Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis Econometrics & Operations Research

# Heterogeneity Analysis in Microcredit Financing: Application of Machine Learning Inference Methods

Sem van Embden (599366se)

| | |
|---|---|
| Supervisor: | (Max) M.F.O. Welz |
| Second assessor: | Prof. dr. Ph.H.B.F. (Philip Hans) Franses |
| Date final version: | 1st July 2024 |

**Abstract**

Microcredit financing has been shown to be a powerful tool for economic empowerment and poverty reduction. In this study, we use newly developed methods to analyze, estimate, and infer the heterogeneous effects in two randomized control trials (RCTs) conducted in Morocco and Bosnia and Herzegovina, where microcredit financing was introduced. By making use of key features such as the Best Linear Predictor (BLP), Group Average Treatment Effects (GATES), and Classification Analysis (CLAN), we estimate and interpret the varied treatment effects of microcredit on outcome variables related to e.g. outstanding credit, employment, and business outcomes. Using machine learning methods such as Random Forest and Boosting to estimate the Conditional Average Treatment Effect (CATE) I show that in Morocco, microcredit significantly increases the loan amount ($p < 0.01$) by 1, although the impact on output, profits and consumption is modest increasing by 5.2, 1.4 and -62 Moroccan Dhirams, respectively, and not statistically significant (p-values of 0.153, 0.486, and 0.425), with the significant heterogeneity in the amount of loans and profit ($p < 0.1$) primarily driven by fixed effects present in the data rather than individual characteristics. In Bosnia and Herzegovina, microcredit also shows varied impacts in outcome variables with most notably a significant increase in the amount of loans by 0.4 ($p < 0.01$), and decrease in savings by 456 Bosnian Marks ($p < 0.1$), yet the heterogeneity is less pronounced, only being close to significant for savings ($p = 0.13$). These findings highlight the importance of considering heterogeneous effects in microfinance program evaluations, suggesting that while microcredit can aid economic development, its effectiveness may depend on targeting and contextual factors.

# 1 Introduction

Microfinance, particularly microcredit, has become a significant instrument for economic empowerment and poverty reduction. Its roots trace back to Bangladesh, where the Grameen Bank, founded by Muhammad Yunus in 1983, aimed to provide small loans to the impoverished on favorable terms. For his work, Yunus was awarded the Nobel Peace Prize in 2006[1].

Numerous randomized control trials (RCTs) and studies have explored the effectiveness of microcredit. These studies have generally found that microcredit significantly improves the livelihoods of poor individuals who lack access to formal financial institutions. For instance, Aslam et al. (2020) demonstrated the positive impacts of microfinance on Grameen Bank borrowers, while Alemu and Ganewo (2023) found that microcredit significantly increased the income of borrowers in Ethiopia.

Additional studies have evaluated microcredit impacts through RCTs conducted in various countries, including Mexico, Mongolia, Bosnia and Herzegovina, Morocco, and Nepal (Angelucci et al., 2015; Attanasio et al., 2015; Augsburg et al., 2015; Banerjee et al., 2015; Crépon et al., 2015; Dhungana et al., 2022).

The effects of microcredit, however, are often moderate rather than great and vary depending on individual characteristics (Armendáriz de Aghion and Morduch, 2005). This heterogeneity suggests that while some individuals benefit greatly from microcredit, others see less impact. This study seeks to measure this heterogeneity using data from two RCTs: one by Crépon et al. (2015) and another by Augsburg et al. (2015).

The concept of heterogeneity in microcredit effects is not new. Kolstad et al. (2017) analyzed the relationship between group heterogeneity and microcredit group exit rates in Angola, while Banerjee et al. (2015) found that heterogeneity persisted years after microfinance was introduced, with more significant impacts on those who had pre-existing businesses compared to those without.

