Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis Econometrics and Operational Research

Uncovering Treatment Effect Heterogeneity in Facebook COVID-19 Vaccination Campaigns

Emma de Ridder (606504)

Supervisor:	M.F.O. Welz MSc
Second assessor:	Prof.dr. P.H.B.F. Franses
Date final version:	30th June 2024

Abstract

This study examines the heterogeneity in the effects of large-scale social media advertising campaigns on COVID-19 vaccination rates in France. Using the methodological framework of Chernozhukov et al. (2023), which focus on specific features of the conditional average treatment effect (CATE) rather than full estimation, I analyse the impact of two types of targeted messaging on different population groups. The findings reveal significant heterogeneity in both positive and negative directions. Vaccination rates were significantly negatively impacted by both messaging campaigns among individuals living in low-income areas, whereas positive vaccination outcomes were observed in wealthier areas. These insights highlight the importance of tailored health campaigns that address the specific needs to different groups in order to effectively improve COVID-19 vaccination rates, and, more broad, general public health outcomes.

1 Introduction

Randomised Controlled Trials (RCTs) assign subjects randomly to either a control or an experimental group, where they receive a different treatment based on their assignment. The primary advantage of RCTs is their ability to provide unbiased estimates of treatment effects, making them widely utilised in social and economic program evaluations, medical treatments, marketing, psychological treatments and policy interventions. Examples include speech treatment for patients with Parkinson's disease (Ramig et al., 2018), the effect of dance movement therapy on stress management improvement and reduction (Bräuninger, 2012), and drug treatments on chronic insomniac adults (Buscemi et al., 2007). See the works of Glewwe and Kremer (2006), Duflo et al. (2004) and Karlan and Zinman (2009) for more examples of RCTs in a (development) economic context. While RCTs can provide valuable simple average treatment effects, understanding treatment effect heterogeneity—such as its effect on age, gender, income level, socio-economic status, etc.—is a major point of contention for researchers and policymakers and can provide valuable insights. To illustrate, Duflo et al. (2021) find that secondary school scholarships for financially disadvantaged students in rural Ghana lead to a 27 percentage point increase in secondary school completion rates, along with increased enrolment in tertiary education and higher rates of securing a job in the public section, with women benefiting particularly strongly in the latter two areas.

Assessing the heterogeneity of treatment effects is crucial for two reasons: it not only provides a better understanding of the underlying drivers behind an effect of a program, allowing experts to tailor interventions more effectively, but it is also essential for determining whether the program's impact can be generalised to populations that have different characteristics. In the past, studies have assessed this heterogeneity in RCTs by pre-determining the subgroups of covariates (Turner et al. (2012) and Twisk et al. (2018)). However, restricting the analysis to pre-sorting the subgroups can result in an overwhelming number of subgroups to check and the loss of potentially valuable information. Conversely, choosing subgroups ex-post exposes the study to the risk of overfitting.

Recently, the use of machine learning (ML) tools has been proposed to resolve these issues and estimate the conditional average treatment effects (CATE). To illustrate, Athey et al. (2019) propose generalised random forests, Athey and Imbens (2015) introduce five tree-based algorithms for estimating the CATE, and Künzel et al. (2019) introduce the X-learner. In causal inference, the CATE is a key metric that measures how the effects of a treatment vary among individuals with different observed characteristics, thereby revealing the responses of different population groups to the treatment. However, a downside of using ML tools in estimating the CATE is that they struggle to obtain uniformly valid causal effects inferences and, therefore, struggle to consistently estimate the CATE (Chernozhukov et al., 2023). This difficulty particularly arises in high-dimensional settings, where the complexity of ML methods and the large number of data features can weaken statistical significance.

The approach by Chernozhukov et al. (2023) diverges fundamentally from directly estimating the CATE with high precision across all scenarios, often leading to biased or inconsistent results due to the inherent challenges. Instead, inspired by the work of Genovese and Wasserman (2008), they propose focusing on the estimation and inference on *features* of the CATE.

This methodological shift allows for more reliable and valid estimates without overcommitting to estimating the CATE with the highest precision, and the researchers hereby overcome the limitation of ML tools in non-parametric inference settings. The purpose of their study is to develop a generic approach to using any ML tool to predict and make inference on heterogeneous treatment by targeting three features of the CATE: the Best Linear Predictor (BLP) of the CATE, the Sorted Group Average Treatment Effects (GATES), and the Classification Analysis (CLAN). The researchers account for estimation and splitting uncertainty by taking a multitude of sample splitting and then aggregating the results using quantile-aggregated inference. A goodness-of-fit measure for the BLP and GATES evaluates and selects among the ML methods applied.

This study applies the approach by Chernozhukov et al. (2023) to a large-scale RCT conducted in France, studying the impact of Facebook advertising campaigns on COVID-19 vaccination.[1] Despite the widespread availability of COVID-19 vaccination, 78% of the population in France was fully vaccinated (two doses), and hospital admissions continued to increase as of December 2022, according to the Santé Publique France (2022). In addition to that, France's regions show large disparities, highlighting the need for effective strategies to encourage uptake. In an effort to boost COVID-19 vaccination, Ho et al. (2023) collaborated with healthcare professionals to post short video messages that addressed common doubts and misconceptions around COVID-19 vaccination. Leveraging Facebook's advertising tools, the campaign randomly targeted users in areas with vaccination rates below 80 percent of France's national average. Two outreach strategies were then employed: "direct messages" were posted to a large audience of Facebook users, while "friends messages" encouraged users to share information about the vaccine with friends. The campaign reached around 11.5 million distinct French Facebook users in total and achieved high engagement metrics, yet the results indicated no statistically significant impact on vaccination rates across any of the strategies employed. The observed null effect suggests that, although the content was highly engaging, users were not sufficiently motivated to get vaccinated. Considering these findings, I seek to investigate the data more thoroughly using ML techniques to uncover obscured heterogeneity in treatment effects that may be overlooked by the aggregate analysis in Ho et al. (2023). It is plausible that different population subgroups might respond differently to the same intervention due to factors such as demographic characteristics, socioeconomic status, and pre-existing attitudes towards vaccination. Understanding which subgroups are more receptive or resistant to vaccination can help develop more targeted interventions to promote COVID-19 vaccination more effectively. This can ultimately improve vaccination rates and overall public health outcomes, as well as provide valuable insights for future health campaigns. My findings reveal that the 20% most affected areas saw a reduction of 72.1 first-dose vaccinations per week and a reduction of 55.5 completed vaccinations per week for the direct messages. For the friends messages, these numbers are 78.1 and 59.0, respectively. Further analysis reveals that these areas are characterised by lower-income individuals. The study highlights the counterproductive effect the treatments have had on promoting COVID-19 vaccination in this population group.

The remainder of this paper is structured as follows. Section 2 covers literature on the integ-

[1] Chernozhukov et al. (2023) apply their approach to a study that assesses the heterogeneity in the effect of microcredit availability. My replication of this study can be found in Appendix F.

ration of ML and RCTs, sample splitting, and previous literature on heterogeneity in COVID-19 vaccination rates. This will be followed by Section 3, which describes the creation of the ML proxy predictor, the construction of the key features of the CATE and quantile aggregation. Thereafter, Section 4 introduces the empirical application, where I use four ML tools to obtain and analyse the BLP, GATES and CLAN and identify which subgroups of areas in France are affected most by the two employed treatments. Finally, Section 5 summarises the paper.

2 Literature Review

The intersection of ML and RCTs marks a significant advancement in data analysis, providing unbiased treatment effects while managing high-dimensional data and uncovering complex patterns. Besides, the intersection allows for identifying heterogeneous effects without requiring strict assumptions. For examples, see Imai and Ratkovic (2013) and Athey and Imbens (2016). In contrast, less complex non-ML tools address heterogeneity by imposing strong assumptions about sparsity and function forms of models (examples can be found in Chernozhukov et al. (2018), Hansen et al. (2018), Belloni et al. (2015) and Dezeure et al. (2017)). ML-based or not, these methods focus on the full estimation of the CATE. The approach by Chernozhukov et al. (2023) is different, addressing the challenge of non-parametric inference by focusing on the *features* of the CATE rather than the function itself. In doing so, the researchers avoid making unrealistic and hard-to-check assumptions. Their research underscores the importance of uncovering heterogeneous effects, as failing to do so can be harmful, particularly in medical treatments (Kravitz et al., 2004).

This study applies the approach of Chernozhukov et al. (2023) to an RCT that evaluates the heterogeneous effects of large-scale social media Facebook advertising campaigns on COVID-19 vaccination rates increase. This RCT, previously studied by Ho et al. (2023), posted recorded video messages on Facebook addressing misconceptions about COVID-19 vaccination in areas in France and US where the vaccination rates were relatively low compared to the national average. Previous studies have shown proof that COVID-19 vaccination rates exhibit heterogeneity across different populations and regions. For instance, Sun and Monnat (2022) find that rural counties in the US, particularly those depending on farming and mining, have significantly lower vaccination rates compared to urban areas. Factors such as education, health care infrastructure, and Trump vote share contribute to this, the latter of which is also supported by Albrecht (2022), who finds that counties dominated by Republic votes have significantly lower vaccination rates. Brown et al. (2021) use the COVID-19 Community Vulnerability Index to identify commonalities between county vulnerability and vaccination rates, concluding that counties where people generally struggle with challenges relating to covering housing, transportation, household arrangements, and disability have significantly lower vaccination rates. Furthermore, demographic factors such as gender, age, race, and education influence the willingness to receive COVID-19 vaccinations. Studies by Malik et al. (2020) and Kreps et al. (2020) find that particularly men and individuals with college degrees are among the most accepting of the COVID-19 vaccine, whereas black Americans show significantly low uptake. These disadvantaged subpopulations face greater risks of infections, have lower access to health care, and, as a result, experience higher COVID-19 mortality rates (Mackey et al. (2021) and Saunders et al. (2021)). Strategic approaches are needed to overcome these barriers to vaccination and ensure that vulnerable populations are vaccinated.

Numerous studies have explored the efficacy of treatment programs to promote vaccination. For example, Dai et al. (2021) used text-based reminders that highlight the importance of getting vaccinated and provide information to simplify the process. They conclude that this boosted vaccination appointments and rates, although the effect decreased with a subsequent reminder. This can be explained by the study of Rabb et al. (2022), finding that text messages do not significantly increase COVID-19 vaccination rates for those individuals who have not been vaccinated five to eight weeks after becoming eligible for vaccination. It suggests that such a vaccination nudge may be effective early in the vaccination campaign and less so for more hesitant individuals. The study by Berliner Senderey et al. (2021) finds that reminders emphasising the personal benefit of vaccination over social benefit are more effective in increasing vaccination rates. This is supported by Batteux et al. (2022), who conduct a systematic review of studies relating to interventions to increase COVID-19 vaccination, finding that personalising communications and sending booking reminders via text message increases vaccine uptake. In Sweden, Campos-Mercade et al. (2021) finds that a money reward of 200 Swedish kronor increased vaccination rates by 4.2 percentage points. The paper by Ho et al. (2023) attempts another approach: utilising Facebook's advertising features. My study extends Ho et al. (2023) by examining whether different subgroups of the population in these French areas react differently to these interventions.

To ensure robustness of the results, Chernozhukov et al. (2023) take inspiration from Davis and Heller (2020) by using a "hold-out" sample and generate ML predictions via multiple sample splitting, similar to the approach of Abadie et al. (2018). Multiple sample splitting mitigates the issues of single sample splitting—including loss of statistical power and efficiency—but produces different estimates, p-values, and confidence intervals for each split (Romano and DiCiccio, 2019). Several studies have developed procedures to achieve stable and accurate inference. For instance, Ritzwoller and Romano (2023) propose a method that aggregates these metrics using a running average, balancing the trade-off between quickly reducing residual randomness and achieving no significant improvement in the error rate. Another aggregation method is proposed by Meinshausen et al. (2009), who aggregate p-values across multiple random splits using quantiles. Chernozhukov et al. (2023) build on these ideas, using median quantile aggregation with adjustments for splitting uncertainty to provide reliable estimates.

3 Methodology

Using the terminology from Chernozhukov et al. (2023), this section presents their methodology for estimation and inference of the key features of the CATE. Accordingly, I explain their approach to constructing ML proxy predictors and using these to develop the BLP, GATES and CLAN features of the CATE. Furthermore, I present their measure for choosing and selecting among ML tools and discuss features of quantile aggregation. The structure of this section is based on the algorithm provided in Chernozhukov et al. (2023), which is also detailed in Appendix A.

3.1 Model and Key Causal Functions

Let Data $:= (Y_i, Z_i, D_i)_{i=1}^N$ be the observed data consisting of N i.i.d. observations of the random vector (Y, Z, D) drawn from a probability distribution P. Here, Y is the outcome of interest, Z is a multidimensional vector of covariates that characterise the observed units studied in the dataset, and D is a binary treatment indicator. Chernozhukov et al. (2023) assume unconfoundedness of the covariates and random treatment assignment of each unit conditional on Z. The regression model for the observed outcome Y is given by:

$$Y := b_0(Z) + Ds_0(Z) + U, \qquad \mathbb{E}[U \mid Z, D] = 0 \tag{3.1}$$

where $b_0(Z) := \mathbb{E}[Y \mid D = 0, Z]$ represents the baseline conditional average (BCA) and $s_0(Z) := \mathbb{E}[Y \mid D = 1, Z] - \mathbb{E}[Y \mid D = 0, Z]$ defines the conditional average treatment effect (CATE). The expectation operator is denoted by $\mathbb{E}$.

Essentially, Y serves as a foundation for estimating causal effects. The model is designed to systematically quantify the effects of a treatment across different subgroups, based on their unique covariate profile Z. By doing so, Y not only measures the overall treatment effectiveness, but can also identify the subgroups that are most significantly impacted by it. Predictive ML methods leverage the regression model by learning $\mathbb{E}[Y \mid D, Z]$, and then estimating the CATE, $s_0(Z)$.

