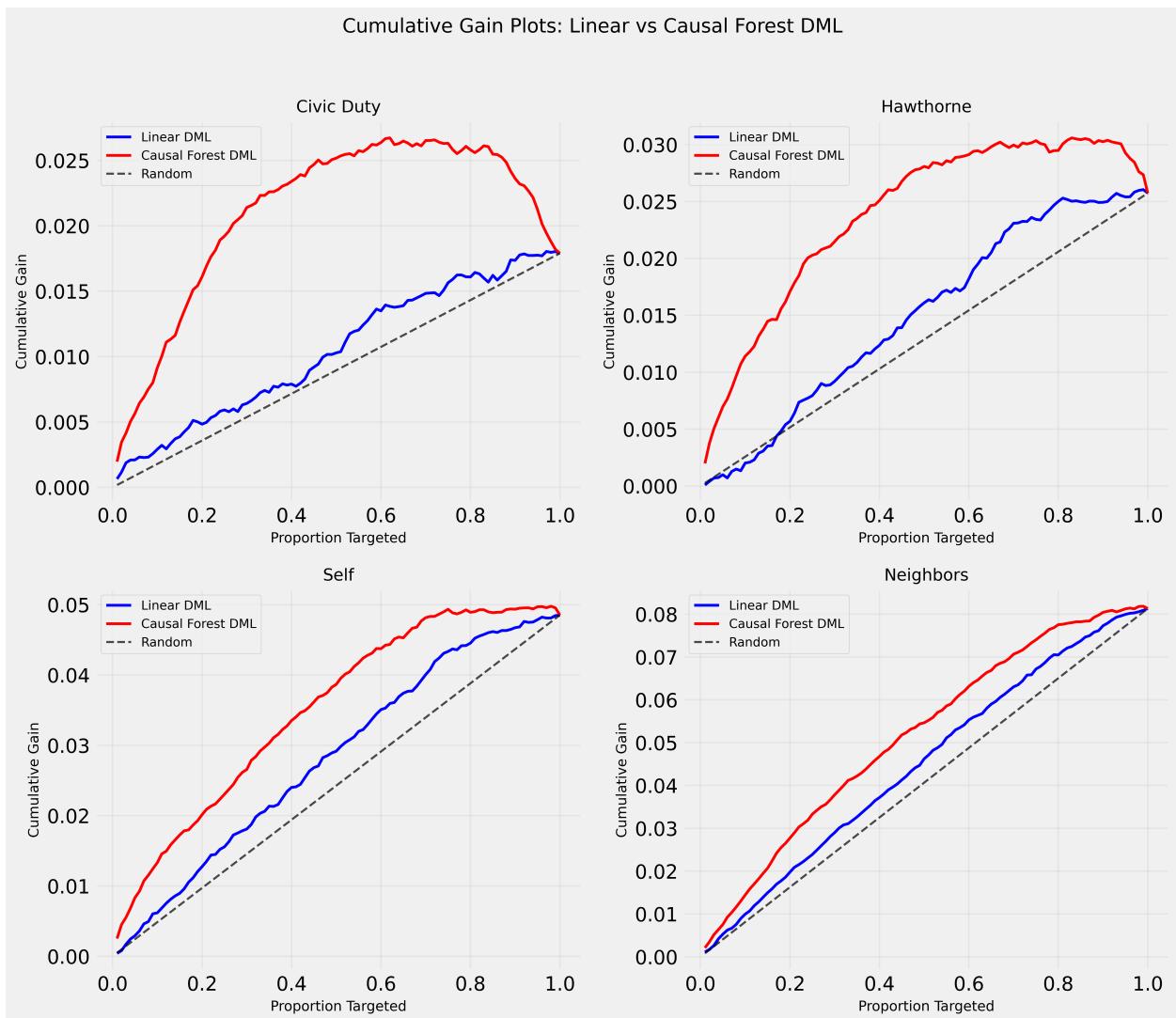


Figure 8: Cumulative Gain Plot for Treatment Effects

References

- Atienza-Barthelemy, J., J. C. Losada, and R. M. Benito (2025, January). Modeling Information Diffusion on Social Media: The Role of the Saturation Effect. *Mathematics* 13(6), 963. Publisher: Multidisciplinary Digital Publishing Institute.
- Bond, R. M., C. J. Fariss, J. J. Jones, A. D. I. Kramer, C. Marlow, J. E. Settle, and J. H. Fowler (2012, September). A 61-million-person experiment in social influence and political mobilization. *Nature* 489(7415), 295–298. Publisher: Nature Publishing Group.
- Chetty, R., M. O. Jackson, T. Kuchler, J. Stroebel, N. Hendren, R. B. Fluegge, S. Gong, F. Gonzalez, A. Grondin, M. Jacob, D. Johnston, M. Koenen, E. Laguna-Muggenburg, F. Mudekereza, T. Rutter, N. Thor, W. Townsend, R. Zhang, M. Bailey, P. Barberá, M. Bhole, and N. Wernherfelt (2022, August). Social capital I: measurement and associations with economic mobility. *Nature* 608(7921), 108–121. Publisher: Nature Publishing Group.
- Gerber, A. S., D. P. Green, and C. W. Larimer (2008). Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment. *The American Political Science Review* 102(1), 33–48. Publisher: [American Political Science Association, Cambridge University Press].
- Haenschen, K. (2016, August). Social Pressure on Social Media: Using Facebook Status Updates to Increase Voter Turnout: Social Pressure on Social Media. *Journal of Communication* 66(4), 542–563.
- Hong, S.-A. (2023). COVID-19 vaccine communication and advocacy strategy: a social marketing campaign for increasing COVID-19 vaccine uptake in South Korea. *Humanities & Social Sciences Communications* 10(1), 109.
- Lord, D., X. Qin, and S. R. Geedipally (2021, January). Chapter 9 - Models for spatial data. In D. Lord, X. Qin, and S. R. Geedipally (Eds.), *Highway Safety Analytics and Modeling*, pp. 299–334. Elsevier.
- Myers, S. A., C. Zhu, and J. Leskovec (2012, August). Information diffusion and external influence in networks. pp. 33–41.
- Solijonov, A. (2016, December). Voter Turnout Trends around the world | International IDEA.
- Vosoughi, S., D. Roy, and S. Aral (2018, March). The spread of true and false news online. *Science* 359(6380), 1146–1151.