Traditional methods for analyzing heterogeneity include subgroup analysis, which compares effects between treatment groups across various patient subsets (Christensen et al., 2021), and the sorted effects method; a semi-parametric inference method for characterizing heterogeneity (Chernozhukov et al., 2018). Recently, innovative strategies have been developed to estimate and infer key features of heterogeneous effects in randomized experiments, including machine learning techniques to predict effects and sort average effects by impact groups (Chernozhukov et al., 2020). These new methods do not rely on strong assumptions, can be used with various machine learning and statistical techniques, and are valid in high dimensional settings. Furthermore, for inference, these approaches employ a variational inference method where medians of p-values and medians of confidence intervals, resulting from numerous data splits, are adjusted to ensure uniform validity. This method effectively quantifies the uncertainty arising from both parameter estimation and data splitting, offering a robust and reliable means to understand treatment effects across heterogeneous populations. Understanding how these treatment effects vary by covariates such as gender and age is crucial for assessing program impacts and underlying mechanisms.

In this study, I apply these machine learning methods to analyze heterogeneous treatment

[1] https://www.nobelpeaceprize.org/laureates/2006

effects in the RCTs conducted by Crépon et al. (2015) and Augsburg et al. (2015). While the latter briefly considers heterogeneity, focusing on education and business-related variables, this study expands this analysis to include additional baseline variables. I aim to provide a more comprehensive analysis of heterogeneity, replicating the analysis form and applying the methods developed by Chernozhukov et al. (2020).

I find that, for the Moroccan dataset, there is significant heterogeneity present for the outcome variables Amount of Loans and Output. This heterogeneity is however mainly driven by fixed effects existing in the data, rather than individual characteristics of households. Concerning Bosnia, no significant heterogeneity is found. Nonetheless, significant differences with respect to baseline covariates are observed across the affected quantile groups for certain outcome variables.

This study contributes to the literature by demonstrating the utility of machine learning methods in understanding heterogeneous treatment effects and providing nuanced insights for more effective policy design. The comparison between the - at the time - more developed Bosnia and Herzegovina and steadily developing Morocco further emphasizes that microcredit programs must be context-specific to maximize their effectiveness and impact on economic development.

The structure of this paper is as follows: Section 2 discusses the data used, Section 3 outlines the methodology, Section 4 presents the results, and Section 5 concludes the paper.

# 2 Data

The data used in this study is from two randomized control trials (RCTs) conducted in Morocco, and Bosnia and Herzegovina. Further explanation of the data is divided into the two corresponding subsections.

## 2.1 Randomized Experiment in Morocco

I first analyze the presence of heterogeneity in the data from an RCT conducted in Morocco. Originally from the paper by Crépon et al. (2015), this analysis has been done before by Chernozhukov et al. (2017). The RCT was carried out with 81 village pairs with similar characteristics, of which one assigned to control and one to treatment. My data is obtained from the replication package[2] and in total at endline contains 5513 households. This is less than the dataset used by aforementioned authors, whose sample contains 5,551 households.

I consider the treatment as the introduction of microcredit, given that there was no other access to microcredit in these villages before and during the study period. Similar to Chernozhukov et al. (2017), we consider the following outcome variables $Y$: Amount of Loans, Output, Profit, and Consumption. With $D$ being an indicator that identifies whether a household resides in a treated village. The covariates, $Z$, encompass various baseline household characteristics, such as the number of members, the number of adults, the age of the head, and indicators for households engaged in animal husbandry or other non-agricultural activities. Additionally, $Z$ includes whether the household had an outstanding loan in the past 12 months, if the household spouse responded to the survey, if another household member (excluding the head) responded to the

[2] https://github.com/mwelz/GenericML/tree/main/slides/data

survey, and controls for 81 village pair - in which randomization took place - fixed effects. Summary statistics for the data are presented in the Appendix, Table 11.

### 2.2 Randomized Experiment in Bosnia and Herzegovina

The second dataset is from an RCT conducted in Bosnia and Herzegovina, originally analyzed by Augsburg et al. (2015). This field experiment was carried out during 2008 to 2010 in collaboration with a large Bosnian microfinance institution (MFI) established in the mid-nineties, serving a client base of 36,000 across the country at the time of the baseline survey. The MFI extended microcredit to a marginal segment of the population, individuals who would typically be rejected for loans but might be considered if slightly more risk was accepted.