3.2 Sample Splitting

In order to get robust estimation and inference results, and avoid overfitting, Chernozhukov et al. (2023) use repeated random data splitting. Let (A, M) define a random partition of the set of indices $\{1, \ldots, N\}$. That is, Data $= (Y_i, D_i, Z_i)_{i=1}^N$ is split in two disjoint sets $\text{Data}_A = (Y_i, D_i, Z_i)_{i \in A}$ and $\text{Data}_M = (Y_i, D_i, Z_i)_{i \in M}$, which denote an auxiliary sample (A) and a main sample (M), respectively. On the auxiliary sample, A, Chernozhukov et al. (2023) use some machine learner to obtain ML estimators $z \mapsto B(z) = B_A(z)$ and $z \mapsto S(z) = S_A(z)$ of $b_0(z)$ and $s_0(z)$, respectively, which they refer to as the ML proxy predictors. Note that both $S(Z)$ and $B(Z)$ are potentially biased and noisy estimators due to estimation uncertainty from conditioning on subsample A. Hence, these proxy predictors are not required to be consistent estimators. In the main sample, M, the proxies are then used to estimate and make inference on three key features of the CATE function $z \mapsto s_0(z)$. These key features inculde the Best Linear Predictor (BLP) of the CATE, the Sorted Group Average Treatment Effects (GATES) and the Classification Analysis (CLAN). Systematically repeating this process addresses not only estimation uncertainty associated with the auxiliary sample, but also uncertainty that is induced by the random partitioning of the data.

3.3 Best Linear Predictor

Assuming the Best Linear Predictor (BLP) of the CATE $s_0(Z)$ exists, it is defined as

$$\mathsf{BLP}[s_0(Z) \mid S(Z)] := \beta_1 + \beta_2(S(Z) - \mathbb{E}[S(Z)]). \tag{3.2}$$

Here, $\beta_1 = \mathbb{E}[s_0(Z)]$ and $\beta_2 = \text{Cov}[s_0(Z), S(Z)]/\text{Var}[S(Z)]$. In the context of treatment effects, specifically for estimating the CATE, the BLP is a statistical method that finds the best linear approximation of the true CATE, while aiming to minimise the prediction error through $S(Z)$. $\mathsf{BLP}[s_0(Z) \mid S(Z)]$ is an unbiased predictor of $s_0(Z)$. See Chernozhukov et al. (2023) for details.

In general, if one can statistically reject $\beta_2 = 0$, we conclude that $s_0(Z)$ in heterogeneous (i.e. $s_0(Z)$ not constant across observations) and $S(Z)$ is a statically significant predictor of $s_0(Z)$. This implies that $S(Z)$ provides meaningful information about the CATE, despite potential errors. Furthermore, for $\beta_2 = 1$, $S(Z)$ perfectly predicts $s_0(Z)$ and does not contain any discrepancies. However, typically, $\beta_2 \neq 1$, correcting for the noise in $S(Z)$. The coefficients β_1 and β_2 in $\mathsf{BLP}[s_0(Z) \mid S(Z)]$ can be identified through two strategies. The first estimation strategy employs a weighted linear projection to enhance the predictive accuracy of the BLP, with its empirical equivalent being:

$$Y_i = \widehat{\alpha}_0' X_{1i} + \widehat{\alpha}_1(D_i - p(Z_i)) + \widehat{\alpha}_2(D_i - p(Z_i))(S_i - \mathbb{E}_{N,M}[S_i]) + \widehat{\epsilon}_i, \quad i \in M \tag{3.3}$$

with $\mathbb{E}_{N,M}[w(Z_i)\widehat{\epsilon}_i X_i] = 0$ and where $X_i = [X_{1i}', X_{2i}']'$, $X_{1i} := [1, B(Z_i), p(Z_i), p(Z_i)S(Z_i)]'$, $X_{2i} := [D_i - p(Z_i), (D_i - p(Z_i))(S_i - \mathbb{E}_{N,M}[S_i])]$. An alternative strategy can be defined by using the Horvitz-Thompson transform H, which is detailed in Appendix B along with detailed algorithms that implement both estimation strategies.

3.4 Sorted Group Average Treatment Effects

The Sorted Group Average Treatment Effects (GATES) measures the average of $s_0(Z)$ across different heterogeneity groups G_k, for $k = 1, \ldots, K$, based on the values of the ML proxy predictor $S(Z)$. That is,

$$\gamma_k = \mathbb{E}[s_0(Z) \mid G_k], \quad k = 1, \ldots, K. \tag{3.4}$$

where $G_k = \{S \in I_k\}$, with non-overlapping intervals $I_k = [\ell_{k-1}, \ell_k)$ and $-\infty = \ell_0 < \ell_1 < ... < \ell_K = +\infty$, creating partitions based on the range of the predicted treatment effect S into K equally-sized distinct groups. Following this, Chernozhukov et al. (2023) define the following estimation strategy:

$$Y_i = \widehat{\alpha}_0' X_{1i} + \widehat{\alpha}' W_{2i} + \widehat{\nu}_i, \quad i \in M, \quad \mathbb{E}_{N,M}[w(Z_i)\widehat{\nu}_i W_i] = 0 \tag{3.5}$$

where $W_i = [X_{1i}', W_{2i}']'$, $X_{1i} := (B(Z_i), p(Z_i)\{1(G_k)_{k=1}^K\})'$ and $W_{2i} := ([D_i - p(Z_i)] \cdot \{1(G_k)\}_{k=1}^K)'$. The estimation strategy using the Horvitz-Thompson transform is detailed in Appendix C. Also refer to Appendix C for details on the algorithms recovering and estimating the GATES parameters $\gamma = (\gamma_k)_{k=1}^K$.

Interestingly, for each group G_k as classified by the GATES, the average treatment effect within that group, γ_k, can be seen as a BLP of $s_0(Z)$ for that specific group. Given the construction of the groups, it is only natural to impose the monotonicity restriction:

$$\mathbb{E}[s_0(Z) \mid G_1] \leq ... \leq \mathbb{E}[s_0(Z) \mid G_K]. \tag{3.6}$$

See Chernozhukov et al. (2017) for further details.

Groups are constructed to maximise the variation in S across different subgroups. This approach effectively captures and clarifies the differences in how various subpopulations respond to a certain treatment. One can apply many alternative methods to create groups based on ML tools applied to auxiliary data, such as an "endogenous stratification" analysis, where grouping is based on predicted baseline response (Abadie et al., 2018). Alternatively, Athey and Imbens (2016) partition the training sample recursively using a causal tree approach.

3.5 Classification Analysis

The Classification Analysis (CLAN) measures the average characteristics of the groups in a subpopulation that are most and least affected by the heterogeneity in a treatment. Under the monotonicity restriction, Chernozhukov et al. (2023) denote the "least affected group" by G_1, whereas G_K defines the "most affected group". Note that the labels "most" and "least" may be switched depending on the context of the analysis.

Denote $g(Y, D, Z)$ a vector of characteristics of an observed unit. The average characteristics of the most and least affected groups are defined by the parameters

$$\delta_1 = \mathbb{E}[g(Y, D, Z) \mid G_1], \quad \text{and} \quad \delta_K = \mathbb{E}[g(Y, D, Z) \mid G_K], \tag{3.7}$$

where binary group indicators G_k, for $k = 1, \ldots, K$, are constructed similarly as for the GATES. The CLAN enables us to quantify the disparity between the most and least affected groups and can be extended to comparing features other than averages, for instance variances or distributions.

3.6 Goodness-of-Fit Measures for Fitting the CATE

To effectively select ML proxies in the main sample, is it beneficial to employ goodness-of-fit measures, given their ability to assess model fit. Guided by the BLP of the CATE, Chernozhukov et al. (2023) propose

$$\Lambda := |\beta_2|^2 \mathrm{Var}(S(Z)) = \mathrm{Corr}(s_0(Z), S(Z))^2 \mathrm{Var}(s_0(Z)). \tag{3.8}$$

This measure essentially quantifies the variability in the actual treatment effect $s_0(Z)$ that is accounted for by the proxy predictor $S(Z)$. Maximising Λ enhances this correlation, indicating a stronger predictive power of $S(Z)$. Similarly, for the GATES analysis, Chernozhukov et al. (2023) maximise

$$\bar{\Lambda} = \mathbb{E}\left(\sum_{k=1}^{K} \gamma_k 1(S \in I_k)\right)^2 = \sum_{k=1}^{K} \gamma_k^2 \, \mathbb{P}(S \in I_k). \tag{3.9}$$

The equation calculates $\bar{\Lambda}$ as the sum of the squared group-specific average treatment effects γ_k^2, weighted by the probability of $S(Z)$ falling into each group interval I_k. For equally-sized groups $G_k = \{S \in I_k\}$, that is $\mathbb{P}(S(Z) \in I_k) = 1/K$ for each $k = 1, \ldots, K$, one chooses the best ML method by maximising

$$\bar{\Lambda} = \frac{1}{K} \sum_{k=1}^{K} \gamma_k^2. \tag{3.10}$$

3.7 Quantile Aggregation

Let θ represent a generic target parameter or functional that this study aims to estimate. This parameter can assume various forms, such as $\theta = \beta_2, \theta = \gamma_k$ or $\theta = \delta_K - \delta_1$, depending on the context of the analysis. I explicitly acknowledge the dependence on the auxiliary sample: $\theta = \theta_A$. Different partitions (A, M) of the set $\{1, ..., N\}$ into sample sizes n and $N - n$ yield different estimands θ_A and estimators $\widehat{\theta}_A$. Consequently, θ_A, $\widehat{\theta}_A$, p-values p_A, confidence intervals $[L_A, U_A]$, and approximate distributions are all random variables. To systematically account for the randomness and therefore uncertainty in these estimators and their distributions, Chernozhukov et al. (2023) employ quantile aggregation, which yields non-random results conditional on the data. More specifically, they focus on using medians for enhanced robustness compared to a single split.

For the median point estimator, Chernozhukov et al. (2023) propose

$$\widehat{\theta} := \mathrm{Med}(\widehat{\theta}_A|\mathrm{Data}). \tag{3.11}$$

The increasing uncertainty from sample splitting is reflected in the confidence interval below, which has a confidence level $1 - 2\alpha$, as opposed to $1 - \alpha$ for a single split:

$$[L, U] := [\mathrm{Med}(L_A|\mathrm{Data}); \mathrm{Med}(U_A|\mathrm{Data})]. \tag{3.12}$$

Testing the null hypothesis with adjusted p-values has a significance level α if

$$\mathbb{P}(p_A \leq \alpha/2|\mathrm{Data}) \geq 1/2 \quad \text{or} \quad p_{0.5} = \mathrm{Med}(p_A|\mathrm{Data}) \leq \alpha/2. \tag{3.13}$$

This implies that in at least 50 percent of the random data splits, the realised p-value p_A is below $\alpha/2$. Consequently, Chernozhukov et al. (2017) define $p = 2p_{0.5}$ as the sample splitting adjusted p-value, where small values provide evidence against the null hypothesis. One-sided alternative hypotheses can now be tested with median p-value $p^{\pm} = M(p_A^{\pm}|Data)$, whereas the two-sided median p-value is $\bar{p} = 2\min(p^+, p^-)$, effectively doubling the smaller one of the two. Utilising median p-values is particularly beneficial for inference, as it minimises the impact of outliers and extreme splits in the data, resulting in more consistent results across many data splits.

4 Results

By applying the approach of Chernozhukov et al. (2023) to an RCT, one can effectively study heterogeneity without pre-determining sources that cause the variation. The methodology is particularly effective in high-dimensional settings, and CLAN estimation provides valuable insights for analysts, experts and researchers. In this empirical application, I study an RCT that evaluates the impact of a large-scale Facebook advertising campaign on COVID-19 vaccinations in France.

4.1 The Impact of Large-Scale Social Media Advertising Campaigns of COVID-19 Vaccinations

The COVID-19 pandemic caused major disruptions to economies, societies and global health. Vaccinations were quickly developed and made available to the public, yet vaccination rates varied widely across countries and regions. This prompted researchers to identify effective strategies to increase COVID-19 vaccination rates, with social media being of particular interest due to its rapid and widespread reach. This is illustrated by de Vere Hunt et al. (2021), for instance, who showed that their messages addressing questions and misconceptions about the COVID-19 vaccine achieved high engagement rates on Facebook, reaching individuals an average of about 5.5 times within a month. The effectiveness of such vaccination promotion campaigns is demonstrated by studies such as Athey et al. (2023) and Evans et al. (2023), whose social media-based campaigns successfully changed individuals' beliefs about vaccination and increased vaccination rates. Building on these findings, this study investigates the effectiveness of video messages from physicians and nurses in France during the height of the Omicron wave in the winter of 2021-2022. These approximately 30-second videos addressed common doubts and misconceptions about COVID-19 vaccination. Using Facebook's advertising tools, the messages were posted in randomly selected areas with vaccination rates below 80 percent of the national average. Two types of messages were employed in the Facebook campaign. "Direct" messages were directly posted to a broad audience of Facebook users. In contrast, "friends" messages were designed to be more personal and involved health experts encouraging viewers to share vaccination information with their friends, along with a link to additional resources and videos on the study's website. This type of message is inspired by studies on alcohol consumption (Collins and Marlatt, 1981) and AIDS prevention (Fisher and Misovich, 1990), which demonstrate the significant influence of social networks on individuals' decision-making. The dataset is taken from Ho et al. (2023), whose study determined that the messages had no significant effects on France's COVID-19 vaccination rates in these areas.

By extending the research of Ho et al. (2023), my aim is to discover the heterogeneous effects of the two types of messages. This analysis is relevant as it uncovers the specific socio-economic factors that influence the effectiveness of the two vaccination promotion messages. Health authorities can then use these insights to better tailor future health campaigns, leading to enhanced public health outcomes in the targeted regions. The dataset includes information from public intermunicipal cooperation establishments (EPCI) and postal codes that correspond to the municipalities within Lyon, Marseille and Paris. Selected areas have vaccination rates that are less than 80 percent of France's national average. The randomisation process divided the areas into three equally sized groups, with each area having a one-third probability of being assigned to the control, direct, or friends group. The two outcome variables in my study are the first-dose COVID-19 vaccinations per week and the number of completed vaccinations (first and second dose) per week. The covariates Z include 43 area-level salary and poverty control variables, such as the average net hourly wage of women in 2020, the number of fiscal households, the share of taxes, and the share of total social benefits. Tables 9 and 10 in Appendix D provide a detailed list of covariates and some corresponding summary statistics, respectively. As a fixed effect, I include the stratification of EPCIs and postal codes based on region, vaccination rates at baseline, and population. Due to a large number of observations, Ho et al. (2023) aggregate the data into 3-by-3-week periods. After removing rows with missing values for the aforementioned variables, the dataset includes $N = 4,810$ observations. Crucial for valid causal inference is undersampling, as it balances the treatment and control groups, ensuring an equal number of observations in each group. Given two treatment types, the undersampled dataset includes a total of $N = 3,242$ observations for both the treatment with direct and friends messages. Both analyses have treatment and control groups, which contain an equal number of observations, namely $1,621$ each, providing a robust dataset for the analysis.