As taken from Augsburg et al. (2015), the baseline survey revealed that the average marginal applicant did not meet 2.6 out of the 6 main requirements for regular loans, with 77% lacking sufficient collateral or failing one or more other requirements. Around one-third of these marginal clients were judged to have weak business proposals, and loan officers were concerned about repayment capacity in about a quarter of the cases. Additionally, 28.2% of the sample lived in urban areas with populations exceeding 50,000. At baseline, 78% had some income from self-employment, distributed across trade (27%), services (29%), agriculture (38%), and manufacturing (6%).

The experiment involved offering loans similar to the MFI's regular products, with an interest rate of 22% Annual Percentage Rate (APR) and maturities averaging 57 weeks. The loan amounts ranged from BAM[3] 300 to BAM 3,000, with a mean of BAM 1,653 (approximately US$1,012 at the baseline exchange rate at that time).

In total after the follow-up, 995 observations remained. In their paper, Augsburg et al. (2015) indicate that there are no statistically significant differences between treatment and control group - except a small difference in the number of household members - and conclude that there is no systematic overall difference between the two groups and no evidence of imbalance.

The outcome variables in this study are divided into subcategories, namely impact on Credit Outstanding at Endline; impact on Self-Employment Activities; impact on Income; impact on Hours Worked by Household Members; and lastly impact on Consumption and Savings. We omit the Social Impact outcome variables, as data on these variables is not available. Furthermore, for the outcome variables related to the number of hours worked, we only focus on the hours worked for all adults and teens (aged 16 to 64), as the data for the age group 16 to 19 does not contain enough non-zero observations to make meaningful inferences, or the data is not available in the case of the number of staff members. An overview of all the outcome variables and their descriptions is presented in the Appendix and summary statistics are given in Table 12 in the Appendix. In this table is also included the summary statistics for the covariates, of which a full list with descriptions is also given in the Appendix.

Not all outcome variables are available for all observations. The authors of the original paper set some of these missing values to zero. In some cases this is logical, as it is likely that the missing values for, par example, the act of starting a business (dummy variable) are zero. However, this is also done for the profit, revenue, and expenses which is not necessarily the

[3]BAM: Bosnia and Herzegovina convertible mark.

case and might heavily impact the results. Therefore, when we apply the methods described in Section 3 to these three outcome variables we drop these around 600 missing observations.

# 3 Methodology

In this section, I outline the methods for estimating and making inferences on heterogeneous treatment effects using machine learning proxies. I describe the Best Linear Predictor (BLP), Group Average Treatment Effects (GATES), and Classification Analysis (CLAN). For more detailed information on these features, I refer to Chernozhukov et al. (2020). The advantage of these methods is that they are not reliant on strong assumptions. Additionally, they are flexible, allowing for the use of numerous machine learning and statistical techniques, and are also valid in high dimensional settings. This approach is versatile regarding the ML method employed and does not depend on its formal properties.

## 3.1 Model

I first define the model by Chernozhukov et al. (2020). I consider data $(Y_i, Z_i, D_i)_{i=1}^N$, which consists of i.i.d. copies of the random vector $(Y, Z, D)$ with probability law $\mathbb{P}$. Here, $Y$ represents the outcome of interest, $D$ is a binary treatment indicator, and $Z$ is a possibly high-dimensional vector of covariates that characterize the observational units. The expectation operator is denoted by $\mathbb{E}$. Furthermore, I let $Y(1)$ and $Y(0)$ be the potential outcomes under treatment and control, respectively.

Then the main causal functions are the Baseline Conditional Average (BCA) defined as

$$b_0(Z) := \mathbb{E}[Y(0) \mid Z],$$

and the Conditional Average Treatment Effect (CATE) defined as

$$s_0(Z) := \mathbb{E}[Y(1) - Y(0) \mid Z] = \mathbb{E}[Y(1) \mid Z] - \mathbb{E}[Y(0) \mid Z].$$

I make the following assumptions. First that the treatment $D$ is randomly assigned conditional on $Z$, with the probability of assignment depending on a subvector of stratifying variables $Z_1 \subseteq Z$:

$$D \perp (Y(1), Y(0)) \mid Z,$$

and secondly that the propensity score $p(Z)$ is known, defined as $P[D = 1 \mid Z] = \mathbb{P}[D = 1 \mid Z_1]$, and is bounded away from zero or one:

$$p(Z) \in [p_0, p_1] \subset (0, 1) \quad a.s.$$

The observed outcome is modeled as:

$$Y = DY(1) + (1 - D)Y(0)$$

Under the stated assumptions, the causal functions are identified by the components of the regression function of $Y$ given $D$ and $Z$:

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

where:

$$b_0(Z) = \mathbb{E}[Y \mid D = 0, Z],$$

$$s_0(Z) = \mathbb{E}[Y \mid D = 1, Z] - \mathbb{E}[Y \mid D = 0, Z].$$

This framework supports the use of predictive machine learning methods to learn $\mathbb{E}[Y|D, Z]$ and subsequently estimate the CATE using the derived formula.

### 3.2 Agnostic Approach

Let $(M, A)$ denote a random partition of the set of indices $\{1, \ldots, N\}$. The strategies considered involve randomly splitting the data $(Y_i, D_i, Z_i)_{i=1}^N$ into a main sample, denoted by $\text{Data}_M = (Y_i, D_i, Z_i)_{i \in M}$, and an auxiliary sample, denoted by $\text{Data}_A = (Y_i, D_i, Z_i)_{i \in A}$. It is assumed but not theoretically required that the main and auxiliary samples are approximately equal in size.

From the auxiliary sample $A$, ML estimators of the baseline and treatment effects are obtained, referred to as proxy predictors:

$$z \mapsto B(z) = B(z; \text{Data}_A) \quad \text{and} \quad z \mapsto S(z) = S(z; \text{Data}_A)$$

These are potentially biased and noisy predictors of $b_0(z)$ and $s_0(z)$, and in principle, consistency for $b_0(z)$ and $s_0(z)$ is not required. These estimates are treated as proxies, which are post-processed to estimate and make inferences on the features of the CATE $z \mapsto s_0(z)$. The analysis conditions on the auxiliary sample $\text{Data}_A$, considering these maps as fixed when working with the main sample.

### 3.3 Best Linear Predictor (BLP)

The goal of the BLP is to estimate the best linear predictor of the CATE using machine learning proxies. The BLP of $s_0(Z)$ is estimated using a weighted linear projection:

$$Y = \alpha' X_1 + \beta_1(D - p(Z)) + \beta_2(D - p(Z))(S - \mathbb{E}[S]) + \varepsilon,$$

where $S = S(Z)$ is the ML proxy predictor mentioned earlier, $X = (X_1, X_2)$ with $X_1 = [1, B(Z)]$ and $X_2 = [D - p(Z), (D - p(Z))(S - \mathbb{E}[S])]$, and $\varepsilon$ the error term.

OLS is then used to estimate the coefficients. The coefficients $\beta_1$ and $\beta_2$ are identified by solving:

$$\mathbb{E}[w(Z)\varepsilon X] = 0,$$

with the weight being defined as:

$$w(Z) = \{p(Z)(1 - p(Z))\}^{-1}.$$

The BLP of the CATE is then given by:

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

Here $\beta_1$ represents the average treatment effect (ATE), while $\beta_2$ captures the heterogeneity (HET) in the treatment effect explained by the machine learning proxy $S(Z)$, indicating how this effect varies with the covariates. Another estimation technique using a Horvitz-Thompson transformation exists, but I will not consider this in our analysis and therefore omit the details in this paper, referring to Chernozhukov et al. (2020) for more information.

### 3.4 Group Average Treatment Effects (GATES)

GATES focuses on estimating the average treatment effects for groups defined by the ML proxy predictor $S(Z)$. The groups $G_k$ are defined by sorting $S(Z)$ into $K$ non-overlapping intervals.

I recover the GATES parameters from the weighted linear projection, where again OLS can be used for estimation:

$$Y = \gamma_0' X_1 + \sum_{k=1}^{K} \gamma_k (D - p(Z)) 1(G_k) + \nu,$$

where $G_k = \{S(Z) \in I_k\}$, i.e. the range of $S(Z)$ is divided into $K$ non-overlapping intervals $I_k = [\ell_{k-1}, \ell_k)$, with boundaries $\ell_1, \ell_2, \ldots, \ell_K$, such that $-\infty = \ell_0 < \ell_1 < \ldots < \ell_K = +\infty$. By dividing the range of $S(Z)$ into intervals, the GATES model captures the heterogeneity in treatment effects across different segments of the population. Each group $G_k$ represents a different range of predicted CATE's, allowing for analysis of how treatment effects vary. The error term is defined by $\nu$.