4.2 Direct Messages

My first focus will be on assessing the heterogeneity of posting direct messages on COVID-19 vaccination rates. Table 1 evaluates the performance of four ML methods in predicting the proxy predictor $S(Z_i)$ for the two outcome variables: dose 1 and completed vaccination. It is evident from the table that Random Forest seems to be the most reliable method, especially for predicting the first outcome variable, although Boosting is a close second for both outcomes. Tree-based methods seem to be superior, likely due to their higher flexibility compared to linear models (such as Elastic Net) and lower propensity to overfit compared to Neural Networks. The comparatively lower performances of these models suggest a limited ability to capture the data patterns effectively. Therefore, I will proceed with the analysis of the BLP, GATES and CLAN using Random Forest and Boosting.

Table 1: Comparison of ML Methods: Large-Scale Direct Messaging Campaign

	Random Forest	Elastic Net	Neural Network	Boosting
Dose 1				
Best BLP (Λ)	5,240	200	437	4,008
Best GATES ($\bar{\Lambda}$)	2,123	87	261	1,888
Completed				
Best BLP (Λ)	3,377	140	240	1,954
Best GATES ($\bar{\Lambda}$)	1,134	69	163	1,202

The table compares the performance of four ML methods—Random Forest, Elastic Net, Neural Network, and Boosting—in predicting outcomes from direct Facebook messages promoting COVID-19 vaccination. The two performance metrics are the weekly number of first-dose COVID-19 vaccinations and the weekly number of full vaccinations. Each method's performance is evaluated by using the best BLP (Λ) and best GATES ($\bar{\Lambda}$) values, representing goodness-of-fit measures. The results are based on medians over 100 splits, reflecting robustness of the results across various data partitions, and indicate that Random Forest and Boosting are particularly effective predictors.

Table 2 reports the estimates of the average treatment effect (ATE) parameter, β_1, and heterogeneity loading (HET) parameter, β_2, for the two outcome variables in the BLP. Confidence intervals are median-adjusted, accounting for variability across the sample splits, and are reported in parentheses. The p-values, adjusted similarly, are reported in brackets and test the hypothesis that the respective outcome parameter is zero. The estimated ATEs of the treatment show that both outcome variables report statistically insignificant results at the 10% level for both ML methods. This indicates that, on average, posting direct messages on Facebook in areas of France with generally low COVID-19 vaccination rates does not significantly affect the number of weekly first-dose vaccinations or the number of weekly completed vaccinations. These findings align with Ho et al. (2023), who find insignificant point estimates of 0.013 during the campaign period and -0.038 post-campaign. Despite high engagement on Facebook in France, people did not follow through with vaccination, thus leading to no significant rise in vaccinations. Ho et al. (2023) suggest that people stick to their existing beliefs and opinions about the vaccine. Additionally, in the summer of 2021, the French government had introduced strong incentives to get vaccinated, such as a COVID-19 certificate. Ho et al. (2023) argue that this polarised the population into those who were vaccinated and those who deliberately resisted vaccination, leaving few people undecided about getting vaccinated.

Table 2: BLP of Large-Scale Direct Messaging Campaign

	Random Forest		Boosting	
	ATE (β_1)	HET (β_2)	ATE (β_1)	HET (β_2)
Dose 1	-0.500	-1.239	1.46	-0.730
	(-12.0, 11.3)	(-1.47, -1.04)	(-10.7, 13.5)	(-0.881, -0.581)
	[0.934]	[0.000]	[0.802]	[0.000]
Completed	-3.00	-0.879	-1.61	-0.512
	(-13.8, 8.13)	(-1.09, -0.674)	(-13.33, 9.63)	(-0.672, -0.364)
	[0.600]	[0.000]	[0.778]	[0.000]

The table reports the BLP of the CATE using two performance indicators: the estimates of the average treatment effect (ATE) and heterogeneity loading (HET) parameters. In the parentheses, 90% confidence intervals are reported, and median-adjusted p-values over 100 splits can be found in the brackets for the hypothesis that the respective outcome parameter is zero. Results suggest that, on average, the direct messages do not significantly impact the number of weekly first doses and weekly completed vaccinations at the 10% level. However, the p-values of the HET indicate that there is significant heterogeneity present in the effect on both outcomes for any reasonable statistical level of significance.

Next, considering the heterogeneity results, the hypothesis that HET is zero at any reasonable significance level is rejected for both weekly first-dose and completed vaccinations across both ML methods. The results suggest that the direct messages have varying impacts on individuals' likelihood of receiving the first dose and completing their vaccinations. Possible explanations might lie in socio-demographic differences, geographical or psychological factors. For example, Kelly et al. (2021) conduct a study investing predictors of willingness to get the COVID-19 vaccine, finding that men and older people (age 65 and older) are more likely to get the vaccine. The negative HET values suggest that for certain groups of individuals within the sample, the impact of the direct message campaign is especially ineffective or even counterproductive. This is an important finding, which I will elaborate on later.

Table 3: GATES of 20% Most and Least Affected Groups for Large-Scale Direct Messaging Campaign

	Random Forest			Boosting		
	20% Most (γ_5)	20% Least (γ_1)	Difference ($\gamma_5 - \gamma_1$)	20% Most (γ_5)	20% Least (γ_1)	Difference ($\gamma_5 - \gamma_1$)
Dose 1	-72.1	68.9	-142	-63.7	66.0	-134
	(-100, -44.2)	(40.7, 97.2)	(-183, -103)	(-91.4, -36.8)	(37.9, 92.6)	(-175, -93.1)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Completed	-55.5	48.0	-103	-56.4	45.9	-104
	(-81.7, -29.5)	(20.7, 74.5)	(-142, -65.7)	(-83.3, -30.0)	(20.9, 72.0)	(-140, -68.7)
	[0.000]	[0.001]	[0.000]	[0.000]	[0.000]	[0.000]

The results in the table report the GATES estimates, dividing the areas in the dataset into $K = 5$ groups based on quantile cutoffs of $S(Z_i)$. Random Forest and Boosting are used to estimate the average effects of the three categories on the outcome variables. The 90% confidence intervals are reported in the parentheses, and median-adjusted p-values over 100 splits are shown in the brackets. Statistically significant results at any reasonable level are reported in all groups. While the least affected group displays positive effects on weekly first doses and weekly completed vaccinations, the most affected group responds negatively. One can conclude that the direct messages have varying impacts across different groups.

The GATES parameters provide more insights into these observations. These estimates are provided in Table 3, where individuals are split into five groups based on quantile cutoffs of $S(Z_i)$. The average effect of each group is estimated, showing significant heterogeneity for both the most and least affected group. For the most affected group, the GATES estimates on the number of weekly first doses and weekly completed vaccinations differ significantly from zero at any meaningful significance level for both ML methods. The negative effects on these outcome variables suggest that this group of individuals, on average, is less likely to get the first vaccine or complete their vaccinations. This seems paradoxical, as it implies that the people most affected by the direct messages are even less inclined to get vaccinated after watching the videos. These individuals might have interpreted the message negatively, felt a backlash, or already had very strong beliefs against vaccination. As a result, the direct messages appear to have had the opposite effect on them. Conversely, the group least affected by seeing the direct messages on Facebook responds positively and is more likely to get vaccinated. These individuals might already have a positive attitude towards vaccination. Further investigation into the difference between the most and least affected groups reveals statistically significant differences from zero for the outcome variables at any reasonable level across both ML methods, reporting negative values.

The results from Table 3 suggest that there is significant heterogeneity present in both the negative and positive directions. This variability suggests that the success of the direct messages varies widely across different areas. It not only underscores the importance of considering individual or group-level characteristics in public health campaigns but also the need for more tailored strategies. These results help explain why the overall ATE in Table 2 is insignificant. It appears that there are two distinct groups: those positively affected and those who are affected negatively. Together, they average out to a net zero effect. One can see this from Figure 1, which is a visual representation of the data in Table 3 for the analysis with Random Forest. A similar figure for the Boosting analysis can be found in Figure 3 in Appendix E.

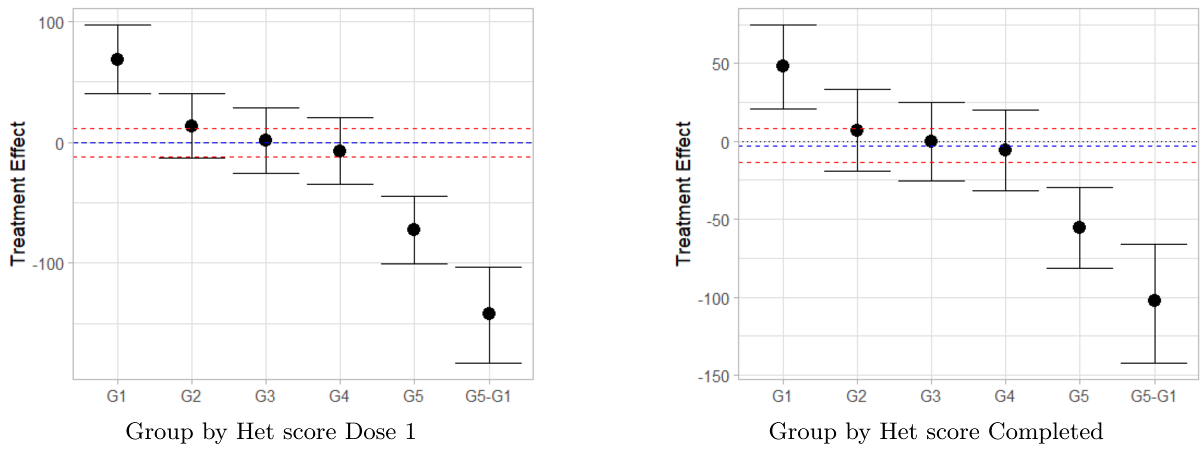


Figure 1: Grouping by Het score based on Random Forest analysis. The black dots represent the GATES estimates with their respective confidence bands. The blue dashed line indicates the ATE and the red dashed lines show the confidence interval. The significant heterogeneity in both opposite directions averages each other out to a net zero effect.

Table 4 displays the average characteristics of the 20% most and least affected groups, based on the same quantiles as Table 3, focusing on four factors: the median standard of living, overall poverty rate, average net hourly wage of people aged 18 to 25 (in 2020), and the average net hourly wage of women (in 2020). The results can help understand the cause of heterogeneity in the treatment effects. My analysis will focus solely on Random Forest results, as Boosting yields statistically insignificant results for all four factors across the outcome variables at any meaningful statistical level. A reasonable explanation for this might be due to the poorer fit observed in Table 1. Table 4 shows that significant results at the 10% level are found only for the median standard of living and the overall poverty rate. For the median standard of living, negative values for both outcome variables suggest that direct messages had a larger negative impact on people who live in areas with lower median incomes. Despite a study by Altman (2021) reporting that doctors and nurses were among the most trusted individuals in providing information about COVID-19 according to US citizens, the messages did not seem to have positively impacted vaccination rates for these groups of individuals. Similar results are observed for the overall poverty rate, suggesting that the most affected groups reside in areas with higher poverty rates compared to the least affected group. One can observe that the likelihood of completing all vaccinations is more negative for these groups compared to receiving the first dose. The results from the table can be attributed to several factors: individuals in poorer communities often have lower trust in public health messages, limited access to healthcare resources, and a generally higher level of vaccine hesitancy. Moreover, they often have a lower understanding of medical and health-related information and greater scepticism towards vaccination promotion campaigns. Conversely, wealthier individuals living in areas where the median living standard is higher respond positively to the direct messages in the advertisements, showing an increase in vaccination rates.

Table 4: CLAN of Large-Scale Direct Messaging Campaign

	Random Forest			Boosting		
	20% Most (δ_5)	20% Least (δ_1)	Difference ($\delta_5 - \delta_1$)	20% Most (δ_5)	20% Least (δ_1)	Difference ($\delta_5 - \delta_1$)
Dose 1						
Median living standard	21,747	22,264	-515	21,809	21,885	-155
	(21391, 22111)	(21914, 22600)	(-1032, -6.00)	(21421, 22149)	(21591, 22183)	(-628, 323)
	[0.000]	[0.000]	[0.047]	[0.000]	[0.000]	[0.555]
Poverty rate	50.7	52.4	-1.62	50.6	51.0	-0.510
	(49.7, 51.7)	(51.4, 53.3)	(-3.05, -0.220)	(49.5, 51.6)	(50.0, 52.1)	(-1.91, 0.890)
	[0.000]	[0.000]	[0.024]	[0.000]	[0.000]	[0.463]
Hourly wage 18-25	10.7	10.7	0.050	10.7	10.7	0.050
	(10.7, 10.8)	(10.6, 10.8)	(-0.06, 0.140)	(10.6, 10.8)	(10.6, 10.7)	(-0.040, 0.130)
	[0.000]	[0.000]	[0.291]	[0.000]	[0.000]	[0.263]
Hourly wage women	14.2	14.4	-0.190	14.1	14.0	0.060
	(13.9, 14.5)	(14.1, 14.6)	(-0.580, 0.240)	(13.8, 14.4)	(13.8, 14.3)	(-0.310, 0.430)
	[0.000]	[0.000]	[0.415]	[0.000]	[0.000]	[0.736]
Completed						
Median living standard	21,744	22,252	-582	21,757	21,825	-72.0
	(21379, 22065)	(21902, 22607)	(-1104, -59.0)	(21393, 22119)	(21509, 22144)	(-569, 423)
	[0.000]	[0.000]	[0.029]	[0.000]	[0.000]	[0.787]
Poverty rate	50.7	52.7	-1.77	50.6	50.9	-0.420
	(49.7, 51.7)	(51.6, 53.8)	(-3.22, -0.380)	(49.5, 51.6)	(50.0, 51.9)	(-1.89, 1.06)
	[0.000]	[0.000]	[0.013]	[0.000]	[0.000]	[0.579]
Hourly wage 18-2025	10.7	10.7	-0.00	10.7	10.7	0.050
	(10.7, 10.8)	(10.7, 10.8)	(-0.110, 0.100)	(10.6, 10.8)	(10.6, 10.7)	(-0.040, 0.130)
	[0.000]	[0.000]	[0.994]	[0.000]	[0.000]	[0.254]
Hourly wage women	14.2	14.5	-0.330	14.1	14.1	0.020
	(13.8, 14.4)	(14.2, 14.8)	(-0.750, 0.100)	(13.8, 14.4)	(13.8, 14.3)	(-0.350, 0.420)
	[0.000]	[0.000]	[0.149]	[0.000]	[0.000]	[0.798]

The table reports the CLAN estimates of the median standard of living, overall poverty rate, average net hourly wage of individuals between 18 and 25, and the average net hourly wage for women for the 20% most and least affected groups are reported for both outcome variables. Note that the median living standard and wages are reported in euros. In parenthesis, the 90% confidence intervals are shown, and median-adjusted p-values over 100 splits are indicated in the brackets. Significant results are found for the first two factors in the Random Forest analysis and indicate that direct messages had a larger negative impact on poorer individuals, increasing the resistance to get vaccinated. Wealthier individuals, on the other hand, show higher COVID-19 vaccination rates after receiving the direct messages on Facebook.