The GATES parameters $\gamma_k$ are identified by:

$$\gamma_k = \mathbb{E}[s_0(Z) \mid G_k].$$

Again, I refer to Chernozhukov et al. (2020) for more detailed information on the GATES, and the Horvitz-Thompson transformation.

### 3.5 Classification Analysis (CLAN)

CLAN aims to identify the characteristics of the most and least affected units based on $S(Z)$. The average characteristics, with $g(Y, D, Z)$ representing a vector of characteristics for each observational unit, of the most and least affected groups $G_1$ and $G_K$ are compared:

$$\delta_1 = \mathbb{E}[g(Y, D, Z) \mid G_1],$$

$$\delta_K = \mathbb{E}[g(Y, D, Z) \mid G_K].$$

The differences $\delta_K - \delta_1$ provide insights into the characteristics that correlate with the heterogeneity in treatment effects.

### 3.6 Variational Estimation and Inference

To ensure robust estimates, the methodology relies on data splitting to account for two sources of uncertainty: estimation uncertainty conditional on the data split and uncertainty induced by the random partitioning of data.

For point estimates, the median of the estimated key features over different random splits is reported:

$$\hat{\theta} = \mathrm{Med}[\hat{\theta}_A \mid \mathrm{Data}].$$

Here $\hat{\theta}_A$ represents the estimate of a key feature of the CATE based on a specific random partition of the data into an auxiliary sample $A$ and a main sample $M$. For confidence intervals:

$$[l, u] = [\mathrm{Med}[L_A \mid \mathrm{Data}], \mathrm{Med}[U_A \mid \mathrm{Data}]]\,,$$

where $[L_A, U_A]$ are the conditional confidence intervals.

This variational inference method, where medians of p-values and medians of confidence intervals - resulting from numerous data splits - are adjusted ensure uniform validity.

### 3.7 Goodness of Fit Measures

The paper employs several goodness of fit measures to evaluate the effectiveness of machine learning (ML) models in predicting Conditional Average Treatment Effects (CATE).

#### 3.7.1 Best Linear Predictor (BLP) Measure

For evaluating the fit of the CATE, Chernozhukov et al. (2020) introduces a measure denoted by $\Lambda$. This measure is defined as:

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

Maximizing $\Lambda$ is equivalent to maximizing the correlation between the ML proxy predictor $S(Z)$ and the true CATE $s_0(Z)$, which in turn is equivalent to maximizing the $R^2$ in the regression of $s_0(Z)$ on $S(Z)$. Therefore, an ML method that attains a higher $\Lambda$ is considered superior.

#### 3.7.2 Group Average Treatment Effects (GATES) Measure

Analogously, for the GATES analysis, the measure $\overline{\Lambda}$ is defined as:

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

This measure captures the part of the variation in $s_0(Z)$ explained by $S(Z)$ across $K$ quantiles. Choosing the ML proxy $S(Z)$ to maximize $\overline{\Lambda}$ is equivalent to maximizing the $R^2$ in the regression of $s_0(Z)$ on the grouped predictor $S(Z)$ (without a constant). If the groups

$G_k = \{S \in I_k\}$ have equal size, the measure simplifies to:

$$\overline{\Lambda} = \frac{1}{K}\sum_{k=1}^{K}\gamma_k^2.$$

An ML method achieving a higher $\overline{\Lambda}$ is preferred as it indicates a better fit.