4.3 Friends Messages

It is also of interest to examine the heterogeneity in the effects of the friends messaging campaign on COVID-19 vaccination rates and determine whether encouragement from health experts to convince friends is a more effective method than receiving less personal video messages addressing misconceptions about the vaccine. Table 5 shows that Random Forest excels in predicting the outcome variables for this treatment. Although Boosting does perform as well as it did in Table 5, I will continue the analysis with both Random Forest and Boosting. The comparative analysis again demonstrates better performance of the tree-based ML models.

Table 5: Comparison of ML Methods: Large-Scale Friends Messaging Campaign

	Random Forest	Elastic Net	Neural Network	Boosting
Dose 1				
Best BLP (Λ)	4,121.6	200.2	296.8	970.8
Best GATES ($\bar{\Lambda}$)	1,986.13	80.05	340.98	1,423.60
Completed				
Best BLP (Λ)	1,603.5	424.3	148.2	365.0
Best GATES ($\bar{\Lambda}$)	1,081.26	69.29	158.08	875.13

The table compares the performance of four ML methods in predicting outcomes from Facebook messages encouraging users to convince friends to promote COVID-19 vaccination. The two performance metrics are the number of weekly first-dose COVID-19 vaccinations and the weekly number of full vaccinations. Each method's performance is evaluated by two goodness-of-fit measures, best BLP and best GATES, with results based on medians over 100 splits. One can observe that tree-based ML methods—Random Forest and Boosting—are particularly effective predictors.

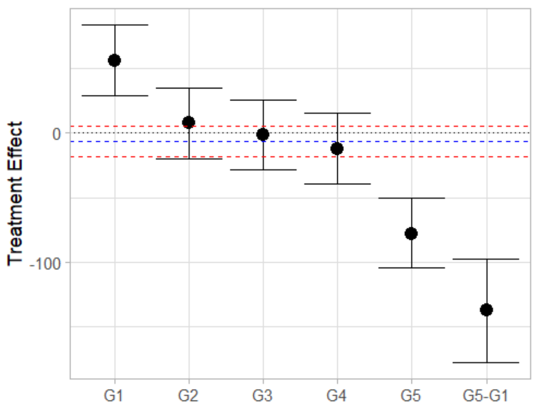
Table 6 reports the estimates of ATE and HET parameters in the BLP for the two outcome variables. Results are similar to those observed in Table 2, showing statistically insignificant results for the ATE estimates at the 10% level across both outcome variables for both ML methods. This indicates that, on average, posting messages on Facebook that encourage users to share links and provide information to promote COVID-19 vaccination does not significantly affect the number of first-dose COVID-19 vaccinations per week or the number of weekly completed vaccinations. Ho et al. (2023) reached the same conclusion, reporting insignificant point estimates of 0.05 and 0.047 during the friends messaging campaign and post-campaign, respectively. This can again be attributed to the fact that people seeing the messages were already convinced not to get vaccinated and, logically, were not likely to encourage others to do so. Considering the heterogeneity results, both outcomes are statistically significant at any reasonable statistical level across the ML methods, suggesting that these friend messages have heterogeneous, yet negative, impacts on the number of weekly first doses and weekly completed vaccinations. It suggests that this kind of treatment had a particularly strong effect on certain individuals in the sample, negatively affecting their likelihood of getting vaccinated.

Table 6: BLP of Large-Scale Friends Messaging Campaign

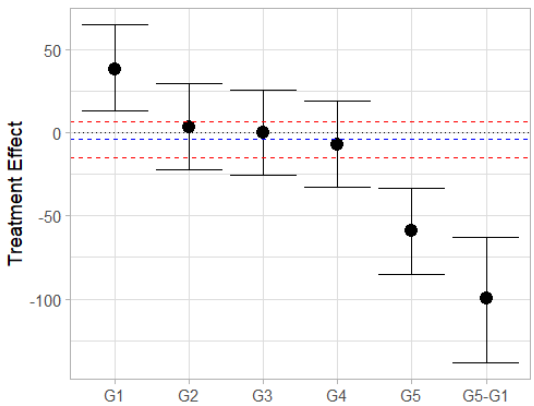
	Random Forest		Boosting	
	ATE (β_1)	HET (β_2)	ATE (β_1)	HET (β_2)
Dose 1	-6.226	-1.070	-7.510	-0.366
	(-17.916, 5.428)	(-1.283, -0.844)	(-19.600, 4.427)	(-0.519, -0.234)
	[0.288]	[0.000]	[0.227]	[0.000]
Completed	-4.266	-0.606	-6.121	-0.138
	(-15.115, 6.493)	(-0.827, -0.397)	(-17.336, 4.880)	(-0.281, 0.016)
	[0.425]	[0.000]	[0.278]	[0.076]

The table reports the BLP of the CATE using the two performance indicators, presenting estimates of the average treatment effect (ATE) and heterogeneity loading (HET) parameters. The 90% confidence intervals are reported in parentheses, and median-adjusted p-values over 100 splits can be found in the brackets. Results suggest that, on average, the friends messages do not significantly impact the weekly number of first doses and weekly completed vaccinations on a 10% level. However, the p-values of the HET indicate that there is significant heterogeneity present in the effect on both outcomes at any meaningful statistical level.

The estimates of the GATES are provided in Table 7, showing significant heterogeneity for γ_1 and γ_5 at any reasonable level of statistical significance. From the results of the table, I can reach the same conclusion as for the direct messages: significant heterogeneity is present in both positive and negative directions. That is, individuals in the most affected group are less likely to get vaccinated, probably due to negatively interpreting the Facebook messages or having pre-existing strong beliefs against COVID-19 vaccination. The least affected group, on the other hand, responds positively, reporting an increase in weekly first-dose and completed vaccinations. The results in Table 7 are visually represented in Figure 2, which again eludes to the average net effect being zero. The same conclusion can be drawn from Figure 4 in Appendix E, which shows the boosting analysis of the GATES parameter from posting the friends messages.



Group by Het score Dose 1



Group by Het score Completed

Figure 2: Grouping by Het score based on Random Forest analysis for the friends messages. The black dot represents the GATES estimate with its respective confidence bands. The blue dashed line indicates the ATE and the red dashed lines show the confidence interval. The most and least affected groups report significant heterogeneity in opposite directions, averaging each other out to a net zero effect. This explains the insignificant ATE results observed in Table 6.

Table 7: GATES of 20% Most and Least Affected Groups For Large-Scale Friends Messaging Campaign

	Random Forest			Boosting		
	20% Most (γ_5)	20% Least (γ_1)	Difference ($\gamma_5 - \gamma_1$)	20% Most (γ_5)	20% Least (γ_1)	Difference ($\gamma_5 - \gamma_1$)
Dose 1	-78.052	56.099	-137.332	-73.347	38.267	-112.268
	(-104.713, -50.355)	(28.939, 83.595)	(-177.485, -97.309)	(-102.038, -44.655)	(10.474, 67.673)	(-151.016, -72.442)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.006]	[0.000]
Completed	-58.950	38.304	-99.983	-56.042	27.190	-83.128
	(-85.287, -33.362)	(12.887, 64.723)	(-138.205, -63.122)	(-81.829, -31.470)	(0.671, 52.232)	(-122.392, -45.602)
	[0.000]	[0.003]	[0.000]	[0.000]	[0.040]	[0.000]

The results in the table report the GATES estimates, dividing the areas into $K = 5$ groups based on quantile cutoffs of $S(Z_i)$. Statistically significant results at any meaningful statistical level are reported in all three groups. Reported in parentheses are the 90% confidence intervals and median-adjusted p-values over 100 splits can be found in the brackets. The least affected group displays positive effects on the weekly number of first doses and completed vaccinations, while the most affected group responds negatively. Similar to the results in Section 4.2, I conclude that the friends messages had varying impacts across the different groups.

So far, the direct and friend messages have reported similar effects on COVID-19 vaccination rates, exhibiting heterogeneity: they negatively impact the most affected group and positively impact the least affected group. The CLAN analysis in Table 8 provides further insights into the characteristics of the individuals most and least affected, thereby highlighting differences between the two treatments. These characteristics include the median standard of living, overall poverty rate, average net hourly wage of people aged 18 to 25 (in 2020), and the average net hourly wage of women (in 2020). From Table 8, different conclusions can be drawn compared to the direct messages. Again, our analysis of the CLAN will focus on Random Forest.

For the first outcome variable, the number of weekly first-dose vaccines, only the factor for the average net hourly wage of women in 2020 reports significant results at the 10% level. It indicates that the messages encouraging users to influence their friends had a larger negative impact in areas where women earn lower average net hourly wages than in areas where women earn higher wages, ultimately discouraging individuals from getting their first dose. It suggests that, in these areas, this treatment was particularly ineffective and even counterproductive in promoting vaccination rates. Regarding the number of weekly completed COVID-19 vaccinations, the statistically significant negative results at the 10% level for the overall poverty rate, the average net hourly wage of people aged 18 to 25, and the average net hourly wage of women suggest that the friends messages posted to Facebook had a larger negative impact on the weekly completed vaccinations for individuals living in areas with these characteristics. Factors such as a lack of trust in the healthcare system, higher susceptibility to disinformation leading to vaccine hesitancy, and prioritising basic daily needs over vaccination due to hardships can account for these results, hindering vaccination efforts.

Table 8: CLAN of Large-Scale Friends Messaging Campaign

	Random Forest			Boosting		
	20% Most (δ_5)	20% Least (δ_1)	Difference ($\delta_5 - \delta_1$)	20% Most (δ_5)	20% Least (δ_1)	Difference ($\delta_5 - \delta_1$)
Dose 1						
Median living Standard	22,246	22,264	-367	21,770	21,884	225
	(21843, 22694)	(21843, 22694)	(-897, 138)	(21467, 22113)	(21544, 22245)	(-272, 718)
	[0.000]	[0.000]	[0.158]	[0.000]	[0.000]	[0.339]
Poverty rate	52.1	52.4	-0.984	50.9	50.6	0.538
	(50.9, 53.2)	(51.3, 53.6)	(-2.54, 0.572)	(49.8, 51.9)	(49.5, 51.8)	(-0.915, 1.98)
	[0.000]	[0.000]	[0.215]	[0.000]	[0.000]	[0.429]
Hourly wage 18-25	10.8	10.8	-0.058	10.7	10.7	0.017
	(10.7, 10.8)	(10.7, 10.9)	(-0.153, 0.031)	(10.6, 10.7)	(10.5, 10.8)	(-0.071, 0.105)
	[0.000]	[0.000]	[0.197]	[0.000]	[0.000]	[0.663]
Hourly wage women	14.6	14.8	-0.525	14.0	14.1	-0.029
	(14.2, 14.9)	(14.4, 15.2)	(-0.930, -0.085)	(13.8, 14.3)	(13.9, 14.4)	(-0.421, 0.366)
	[0.000]	[0.000]	[0.018]	[0.000]	[0.000]	[0.887]
Completed						
Median living Standard	22,207	22,252	-370	21,769	22,104	147
	(21758, 22647)	(21758, 22647)	(-956, 215)	(21452, 22115)	(21740, 22483)	(-354, 638)
	[0.000]	[0.000]	[0.215]	[0.000]	[0.000]	[0.538]
Poverty rate	52.4	51.1	-1.37	51.1	51.1	-0.156
	(51.3, 53.6)	(50.1, 52.1)	(-2.90, 0.160)	(50.1, 52.1)	(49.50, 51.8)	(-1.54, 1.24)
	[0.000]	[0.000]	[0.079]	[0.000]	[0.000]	[0.826]
Hourly wage 18-25	10.8	10.8	-0.107	10.7	10.7	0.000
	(10.7, 10.9)	(10.7, 10.9)	(-0.200, -0.004)	(10.6, 10.7)	(10.6, 10.7)	(-0.095, 0.085)
	[0.000]	[0.000]	[0.044]	[0.000]	[0.000]	[0.981]
Hourly wage women	14.6	14.8	-0.525	14.1	14.1	-0.029
	(14.2, 14.9)	(14.4, 15.2)	(-0.930, -0.085)	(13.8, 14.3)	(13.9, 14.4)	(-0.421, 0.366)
	[0.000]	[0.000]	[0.018]	[0.000]	[0.000]	[0.887]

The table presents the CLAN estimates of four characteristics for the two outcome variables. These characteristics are: median standard of living, overall poverty rate, average net hourly wage of individuals between 18 and 25, and the average net hourly wage for women. These characteristics are reported for the 20% most and least affected groups. Note that the median living standard and wages are reported in euros. Furthermore, 90% confidence intervals are provided in parentheses, and median-adjusted p-values over 100 splits are shown in the brackets. The results from the Random Forest analysis suggest that particularly areas where women earn lower average hourly wages experience a significant negative impact on vaccination rates in the targeted areas.