### 3.7.3 Empirical Versions

The empirical versions of the above measures are calculated as:

$$\hat{\Lambda} = |\hat{\beta}_2|^2 \mathbb{E}_{N,M}(S_i - \mathbb{E}_{N,M}S_i)^2,$$

$$\hat{\overline{\Lambda}} = \sum_{k=1}^{K}\hat{\gamma}_k^2 \mathbb{E}_{N,M} 1\{S_i \in I_k\}.$$

Here $S_i$ denotes the value of the ML proxy predictor $S(Z)$ for the $i$-th observation. It is the predicted value of the CATE for each individual in the dataset. Furthermore, $E_{N,M}$ represents the empirical expectation over the main sample $M$ of size $n$. Specifically, for a given function $h(X_i)$, the empirical expectation is computed as:

$$E_{N,M}[h(X_i)] = \frac{1}{n}\sum_{i\in M} h(X_i).$$

These empirical versions are used to compare and select among different ML methods. From hereafter written as $\Lambda$ and $\hat{\Lambda}$

## 3.8 Estimation Algorithm

The following steps outline the procedure for estimating and making inferences on heterogeneous treatment effects using the BLP, GATES, and CLAN methods described earlier. Below algorithm ensures robust and valid estimates by incorporating repeated data splitting and machine learning proxies and is taken and modified from Chernozhukov et al. (2020).

### 3.8.1 Algorithm

The inputs are given by the data $\{(Y_i, D_i, Z_i, p(Z_i))\}$ on units $i \in [N] = \{1, \ldots, N\}$. Fix the number of splits $N_S$ and the significance level $\alpha$, e.g. $N_S = 250$ and $\alpha = 0.05$. Fix a set of ML or Causal ML methods.

## 3.9 Application

I apply the methodology to the two datasets, Morocco, and Bosnia and Herzegovina, to estimate and make inferences on the heterogeneous treatment effects of microcredit availability using the R package GenericML[4]. The code to replicate the analysis on the Bosnia and Herzegovina dataset can be found on GitHub[5]. I use the ML methods Random Forest, Elastic Net, Support Vector

[4] https://github.com/mwelz/GenericML

[5] https://github.com/svembden/Thesis_ML_Inference_Morocco_and_Bosnia

**Algorithm 1** Inference Algorithm

1: **Step 1:** Randomly partition the dataset $\{(Y_i, D_i, Z_i)\}_{i=1}^N$ into two parts: a main sample $\text{Data}_M$ and an auxiliary sample $\text{Data}_A$. Repeat this process $N_S$ times.
2: **for** each split $s = 1, 2, \dots, N_S$ **do**
3: **Step 2:** Using the auxiliary sample $\text{Data}_A$, train each machine learning (ML) method and output predictions $B$ and $S$ for the main sample $\text{Data}_M$.
4: **Step 3:** Choose the best ML method or aggregate multiple ML methods.
5: **Step 4:** Estimate the Best Linear Predictor (BLP) parameters.
6: **Step 5:** Estimate the Group Average Treatment Effects (GATES) parameters.
7: **Step 6:** Estimate the Characteristics of Heterogeneous Effects (CLAN) parameters.
8: **Step 7:** Compute the goodness of fit measures in the main sample $\text{Data}_M$.
9: **end for**

Machines, and Boosting in our estimation, making use of the mlr3 package in R. All code is run on a 6 core AMD Radeon CPU. All estimations are run with a propensity score ($p(Z)$) of 0.5 and 100 splits.

# 4 Results

In this section I consider the results of the application of the methods and algorithm developed by Chernozhukov et al. (2020).

## 4.1 Morocco

I first consider the results of the Crépon et al. (2015) study. Similar to Chernozhukov et al. (2017), I find an unconditional average treatment effect of 1127, 5222, 1839, -31 for the outcome variables Amount of Loans, Output, Profit, and Consumption, respectively.

Table 1 shows the results of the first step of Algorithm 1. According to the results, the Random Forest and Elastic Net methods generally perform the best across multiple outcome variables, again similar to the findings of Chernozhukov et al. (2017). Therefore, I will focus on these two methods for further analysis. The difference in the values of $\Lambda$ and $\bar{\Lambda}$ with the original paper is due to the use of different machine learning methods, packages, or arguments.