Overall, the results show that the direct messaging campaign was ineffective and even counterproductive, especially in poorer areas with lower median incomes. This can be attributed to various factors, including deep-rooted distrust in public health institutions, limited access to healthcare resources, and higher susceptibility to misinformation in these communities. Similarly, the friends messaging campaign demonstrated counterproductive effects on vaccination rates in areas where women earn lower average hourly wages. Policymakers can take valuable lessons from both findings. It shows that a one-size-fits-all approach is insufficient, and socio-economic differences need to be considered when designing future public health campaigns to ensure their success.

While this study provides valuable insights, several limitations need to be acknowledged. First, the study relies on Facebook advertising data, which constrains the findings to people who actively use the platform. As a result, the sample may not accurately represent the general population. For instance, older individuals or those without internet access could be underrepresented. Second, the intervention period—the winter of 2021-2022—coincides with government prevention measures such as the introduction of a vaccine pass to access various public places and mandatory face mask-wearing. This makes it harder to attribute effects solely to a specific treatment, as the measures could have independently influenced individuals' vaccination behaviour in the targeted regions. Third, Ho et al. (2023) aggregate the data into 3-by-3 week periods, but this may conceal temporal dynamics and more detailed area-specific variations in response to the interventions. Future research should address these limitations by extending the intervention period or utilising other social media platforms, thereby providing a more enhanced understanding of the impact of large-scale social media campaigns on COVID-19 vaccination rates.

5 Conclusion

In randomised controlled trials (RCTs), full estimation of the conditional average treatment effect (CATE) often requires strong assumptions to obtain consistent estimates. The methodological approach of Chernozhukov et al. (2023) instead focuses on features of the CATE. These include the Best Linear Predictor (BLP) of the CATE, the Sorted Group Average Treatment Effects (GATES) and the Classification Analysis (CLAN). Complemented by machine learning (ML) techniques, this allows for reliable inference on heterogeneous treatment effects. Building on the the framework of Chernozhukov et al. (2023), this study investigates the heterogeneity in the effects of large-scale Facebook advertising campaigns on COVID-19 vaccination in France, particularly in areas where vaccination rates are below 80 percent of the national average. Two types of messages are employed: broad-reaching direct messages and more personalised messages that encourage users to convince their friends to get vaccinated (referred to as friends messages). My analysis focuses on two outcome variables: the number of weekly first-dose and completed vaccinations. The study is an extension of the work by Ho et al. (2023), who explored whether these advertising messages increased vaccination rates. Their results reported insignificant and very small effects. In contrast, this study aims to investigate more thoroughly whether there are potential heterogeneous effects of the two types of messages.

Similar to Ho et al. (2023), I find no significant effect on the number of first dose and completed vaccinations for both messaging campaigns on average. However, the findings reveal significant heterogeneity in both positive and negative directions for both campaigns. Specifically, the direct messaging campaign highlighted particular adverse effect in poorer areas, indicating this treatment was counterproductive in stimulating vaccination rates among poorer individuals in France. Vulnerable populations may experience increased resistance against COVID-19 vaccinations due to a mistrust of health care authorities or pre-existing negative opinions about vaccination. Similarly, the friends messages particularly negatively affected areas where women earn lower average hourly wages. Socio-economic disadvantages and pre-existing beliefs can again be attributed to the effectiveness of this intervention. Overall, the result suggest that both messaging campaigns led to a decrease in the weekly number of first-dose and completed vaccinations among lower-income individuals. Conversely, individuals in wealthier areas responded positively to both treatments. The research underscores the importance of tailoring campaigns to address the specific needs and circumstances of different demographic groups. Understanding the characteristics that contribute to the heterogeneity in treatment effects can aid policymakers and health professionals in developing more targeted and effective strategies to promote COVID-19 vaccination and address future public health challenges. Future research should continue to explore these population dynamics, investigating alternative approaches to effectively reach and encourage these hard-to-convince individuals to get vaccinated despite pre-existing beliefs, a lack of medical understanding, or other factors influencing their decision to not get vaccinated.

References

Alberto Abadie, Matthew M Chingos, and Martin R West. Endogenous stratification in randomized experiments. *Review of Economics and Statistics*, 100(4):567–580, 2018.

Don Albrecht. Vaccination, politics and covid-19 impacts. *BMC Public Health*, 22(1):96, 2022.

Drew Altman. Why doctors and nurses can be vital vaccine messengers. *Kaiser Family Foundation*, 2021.

Manuela Angelucci, Dean Karlan, and Jonathan Zinman. Win some lose some? evidence from a randomized microcredit program placement experiment by compartamos banco. Technical report, National Bureau of Economic Research, 2013.

Susan Athey and Guido Imbens. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360, 2016.

Susan Athey and Guido W Imbens. Machine learning methods for estimating heterogeneous causal effects. *stat*, 1050(5):1–26, 2015.

Susan Athey, Julie Tibshirani, and Stefan Wager. Generalized random forests. 2019.

Susan Athey, Kristen Grabarz, Michael Luca, and Nils Wernerfelt. Digital public health interventions at scale: The impact of social media advertising on beliefs and outcomes related to covid vaccines. *Proceedings of the National Academy of Sciences*, 120(5):e2208110120, 2023.

Orazio Attanasio, Britta Augsburg, Ralph De Haas, Emla Fitzsimons, and Heike Harmgart. Group lending or individual lending? evidence from a randomised field experiment in mongolia. Technical report, WZB Discussion Paper, 2014.

Britta Augsburg, Ralph De Haas, Heike Harmgart, and Costas Meghir. Microfinance, poverty and education. 2012.

Abhijit Banerjee. Microcredit under the microscope: What have we learned in the past two decades, and what do we need to know? *Annu. Rev. Econ.*, 5(1):487–519, 2013.

Abhijit Banerjee, Esther Duflo, Rachel Glennerster, and Cynthia Kinnan. The miracle of microfinance? evidence from a randomized evaluation. *American economic journal: Applied economics*, 7(1):22–53, 2015.

Abhijit Banerjee, Emily Breza, Esther Duflo, and Cynthia Kinnan. Do credit constraints limit entrepreneurship? heterogeneity in the returns to microfinance. *Heterogeneity in the Returns to Microfinance (September 1, 2017). Global Poverty Research Lab Working Paper*, (17-104), 2017.

Eleonore Batteux, Freya Mills, Leah Ffion Jones, Charles Symons, and Dale Weston. The effectiveness of interventions for increasing covid-19 vaccine uptake: A systematic review. *Vaccines*, 10(3):386, 2022.

Alexandre Belloni, Victor Chernozhukov, and Kengo Kato. Uniform post-selection inference for least absolute deviation regression and other z-estimation problems. *Biometrika*, 102(1): 77–94, 2015.

Adi Berliner Senderey, Reut Ohana, Shay Perchik, Ido Erev, and Ran Balicer. Encouraging uptake of the covid-19 vaccine through behaviorally informed interventions: National real-world evidence from israel. *Available at SSRN 3852345*, 2021.

Iris Bräuninger. Dance movement therapy group intervention in stress treatment: A randomized controlled trial (rct). *The Arts in Psychotherapy*, 39(5):443–450, 2012.

Clare C Brown, Sean G Young, and George C Pro. Covid-19 vaccination rates vary by community vulnerability: A county-level analysis. *Vaccine*, 39(31):4245–4249, 2021.

Nina Buscemi, Ben Vandermeer, Carol Friesen, Liza Bialy, Michelle Tubman, Maria Ospina, Terry P Klassen, and Manisha Witmans. The efficacy and safety of drug treatments for chronic insomnia in adults: a meta-analysis of rcts. *Journal of general internal medicine*, 22: 1335–1350, 2007.

Pol Campos-Mercade, Armando N Meier, Florian H Schneider, Stephan Meier, Devin Pope, and Erik Wengström. Monetary incentives increase covid-19 vaccinations. *Science*, 374(6569): 879–882, 2021.

Victor Chernozhukov, Mert Demirer, Esther Duflo, and Ivan Fernandez-Val. Generic machine learning inference on heterogenous treatment effects in randomized experiments. 2017.

Victor Chernozhukov, Iván Fernández-Val, and Ye Luo. The sorted effects method: Discovering heterogeneous effects beyond their averages. *Econometrica*, 86(6):1911–1938, 2018.

Victor Chernozhukov, Mert Demirer, Esther Duflo, and Iván Fernández-Val. Fisher-schultz lecture: Generic machine learning inference on heterogenous treatment effects in randomized experiments, with an application to immunization in india, 2023.

R Lorraine Collins and G Alan Marlatt. Social modeling as a determinant of drinking behavior: Implications for prevention and treatment. *Addictive behaviors*, 6(3):233–239, 1981.

Bruno Crépon, Florencia Devoto, Esther Duflo, and William Parienté. Estimating the impact of microcredit on those who take it up: Evidence from a randomized experiment in morocco. *American Economic Journal: Applied Economics*, 7(1):123–150, 2015.

Hengchen Dai, Silvia Saccardo, Maria A Han, Lily Roh, Naveen Raja, Sitaram Vangala, Hardikkumar Modi, Shital Pandya, Michael Sloyan, and Daniel M Croymans. Behavioural nudges increase covid-19 vaccinations. *Nature*, 597(7876):404–409, 2021.

Jonathan MV Davis and Sara B Heller. Rethinking the benefits of youth employment programs: The heterogeneous effects of summer jobs. *Review of economics and statistics*, 102(4):664–677, 2020.

I de Vere Hunt, T Dunn, M Mahoney, M Chen, L Zhang, S Bousheri, D Bernard, and E Linos. A social media-based campaign to promote covid-19 vaccine uptake in underrepresented groups in the us. *European Journal of Public Health*, 31(Supplement_3):ckab165–170, 2021.

Ruben Dezeure, Peter Bühlmann, and Cun-Hui Zhang. High-dimensional simultaneous inference with the bootstrap. *Test*, 26:685–719, 2017.

Esther Duflo, Michael Kremer, and Jonathan Robinson. Understanding technology adoption: Fertilizer in western kenya, preliminary results from field experiments. *Unpublished manuscript, Massachusetts Institute of Technology*, 2004.

Esther Duflo, Pascaline Dupas, and Michael Kremer. The impact of free secondary education: Experimental evidence from ghana. Technical report, National Bureau of Economic Research, 2021.

W Douglas Evans, Jeffrey B Bingenheimer, Michael Long, Khadidiatou Ndiaye, Dante Donati, Nandan M Rao, Selinam Akaba, Ifeanyi Nsofor, and Sohail Agha. Outcomes of a social media campaign to promote covid-19 vaccination in nigeria. *Plos one*, 18(9):e0290757, 2023.

Jeffrey D Fisher and Stephen J Misovich. Social influence and aids-preventive behavior. In *Social influence processes and prevention*, pages 39–70. Springer, 1990.

Christopher Genovese and Larry Wasserman. Adaptive confidence bands. 2008.

Paul Glewwe and Michael Kremer. Schools, teachers, and education outcomes in developing countries. *Handbook of the Economics of Education*, 2:945–1017, 2006.

Christian Hansen, Damian Kozbur, and Sanjog Misra. Targeted undersmoothing. *arXiv preprint arXiv:1706.07328*, 2018.

Lisa Ho, Emily Breza, Abhijit Banerjee, Arun G Chandrasekhar, Fatima C Stanford, Renato Fior, Paul Goldsmith-Pinkham, Kelly Holland, Emily Hoppe, Louis-Maël Jean, et al. The impact of large-scale social media advertising campaigns on covid-19 vaccination: Evidence from two randomized controlled trials. In *AEA Papers and Proceedings*, volume 113, pages 653–658. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, 2023.

Kosuke Imai and Marc Ratkovic. Estimating treatment effect heterogeneity in randomized program evaluation. 2013.

Dean Karlan and Jonathan Zinman. Observing unobservables: Identifying information asymmetries with a consumer credit field experiment. *Econometrica*, 77(6):1993–2008, 2009.

Dean Karlan and Jonathan Zinman. Expanding credit access: Using randomized supply decisions to estimate the impacts. *The Review of Financial Studies*, 23(1):433–464, 2010.

Dean Karlan and Jonathan Zinman. Microcredit in theory and practice: Using randomized credit scoring for impact evaluation. *Science*, 332(6035):1278–1284, 2011.

Bridget J Kelly, Brian G Southwell, Lauren A McCormack, Carla M Bann, Pia DM MacDonald, Alicia M Frasier, Christine A Bevc, Noel T Brewer, and Linda B Squiers. Predictors of willingness to get a covid-19 vaccine in the us. *BMC infectious diseases*, 21:1–7, 2021.

Richard L Kravitz, Naihua Duan, and Joel Braslow. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *The Milbank Quarterly*, 82(4):661–687, 2004.

Sarah Kreps, Sandip Prasad, John S Brownstein, Yulin Hswen, Brian T Garibaldi, Baobao Zhang, and Douglas L Kriner. Factors associated with us adults' likelihood of accepting covid-19 vaccination. *JAMA network open*, 3(10):e2025594–e2025594, 2020.

Sören R Künzel, Jasjeet S Sekhon, Peter J Bickel, and Bin Yu. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the national academy of sciences*, 116(10):4156–4165, 2019.

Katherine Mackey, Chelsea K Ayers, Karli K Kondo, Somnath Saha, Shailesh M Advani, Sarah Young, Hunter Spencer, Max Rusek, Johanna Anderson, Stephanie Veazie, et al. Racial and ethnic disparities in covid-19–related infections, hospitalizations, and deaths: a systematic review. *Annals of internal medicine*, 174(3):362–373, 2021.

Amyn A Malik, SarahAnn M McFadden, Jad Elharake, and Saad B Omer. Determinants of covid-19 vaccine acceptance in the us. *EClinicalMedicine*, 26, 2020.

Rachael Meager. Aggregating distributional treatment effects: A bayesian hierarchical analysis of the microcredit literature. *American Economic Review*, 112(6):1818–1847, 2022.

Nicolai Meinshausen, Lukas Meier, and Peter Bühlmann. P-values for high-dimensional regression. *Journal of the American Statistical Association*, 104(488):1671–1681, 2009.

Nathaniel Rabb, Megan Swindal, David Glick, Jake Bowers, Anna Tomasulo, Zayid Oyelami, Kevin H Wilson, and David Yokum. Evidence from a statewide vaccination rct shows the limits of nudges. *Nature*, 604(7904):E1–E7, 2022.

Lorraine Ramig, Angela Halpern, Jennifer Spielman, Cynthia Fox, and Katherine Freeman. Speech treatment in parkinson's disease: Randomized controlled trial (rct). *Movement Disorders*, 33(11):1777–1791, 2018.

David M Ritzwoller and Joseph P Romano. Reproducible aggregation of sample-split statistics. *arXiv preprint arXiv:2311.14204*, 2023.

Joseph P Romano and Cyrus DiCiccio. *Multiple data splitting for testing.* Department of Statistics, Stanford University Stanford, CA, 2019.

Milda R Saunders, Regina James, Kesi Williams, and Marshall Chin. Addressing disparities in access to high-quality care. *The Science of Health Disparities Research*, pages 321–338, 2021.

Yue Sun and Shannon M Monnat. Rural-urban and within-rural differences in covid-19 vaccination rates. *The Journal of Rural Health*, 38(4):916–922, 2022.

Alessandro Tarozzi, Jaikishan Desai, and Kristin Johnson. On the impact of microcredit: Evidence from a randomized intervention in rural ethiopia. 2013.

Elizabeth L Turner, Pablo Perel, Tim Clayton, Phil Edwards, Adrian V Hernández, Ian Roberts, Haleema Shakur, Ewout W Steyerberg, CRASH Trial Collaborators, et al. Covariate adjustment increased power in randomized controlled trials: an example in traumatic brain injury. *Journal of clinical epidemiology*, 65(5):474–481, 2012.

Jos Twisk, Lisa Bosman, Trynke Hoekstra, Judith Rijnhart, Marieke Welten, and Martijn Heymans. Different ways to estimate treatment effects in randomised controlled trials. *Contemporary clinical trials communications*, 10:80–85, 2018.

A Inference Algorithm

Algorithm 1, as presented in Chernozhukov et al. (2023), provides the general algorithm for inference on heterogeneous effects. Refer to Section 3 for details on the computation of model estimates and evaluation statistics in the final steps of the algorithm.

Algorithm 1 Inference Algorithm with Random Splits

initialise:
Data set $\{(Y_i, D_i, Z_i, p(Z_i))\}$ on units $i \in [N] = \{1, \dots, N\}$, number of splits S, significance level α, set of ML or Causal ML methods
generate:
S random splits in half of $[N]$ into disjoint sets A and M, where $s = \{1, \dots, S\}$
$s \leftarrow 1$
while $s \leq S$ **do**
  Using A, tune and train each ML method to generate B and S
  Output these predictions B and S for M
  Optionally, choose the best or aggregate ML methods via (A.1) or (A.2)
  Estimate the BLP parameters by weighted OLS (3.3) or Horvitz-Thompson (C.1) in M
  Estimate the GATES parameters by weighted OLS (3.5) or Horvitz-Thompson (**??**) in M
  Estimate the CLAN parameters by taking averages (3.7) in M
  Compute goodness-of-fit measures via (3.8) and (3.9) for each model in M
  $s \leftarrow s + 1$
end while
Choose the best ML method based on the median-aggregated goodness-of-fit measures
Calculate the quantile-aggregated point estimates, p-values, and $(1 - 2\alpha)$ confidence bounds of each target parameter
return Final quantile-aggregated point estimates, p-values, $(1 - 2\alpha)$ confidence intervals of each target parameter in the BLP, GATES and CLAN

As shown in Algorithm 1, the best ML method to target the CATE can be chosen at two different stages, based on the sample used for this decision. In the main sample, the best ML method is selected based on the goodness-of-fit measures of either the BLP or GATES, as described in Section 3. Alternatively, the ML method can also be chosen earlier, using the auxiliary sample. In this case, the best ML method is determined by either minimising the error in predicting YH, or through the weighted prediction of Y. More specifically, one can solve either of the following problems, respectively:

$$(B,S) = \arg \min_{B\in\mathsf{B},S\in\mathsf{S}} \sum_{i\in A} w(Z_i)[Y_i - B(Z_i) - (D_i - p(Z_i))\{S(Z_i) - \bar{S}(Z_i)\}]^2, \tag{A.1}$$

$$(B,S) = \arg \min_{B\in\mathsf{B},S\in\mathsf{S}} \sum_{i\in A} [Y_i H_i - B(Z_i)H_i - S(Z_i)]^2, \tag{A.2}$$

where $\bar{S}(Z_i) = |A|^{-1}\sum_{i\in A} S(Z_i)$, and B and S are parameter spaces for $z \mapsto B(z)$ and $z \mapsto S(z)$. See Chernozhukov et al. (2023) for more details.

B Algorithm BLP

The coefficients β_1 and β_2 of $\mathsf{BLP}[s_0(Z)|S(Z)]$ can be identified and estimated through two strategies. The first strategy, outlined in Algorithm 2, considers a weighted linear projection that enhances the predictive accuracy of the BLP by leveraging proxy measures $S(Z)$ and $B(Z)$ (denoted by S and B in Algorithm 2) to reduce noise. Its effectiveness is demonstrated through a simulation example in Chernozhukov et al. (2023), which shows that the estimated BLP can predict the CATE more accurately than simple ML proxies, even when there is heterogeneity present in the CATE.

Algorithm 2 Best Linear Predictor (BLP), Strategy A

initialise:
Data set $\{(Y_i, D_i, Z_i, p(Z_i))\}$ on units $i \in [N] = \{1,\dots,N\}$, number of splits S, significance level α, set of ML or Causal ML methods, α_0
set:
$X_i = (X'_{1i}, X'_{2i})'$
$X_{1i} = [1, B(Z_i), p(Z_i), p(Z_i)S(Z_i)]'$
$X_{2i} = [D_i - p(Z_i), (D_i - p(Z_i))(S_i - \mathbb{E}_{N,M}[S_i])]'$
generate:
S random splits of $[N]$ into disjoint sets A and M, where $s = \{1,\dots,S\}$
$s \leftarrow 1$
while $s \le S$ **do**
Using A, tune and train each ML method to generate $B(Z_j)$ and $S(Z_j)$, for $j \in A$
Output these predictions $B(Z_j)$ and $S(Z_j)$ for $k \in M$
Estimate the BLP parameters by weighted OLS in M:

$$Y_k = \widehat{\alpha}_0' X_{1k} + \widehat{\alpha}_1(D_k - p(Z_k)) + \widehat{\alpha}_2(D_k - p(Z_k))(S_k - \mathbb{E}_{N,M}[S_k]) + \widehat{\epsilon}_k, \quad k \in M$$

Extract $\widehat{\alpha} = (\widehat{\alpha}_1, \widehat{\alpha}_2)'$
$s \leftarrow s + 1$
end while
Choose the best ML method based on the median-aggregated goodness-of-fit measures
Calculate the quantile-aggregated point estimates, p-values, and $(1-2\alpha)$ confidence bounds of $\widehat{\alpha} = (\widehat{\alpha}_1, \widehat{\alpha}_2)'$
return Final quantile-aggregated estimates for $\beta = (\beta_1, \beta_2)$, p-values, $(1-2\alpha)$ confidence intervals for parameters

The second strategy to identify and estimate β_1 and β_2 uses the Horvitz-Thompson transform H, defined as $H = H(D,Z) = \frac{D - p(Z)}{p(Z)(1-p(Z))}$, as detailed in Chernozhukov et al. (2023). The empirical implementation of this estimation strategy is:

$$Y_i H_i = \widehat{\mu}_0' X_{1i} H_i + \widehat{\mu}_1 + \widehat{\mu}_2 (S_i - \mathbb{E}_{N,M}[S_i]) + \widehat{\epsilon}_i, \quad i \in M \tag{B.1}$$

with $\mathbb{E}_{N,M}[\widehat{\epsilon}_i \tilde{X}_i] = 0$ and $\tilde{X}_i := (X_{1i}' H_i, \tilde{X}_{2i}')'$, $X_{1i} := [1, B(Z_i), p(Z_i), p(Z_i)S(Z_i)]'$, and $\tilde{X}_{2i} := (1, S_i - \mathbb{E}_{N,M}[S_i])'$. Note that $S := S(Z)$ and $B := B(Z)$ in Algorithm 3.

Algorithm 3 Best Linear Predictor (BLP), Strategy B

---

initialise:
Data set $\{(Y_i, D_i, Z_i, p(Z_i))\}$ on units $i \in [N] = \{1, \dots, N\}$, number of splits S, significance level α, set of ML or Causal ML methods, μ_0
set:
$\tilde{X}_i = (X_{1i}' H_i, \tilde{X}_{2i}')'$
$X_{1i} = [1, B(Z_i), p(Z_i), p(Z_i)S(Z_i)]'$
$\tilde{X}_{2i} = (1, S_i - \mathbb{E}_{N,M}[S_i])'$
generate:
S random splits of $[N]$ into disjoint sets A and M, where $s = \{1, \dots, S\}$
$s \leftarrow 1$
while $s \leq S$ **do**
    Using A, tune and train each ML method to generate $B(Z_j)$ and $S(Z_j)$, for $j \in A$
    Output these predictions $B(Z_j)$ and $S(Z_j)$ for $k \in M$
    Estimate the BLP parameters by weighted OLS in M:

$$Y_k H_k = \widehat{\mu}_0' X_{1k} H_i + \widehat{\mu}_1 + \widehat{\mu}_2 (S_k - \mathbb{E}_{N,M}[S_k]) + \widehat{\epsilon}_k, \quad k \in M$$

    Extract $\widehat{\mu} = (\widehat{\mu}_1, \widehat{\mu}_2)'$
    $s \leftarrow s + 1$
end while
Choose the best ML method based on the median-aggregated goodness-of-fit measures
Calculate the quantile-aggregated point estimates, p-values, and $(1 - 2\alpha)$ confidence bounds of $\widehat{\mu} = (\widehat{\mu}_1, \widehat{\mu}_2)'$
return Final quantile-aggregated estimates for $\beta = (\beta_1, \beta_2)$, p-values, $(1 - 2\alpha)$ confidence intervals for parameters

---

Chernozhukov et al. (2023) present two theorems demonstrating that the coefficients $\widehat{\alpha} = (\widehat{\alpha}_1, \widehat{\alpha}_2)'$ from strategy A and $\widehat{\mu} = (\widehat{\mu}_1, \widehat{\mu}_2)'$ from strategy B correctly identify the BLP coefficients $\beta = (\beta_1, \beta_2)$, respectively. In other words, both $\widehat{\alpha}$ and $\widehat{\mu}$ solve the best linear approximation problem for $s_0(Z)$ in their respective strategies. For more details, including proofs and a comparative analysis of both estimation strategies, I refer to Chernozhukov et al. (2023).

C Algorithm GATES

The identification and estimation strategies for the GATES parameters $\gamma = (\gamma_k)_{k=1}^K$ mimic those used for the BLP. Strategy A, detailed in Algorithm 4, utilises a weighted linear projection to estimate the GATES parameters. In contrast, strategy B uses a linear projection of the Horvitz-Thompson transformed variables. Both strategies are outlined in Algorithms 4 and 5, respectively. Note that $S := S(Z)$ and $B := B(Z)$.

Algorithm 4 Sorted Group Average Treatment Effects (GATES), Strategy A

initialise:
Data set $\{(Y_i, D_i, Z_i, p(Z_i))\}$ on units $i \in [N] = \{1, \ldots, N\}$, number of splits S, significance level α, set of ML or Causal ML methods, α_0
set:
$W_i = [X_{1i}', W_{2i}']'$
$X_{1i} = (B(Z_i), p(Z_i)\{1(G_k)_{k=1}^K\})'$
$W_{2i} = ([D_i - p(Z_i)] \cdot \{1(G_k)\}_{k=1}^K)'$
generate:
S random splits of $[N]$ into disjoint sets A and M, where $s = \{1, \ldots, S\}$
$s \leftarrow 1$
while $s \leq S$ **do**
Using A, tune and train each ML method to generate $B(Z_j)$ and $S(Z_j)$, for $j \in A$
Output these predictions $B(Z_j)$ and $S(Z_j)$ for $k \in M$
Estimate the GATES parameters by weighted OLS in M:

$$Y_k = \widehat{\alpha}_0' X_{1k} + \widehat{\alpha}' W_{2k} + \widehat{\nu}_k, \quad k \in M$$

Extract $\widehat{\alpha} = (\widehat{\alpha}_1, \ldots, \widehat{\alpha}_K)'$
$s \leftarrow s + 1$
end while
Choose the best ML method based on the median-aggregated goodness-of-fit measures
Calculate the quantile-aggregated point estimates, p-values, and $(1 - 2\alpha)$ confidence bounds of $\widehat{\alpha} = (\widehat{\alpha}_1, \widehat{\alpha}_2)'$
return Final quantile-aggregated estimates for $\gamma = (\gamma_1, \ldots, \gamma_K)$, p-values, $(1-2\alpha)$ confidence intervals for parameters

For estimation strategy B, Chernozhukov et al. (2023) define the empirical implementation by:

$$Y_i H_i = \widehat{\mu}_0' X_{1i} H_i + \widehat{\mu}' \tilde{W}_{2i} + \widehat{\upsilon}_i, \quad i \in M \tag{C.1}$$

with $\mathbb{E}_{N,M}[\widehat{\upsilon}_i \tilde{W}_i] = 0$ and where $\tilde{W}_i = (X_{1i}' H_i, \tilde{W}_{2i}')$, $X_{1i} := (B(Z_i), p(Z_i)\{1(G_k)_{k=1}^K\})'$, and $\tilde{W}_{2i} := [\{1(G_k)\}_{k=1}^K]'$. Details can be found in Chernozhukov et al. (2023) and Algorithm 5.

Algorithm 5 Sorted Group Average Treatment Effects (GATES), Strategy B

initialise:
Data set $\{(Y_i, D_i, Z_i, p(Z_i))\}$ on units $i \in [N] = \{1, \ldots, N\}$, number of splits S, significance level α, set of ML or Causal ML methods, μ_0
set:
$\tilde{W}_i = (X_{1i}' H_i, \tilde{W}_{2i}')$
$X_{1i} = (B(Z_i), p(Z_i)\{1(G_k)_{k=1}^K\})'$
$\tilde{W}_{2i} = [\{1(G_k)\}_{k=1}^K]'$
generate:
S random splits of $[N]$ into disjoint sets A and M, where $s = \{1, \ldots, S\}$
$s \leftarrow 1$
while $s \leq S$ **do**
Using A, tune and train each ML method to generate $B(Z_j)$ and $S(Z_j)$, for $j \in A$
Output these predictions $B(Z_j)$ and $S(Z_j)$ for $k \in M$
Estimate the GATES parameters in M:

$$Y_i H_i = \widehat{\mu}_0' X_{1i} H_i + \widehat{\mu}' \tilde{W}_{2i} + \widehat{v}_i, \quad i \in M$$

Extract $\widehat{\mu} = (\widehat{\mu}_1, \ldots, \widehat{\mu}_K)'$
$s \leftarrow s + 1$
end while
Choose the best ML method based on the median-aggregated goodness-of-fit measures
Calculate the quantile-aggregated point estimates, p-values, and $(1 - 2\alpha)$ confidence bounds of $\widehat{\alpha} = (\widehat{\alpha}_1, \widehat{\alpha}_2)'$
return Final quantile-aggregated estimates for $\gamma = (\gamma_1, \ldots, \gamma_K)$, p-values, $(1-2\alpha)$ confidence intervals for parameters

Chernozhukov et al. (2023) present a formal theorem regarding parameter identification, asserting that α_k, μ_k and γ_k are equal and identify the GATES as:

$$\alpha_k = \mu_k = \gamma_k = \mathbb{E}[s_0(Z)|G_k], \quad k = 1, \ldots, K. \tag{C.2}$$

D Summary Statistics Direct and Friends Messaging Campaign

Table 9 provides a detailed list of control variable descriptions that are used in the analysis. The variables are measured in the year 2020.

Table 9: Description of Control Variables

Variable	**Description**
SNHM20	Average net hourly wage in 2020 (euros)
SNHMC20	Average net hourly wage of executives, intellectual professions, and salaried business leaders in 2020 (euros)
SNHMP20	Average net hourly wage of intermediate professions in 2020 (euros)
SNHME20	Average net hourly wage of employees in 2020 (euros)
SNHMO20	Average net hourly wage of workers in 2020 (euros)
SNHMF20	Average net hourly wage of women in 2020 (euros)
SNHMFC20	Average net hourly wage of female executives or those in intellectual professions and salaried business leaders in 2020 (euros)
SNHMFP20	Average net hourly wage of women in intermediate professions in 2020 (euros)
SNHMFE20	Average net hourly wage of female employees in 2020 (euros)
SNHMFO20	Average net hourly wage of female workers in 2020 (euros)
SNHMH20	Average net hourly wage of men in 2020 (euros)
SNHMHC20	Average net hourly wage of male executives or those in intellectual professions and salaried business leaders in 2020 (euros)
SNHMHP20	Average net hourly wage of men in intermediate professions in 2020 (euros)
SNHMHE20	Average net hourly wage of male employees in 2020 (euros)
SNHMHO20	Average net hourly wage of male workers in 2020 (euros)
SNHM1820	Average net hourly wage of people aged 18 to 25 in 2020 (euros)
SNHM2620	Average net hourly wage of people aged 26 to 50 in 2020 (euros)
SNHM5020	Average net hourly wage of people over 50 in 2020 (euros)
SNHMF1820	Average net hourly wage of women aged 18 to 25 in 2020 (euros)
SNHMF2620	Average net hourly wage of women aged 26 to 50 in 2020 (euros)
SNHMF5020	Average net hourly wage of women over 50 in 2020 (euros)
SNHMH1820	Average net hourly wage of men aged 18 to 25 in 2020 (euros)
SNHMH2620	Average net hourly wage of men aged 26 to 50 in 2020 (euros)
SNHMH5020	Average net hourly wage of men over 50 in 2020 (euros)
NBMENFISC18	Number of fiscal households
NBPERSMENFISC18	Number of people in fiscal households
MED18	Median standard of living (€)
PIMP18	Percentage of taxable households (%)
TP6018	Poverty rate - Overall (%)
PACT18	Share of activity income (%)
PTSA18	of which share of wages and salaries (%)
PCHO18	of which share of unemployment benefits (%)
PBEN18	of which share of income from self-employment (%)
PPEN18	Share of pensions, retirements, and annuities (%)
PPAT18	Share of property income and other income (%)
PPSOC18	Share of total social benefits (%)
PPFAM18	of which share of family benefits (%)
PPMINI18	of which share of minimum social benefits (%)
PPLOGT18	of which share of housing benefits (%)
PIMPOT18	Share of taxes (%)
RD18	1st decile of standard of living (€)
D118	9th decile of standard of living (€)
D918	Interdecile ratio (9th decile/1st decile)

Descriptions for control variables used in the analysis. All monetary values are reported in euros.

Table 10: Descriptive Statistics of Areas for the Direct Messaging Campaign

	All	Treated	Control
Outcome Variables			
Number of Dose 1 Vaccinations (New Week)	82.343	89.902	78.461
Number of Completed Vaccinations (New Week)	91.079	99.681	86.661
Covariates			
Average Net Hourly Wage	15.379	15.309	15.414
Average Net Hourly Wage of People Aged 18 to 25	10.647	10.648	10.646
Median Standard of Living	21,938	21,947	21,934
Overall Poverty Rate	13.986	13.968	13.994
Share of Unemployment Benefits	2.983	2.948	3.001
Share of Income from Self-Employment	5.566	5.620	5.538
Share of Total Social Benefits	5.526	5.499	5.540
Share of Taxes	-17.009	-17.020	-17.003

Descriptive statistics for the outcome variables and a selection of covariates for the direct treatment. They are categorised into three groups: all observations, treated group observations, and control group observations.

Table 11: Descriptive Statistics of Areas for the Friends Messaging Campaign

	All	Treated	Control
Outcome Variables			
Number of Dose 1 Vaccinations (New Week)	82.343	81.771	82.634
Number of Completed Vaccinations (New Week)	91.079	90.382	91.433
Covariates			
Average Net Hourly Wage	15.379	15.435	15.350
Average Net Hourly Wage of People Aged 18 to 25	10.647	10.663	10.638
Median Standard of Living	21,938	22,002	21,906
Overall Poverty Rate	13.986	13.972	13.992
Share of Unemployment Benefits	2.983	3.010	2.969
Share of Income from Self-Employment	5.566	5.489	5.605
Share of Total Social Benefits	5.526	5.502	5.539
Share of Taxes	-17.009	-17.059	-16.984

Descriptive statistics for the outcome variables and selected covariates for the friends messages, categorized into three groups: all observations, treated group observations, and control group observations. The first three covariates are measured in euros for the year 2020.

E Boosting Analysis on the GATES for the Direct and Friends Messaging Campaign

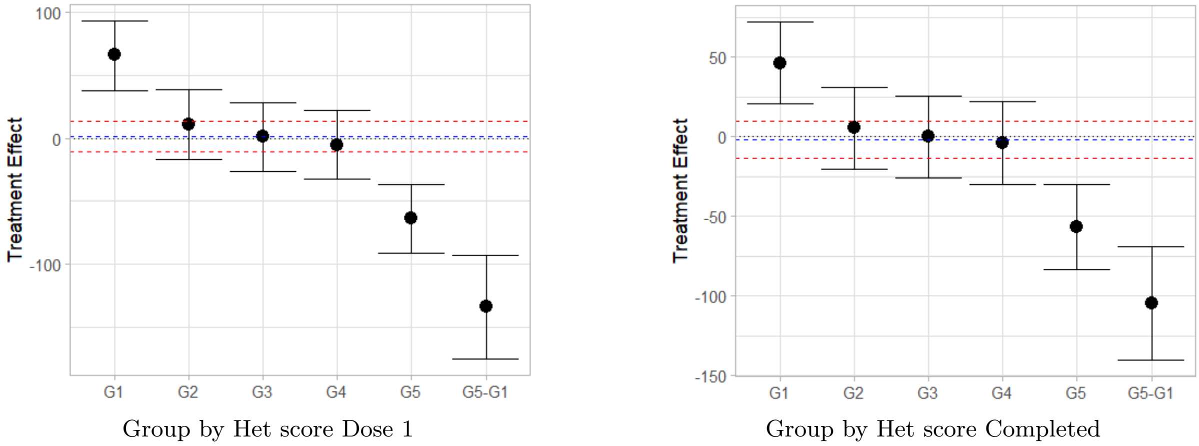


Figure 3: Grouping by Het score based on Boosting analysis for the direct messages. Het black dot represents het GATES estimate, with its respective confidence bands. The blue dashed line indicates the ATE and the red dashed lines show its confidence interval.

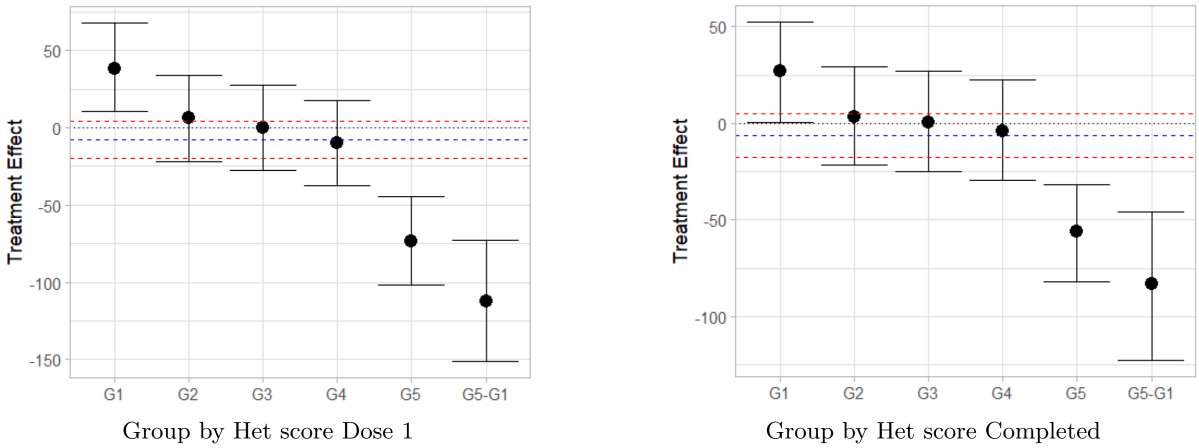


Figure 4: Grouping by Het score based on Boosting analysis for the friends messages. Het black dot represents het GATES estimate, with its respective confidence bands. The blue dashed line indicates the ATE and the red dashed lines show its confidence interval.

F Evaluation of Heterogeneity in the Effect of Microcredit Availability

The purpose of giving out microcredit is to give low-income individuals the opportunity to start their own business or become self-employed, thereby improving their standard of living. Borrowers tend to be from less developed countries. While the premise may seem valid, there have been numerous studies that have questioned the effectiveness and impact of microcredit. A summary of recent literature on microcredit availability can be found in Banerjee (2013). Research by Attanasio et al. (2014), Banerjee et al. (2015), Angelucci et al. (2013) and Tarozzi et al. (2013) show that, while access to microcredit positively impacts investments on self-employment activities, it does not lead to an overall increase in consumption, total income or profit. In addition, these studies often find no significant increase in business profits or income from self-employment activities. Experiments conducted on an individual level confirm these findings (Augsburg et al. (2012), Karlan and Zinman (2010) and Karlan and Zinman (2011)). In general, studies report that the difference in microcredit take-up between the treatment and control group is low, as exemplified by Crépon et al. (2015), who find a 13% microcredit take-up in the treatment group and 17% in a subsample of villages that was regarded as having a "higher probabiliy" of taking up microcredit.

Given these generally weak effects of microcredit availability across all units in the sample, an important question arises: could there be significant variations in how different units are affected, but that this is masked by average effects? Investigating this potential heterogeneity can provide insights into the effectiveness of microcredit on individuals' welfare, offering important implications for policy designs and targetting groups that would benefit most from microcredit availability. In fact, previous papers on heterogeneous treatment effects in microcredit availability have found interesting results, prompting the hypothesis that heterogeneity is indeed present in these experiments. To illustrate, Banerjee et al. (2017), who follow up on the study by Banerjee et al. (2015), show that microcredit has a much larger impact on business outcomes for individuals who had already started a business prior to receiving microcredit, compared to those without prior businesses. Meager (2022) supports this by proving that "had a prior business" is a generalisable and robust predictor. Additionally, Crépon et al. (2015) classify households in three categories based on their likelihood of borrowing pre-intervention and find that, among those most likely to borrow, microcredit access does not significantly impact income and consumption.

This empirical application replicates the research by Chernozhukov et al. (2017), who apply their generic approach to a study examing the heterogeneous effects of microcredit availability on borrowing and self-employment activities. This study, conducted between 2006 and 2007, included 162 rural villages in Morocco. These villages were divided into 81 pairs based on similar characteristics such as the number of households, existing infrastructure, and type of agriculture activities. In each pair, one village was randomly selected for the treatment—access to microcredit—while its counterpart acted as a control. After two years of intervention, household-level data from $N = 5,524$ households was collected through surveys. The primary goal of the study is to assess whether there are any changes in overall loan amount, output from self-employment activities, profit from self-employment activities, and monthly consumption (all in MAD[2]). Each of these outcome variables is represented by Y. Furthermore, D indicates whether a household is part of a treated village with access to microcredit, and Z includes eight household characteristics, including the number of household members, the age of the household's head, and whether or not the household has borrowed from any source. Along with that, Z also includes corresponding dummy variables and 81 fixed effects to control for the unique characteristics within each village pair, effectively isolating the treatment effect of microcredit availability. Furthermore, the study uses grouping variables: a village pair identifier and an identifier for the individual villages within each pair, which are used to control for fixed effects

[2] Moroccan Dirams

Table 12: Descriptive Statistics of Households

	All	Treated	Control
Outcome Variables			
Total Amount of Loans	2,364	2,934	1,807
Total output from self-employment activities (past 12 months)	32,559	35,199	29,976
Total profit from self-employment activities (past 12 months)	10,120	11,050	9,211
Total monthly consumption	3,012	2,996	3,027
Baseline Covariates			
Number of Household Members	3.878	3.870	3.887
Number of Members 16 Years Old or Older	2.604	2.600	2.608
Head Age	35.953	35.899	36.005
Declared Animal Husbandry Self-employment Activity	0.415	0.426	0.404
Declared Non-agricultural Self-employment Activity	0.146	0.128	0.163
Borrowed from Any Source	0.211	0.225	0.197
Spouse of Head Responded to Self-employment Section	0.067	0.074	0.061
Member Responded to Self-employment Section	0.044	0.047	0.041

Descriptive statistics for the four outcome variables and eight baseline covariates, categorised into three groups: all observations, treated group observations, and control group observations. Other covariates not shown in this table include the respective dummy variables of the eight baseline covariates, and 81 village pairs, totaling 96 covariates used in the analysis. Note that all monetary variables are expressed in Moroccan Dirams (MAD)

and clustering. Standard errors are clustered on the village level. Observations with missing data from the variables mentioned are excluded, resulting in a total of $N = 5,513$ villages in the dataset. Table 12 presents some descriptive statistics for the outcome variables and the eight baseline covariates.

The performance of four ML methods in predicting the proxy predictor $S(Z_i)$ across the four outcome variables is presented in Table 13. The results from this table indicate that Random Forest and Elastic Net seem to be the most reliable methods across all outcome variables for both goodness-of-fit measures. Although Neural Network and Boosting are generally effective, their relatively lower performances suggest that they may not be able to capture data patterns and relationships as effectively as Random Forest and Elastic Net. Accordingly, further analysis will focus on the latter two ML methods.

Table 13: Comparison of ML Methods: Microfinance Availability

	Random Forest	Elastic Net	Neural Network	Boosting
Amount of Loans				
Best BLP (Λ)	1,818,461	488,156	437,965	820,570
Best GATES ($\bar{\Lambda}$)	2,997,844	2,265,442	2,505,021	3,325,010
Output				
Best BLP (Λ)	16,843,626	73,911,741	38,737,276	11,999,845
Best GATES ($\bar{\Lambda}$)	122,369,634	160,578,356	101,897,011	103,226,183
Profit				
Best BLP (Λ)	17,151,533	8,010,214	10,320,273	3,683,870
Best GATES ($\bar{\Lambda}$)	38,093,159	22,767,522	31,885,759	32,403,059
Consumption				
Best BLP (Λ)	6,482	8,095	18,138	10,600
Best GATES ($\bar{\Lambda}$)	29,672	40,490	38,850	31,692

The table compares the performance of four ML methods—Random Forest, Elastic Net, Neural Network, and Boosting—in predicting microfinance availability, evaluated by four performance indicators. These indicators include the amount of loans, the output from self-employment activities, profit from self-employment activities, and monthly consumption. Each method's performance is evaluated by using the best BLP (Λ) and best GATES ($\bar{\Lambda}$) values, representing goodness-of-fit measures. The results are based on medians over 100 splits, reflecting robustness of the results across various data partitions, and indicate that Random Forest and Elastic Net are particularly effective predictors.

Table 14 reports the estimates of the average treatment effect (ATE) and heterogeneity loading (HET) parameters, β_1 and β_2, respectively, in the BLP for the four outcome variables. Confidence intervals are median-adjusted, accounting for variability across the sample splits, and are reported in the parentheses. The p-values are adjusted similarly and are reported in brackets, testing the hypothesis that the respective outcome parameter is zero. The estimated ATEs of microfinance availability show that the amount of loans is positive and statistically significant at the 1% significance level for both Elastic Net and Random Forest. However, the same cannot be said for the ATEs on output, profit and consumption, reporting statistically insignificant results at the 10% level. This indicates that while, on average, microfinance availability has a significant impact on the amount of money borrowed, it does not significantly affect the output and profit from self-employment activities or monthly consumption.

Next, considering the heterogeneity results, I reject the hypothesis that the HET estimates are zero at the 10% significance level for the amount of loans, output and profit, but this finding only holds with specific ML methods: Random Forest indicates significant heterogeneity in the effect of microfinance availability for the amount of laons and profit, while Elastic Net shows significant heterogeneity for output. Both methods however, do not uncover any heterogeneity for consumption. These results suggest that the availability of microcredit affects outcomes that involve business activities heterogeneously but shows no immediate impact on living standards, as measured by consumption, even for the most benefited households. Chernozhukov et al. (2017) argue that this might be because households use microloans to invest in their business

Table 14: BLP of Microfinance Availability

	Elastic Net		Random Forest	
	ATE (β_1)	HET (β_2)	ATE (β_1)	HET (β_2)
Amount of Loans	1,141	0.200	1,132	0.339
	(308,1994)	(-0.214,0.671)	(321,1964)	(-0.023,0.676)
	[0.008]	[0.294]	[0.005]	[0.066]
Output	5,625	0.278	4,929	0.120
	(-1483,12726)	(0.001,0.564)	(-2390,12162)	(-0.109,0.371)
	[0.118]	[0.048]	[0.181]	[0.324]
Profit	1,793	0.290	1,724	0.204
	(-2336,5879)	(-0.109,0.706)	(-2309,5809)	(-0.018,0.433)
	[0.389]	[0.146]	[0.413]	[0.069]
Consumption	-52.0	0.153	-63.5	0.079
	(-202,93.2)	(-0.234,0.514)	(-219,89.0)	(-0.212,0.406)
	[0.480]	[0.420]	[0.433]	[0.618]

The table reports the BLP of the CATE using the four performance indicators, presenting estimates of the average treatment effect (ATE) and heterogeneity loading (HET) parameters in the BLP, respectively. In the parentheses, 90% confidence intervals are reported, and median-adjusted p-values over 100 splits can be found in the brackets for the hypothesis that the parameter is zero. Results suggest that, on average, microfinance availability significantly impacts the amount of loans, but does not on output, profit and consumption at the 10% level. Furthermore, from the p-values of the HET estimates, one can conclude that there is no significant heterogeneity present in the effect on consumption.

rather than to increase consumption, resulting in different levels of business success and profit.

The estimates of the GATES are provided in Table 15, where households are divided into five groups based on quantile cutoffs of $S(Z_i)$. The average effect of each group is estimated, revealing significant heterogeneity primarily for the most affected group, denoted as γ_5. For this group, the GATES on the amount of loans, output and profit differs significantly from zero at the 10% significance level for both ML methods. The positive estimated effects on these outcome variables suggest that, on average, this group of households borrows more and achieves greater outputs and profits from their self-employment activities. This supports the hypothesis that microcredit availability helps these households in expanding their business ventures and economic activity. Further investigation shows statistically significant differences from zero between the most and least affected groups for the first three outcome variables at the 10% level: Random Forest identifies these differences for the amount of loans and profit, and Elastic Net does so for output. For consumption, both ML methods report statically insignificant results at the 10% level across all groups. These results indicate that, in conjunction with the results from the BLP analysis in Table 14, no heterogeneity is found in the outcome variable consumption. Similar to Chernozhukov et al. (2017), I will therefore omit the results of consumption in the CLAN analysis. Table 16 displays the average characteristics of the 20% most and least affected groups, based on the same quantiles as in Table 15, focusing on three factors: the age of the household head, investment in non-agricultural self-employment activities, and whether the household borrowed from any source. The results can help understand the cause of hetero-

Table 15: GATES of 20% Most and Least Affected Groups for Microcredit Availability

	Elastic Net			Random Forest		
	20% Most (γ_5)	20% Least (γ_1)	Difference ($\gamma_5 - \gamma_1$)	20% Most (γ_5)	20% Least (γ_1)	Difference ($\gamma_5 - \gamma_1$)
Amount of Loans	2,055	530	1,720	3,014	-186	2,924
	(226,3843)	(-1232,2455)	(-1152,4390)	(1008,4985)	(-2176,1954)	(280,5827)
	[0.016]	[0.599]	[0.223]	[0.004]	[0.879]	[0.030]
Output	24,171	-2,634	25,975	20,890	1,543	18,492
	(4160,45799)	(-13643,8751)	(2196,51201)	(1357,40300)	(-15162,16965)	(-6610,44215)
	[0.017]	[0.677]	[0.032]	[0.033]	[0.815]	[0.142]
Profit	6,512	617	6,566	11,976	-335	12,019
	(-1164,15177)	(-7935,8998)	(-6476,19211)	(546,23167)	(-8773,7724)	(-1771,26071)
	[0.094]	[0.825]	[0.312]	[0.036]	[0.947]	[0.084]
Consumption	38.8	-342	343	-15.1	-212	239
	(-281,328)	(-809,141)	(-218,922)	(-362,347)	(-670,198)	(-342,833)
	[0.808]	[0.165]	[0.231]	[0.949]	[0.321]	[0.392]

The results in the table report the GATES analysis on the most affected group, the least affected group, and the difference between these two. Based on quantile cutoffs, $K = 5$ groups are created, and Elastic Net and Random Forest are used to estimate the average affects of the three categories in the table on the four outcome variables. 90% confidence intervals are reported in parentheses, and median-adjusted p-values over 100 splits can be found in the brackets. One can observe that the most affected group reports statistically significant results at the 10% level for the first three metrics. The same can be said for the difference between the most and least affected groups for these three outcome variables, only now just one the ML tools reports statistically significant results at the 10% level. In general, all four metrics report insignificance results for the least affected group and the consumption outcome variable reports only insignificant results across all groups. In conjunction with the results from the BLP analysis in Table 14, one can conclude that the consumption metric does not contain any heterogeneity.

geneity in the treatment effects. From the first outcome variable—amount of loans—the results suggest that households with younger household heads that are involved in non-agricultural self-employment and that borrow less from other sources are likely to borrow more from a microfinance institution. This suggests that households see formal loans from the microfinance institution as substitutes for their overall borrowing rather than complements. For the output, households with similar characteristics are likely to obtain more financial output from their self-employment activities due to microcredit availability. Regarding the profit, only Random Forest yields statistically significant results, specifically for the average non-agricultural self-employment sector and for households that borrow from any source. Important to note here is that the p-value of the difference between δ_5 and δ_1 can be prone to the sensitivity of the ML method that is used to generate the ML proxy. This explains why Random Forest and Elastic Net do not consistently produce the same significance levels in their results. Overall, this study's main findings show that households involved in non-agricultural self-employment activities and those that have borrowed money from other sources particularly benefit from microcredit availability.

Table 16: CLAN of Microfinance Availability

	Elastic Net			Random Forest		
	20% Most (δ_5)	20% Least (δ_1)	Difference ($\delta_5 - \delta_1$)	20% Most (δ_5)	20% Least (δ_1)	Difference ($\delta_5 - \delta_1$)
Amount of Loans						
Head Age	31.3	41.1	-12.0	24.3	37.2	-12.7
	(29.0,33.5)	(38.9,42.7)	(-14.8,-8.69)	(22.0,26.6)	(35.3,39.1)	(-15.8,-9.69)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Non-agricultural self-emp.	0.166	0.069	0.091	0.145	0.121	0.018
	(0.134,0.196)	(0.042,0.090)	(0.052,0.129)	(0.115,0.174)	(0.094,0.149)	(-0.020,0.056)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.362]
Borrowed from Any Source	0.186	0.193	-0.001	0.138	0.281	-0.141
	(0.154,0.217)	(0.159,0.227)	(-0.048,0.047)	(0.109,0.166)	(0.243,0.318)	(-0.187,-0.093)
	[0.000]	[0.000]	[1.000]	[0.000]	[0.000]	[0.000]
Output						
Head Age	35.9	38.5	-2.33	33.4	30.8	3.02
	(33.9,38.0)	(36.4,40.6)	(-5.21,0.547)	(31.2,35.6)	(28.6,33.0)	(-0.216,6.18)
	[0.000]	[0.000]	[0.112]	[0.000]	[0.000]	[0.060]
Non-agricultural self-emp.	0.330	0.025	0.302	0.255	0.116	0.141
	(0.291,0.369)	(0.012,0.038)	(0.262,0.343)	(0.217,0.290)	(0.089,0.143)	(0.095,0.187)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Borrowed from Any Source	0.196	0.242	-0.048	0.188	0.200	-0.015
	(0.163,0.229)	(0.206,0.278)	(-0.096,0.000)	(0.155,0.219)	(0.166,0.233)	(-0.062,0.032)
	[0.000]	[0.000]	[0.051]	[0.000]	[0.000]	[0.529]
Profit						
Head Age	36.2	35.5	0.623	31.5	33.0	-1.63
	(34.1,38.2)	(33.4,37.6)	(-2.31,3.56)	(29.4,33.6)	(30.7,35.4)	(-4.82,1.57)
	[0.000]	[0.000]	[0.654]	[0.000]	[0.000]	[0.318]
Non-agricultural self-emp.	0.160	0.127	0.020	0.176	0.119	0.056
	(0.130,0.191)	(0.096,0.157)	(-0.024,0.063)	(0.144,0.208)	(0.092,0.147)	(0.014,0.098)
	[0.000]	[0.000]	[0.305]	[0.000]	[0.000]	[0.005]
Borrowed from Any Source	0.206	0.183	0.013	0.152	0.196	-0.043
	(0.171,0.240)	(0.151,0.215)	(-0.033,0.062)	(0.122,0.182)	(0.163,0.229)	(-0.089,-0.003)
	[0.000]	[0.000]	[0.548]	[0.000]	[0.000]	[0.067]

The table presents the CLAN estimates for three characteristics across three outcomes variables. The characteristics include the age of the household head, investment in non-agricultural self-employment activities, and whether the household has borrowed from any source. The 20% most and least affected groups are reported for all outcomes, with 90% confidence intervals in parentheses, and median-adjusted p-values in brackets. Significant results across all three outcomes are found for the latter two characteristics, indicating that microcredit avilability is especially beneficial for households with these traits.

G Programming Code

The R scripts used for both empirical analyses in this study are made available. Each analysis includes a preprocessing script that loads and preprocesses the data, and another script that performs that analysis using the `GenericML` framework. Packages such as `mlr3`, `ggplot2L`, and `GenericML` are employed to ensure robust analysis. Computations were performed on a Windows 11 Home system with 4 cores, using R version 4.4.0.