Table 2 shows the results of the BLP for the CATE using machine learning proxies $S(Z)$ for the four different outcome variables. The table provides estimates for the coefficients $\beta_1$ and $\beta_2$, which represent the Average Treatment Effect (ATE) and the heterogeneity loading (HET) parameters in the BLP model, respectively. Adjusted p-values for the hypothesis that the parameter equals zero are displayed in brackets, with the confidence intervals shown in parentheses. The ATE's are conistent with those of Augsburg et al. (2015).

Table 1: Comparison of ML Methods: Moroccan Microfinance Availability

| | **Elastic Net** | **Boosting** | **SVM** | **Random Forest** |
|---|---|---|---|---|
| **Amount of Loans** | | | | |
| Best BLP ($\Lambda$) | 619,553 | 546,749 | 490,394 | 2,159,715 |
| Best GATES ($\bar{\Lambda}$) | 2,116,754 | 2,013,715 | 2,218,404 | 2,838,066 |
| **Output** | | | | |
| Best BLP ($\Lambda$) | 53,820,920 | 10,758,691 | 35,026,890 | 14,296,479 |
| Best GATES ($\bar{\Lambda}$) | 150,213,345 | 90,597,248 | 94,268,166 | 106,729,646 |
| **Profit** | | | | |
| Best BLP ($\Lambda$) | 8,988,006 | 2,957,196 | 9,076,746 | 14,617,825 |
| Best GATES ($\bar{\Lambda}$) | 20,616,495 | 17,165,554 | 29,702,737 | 35,019,824 |
| **Consumption** | | | | |
| Best BLP ($\Lambda$) | 5,046 | 12,950 | 18,611 | 4,894 |
| Best GATES ($\bar{\Lambda}$) | 36,244 | 27,535 | 36,079 | 31,073 |

Notes: Medians over 100 splits.

Table 2: BLP of Moroccan Microfinance Availability

| | Elastic Net | | Random Forest | |
|---|---|---|---|---|
| | ATE ($\beta_1$) | HET ($\beta_2$) | ATE ($\beta_1$) | HET ($\beta_2$) |
| Amount of Loans | 1,065 | 0.211 | 1,064 | 0.346 |
| | (231, 1,917) | (-0.165, 0.605) | (273, 1,915) | (-0.035, 0.690) |
| | [0.012] | [0.237] | [0.009] | [0.078] |
| Output | 5,292 | 0.256 | 4,584 | 0.106 |
| | (-1,974, 12,693) | (-0.033, 0.537) | (-2,724, 12,040) | (-0.146, 0.340) |
| | [0.153] | [0.077] | [0.209] | [0.372] |
| Profit | 1,473 | 0.281 | 1,413 | 0.184 |
| | (-2,622, 5,658) | (-0.170, 0.672) | (-2,675, 5,564) | (-0.043, 0.427) |
| | [0.486] | [0.170] | [0.482] | [0.103] |
| Consumption | -61.88 | 0.106 | -65.28 | 0.075 |
| | (-203, 91) | (-0.250, 0.442) | (-220, 94) | (-0.247, 0.388) |
| | [0.425] | [0.513] | [0.406] | [0.651] |

Notes: Medians over 100 splits. 90% confidence interval in parenthesis.

P-values for the hypothesis that the parameter is equal to zero in brackets.

All monetary variables are expressed in Moroccan Dirhams, or MAD.

The exchange rate at baseline was US$1 to MAD 8.06.

The results are comparable to the findings of Chernozhukov et al. (2017), the ATE not differing significantly from the results of this paper. However, in these results the ATE of Output does not seem to be significant at the 10% level. This difference could attributed to the use of less intricate machine learning methods, or the fact that our dataset misses some observations compared to those of the original paper. The ATE of the variables Profit and Consumption is not significant at the 10% level. Implying that the treatment effect is not significantly different from zero and drawing the same conclusion as Crépon et al. (2015) that the availability of microcredit does not have a significant effect on the profit and consumption. For the heterogeneity loading parameter, I find two cases where the hypothesis that HET ($\beta_2$) is zero is rejected at the 10% level. For the Amount of Loans the Random Forest method suggests that there is significant heterogeneity in the treatment effect at the 10% level, and for Profit the Elastic Net method suggests that there is significant heterogeneity in the treatment effect at the 10% level. These results are also found in Chernozhukov et al. (2017), but they find that the heterogeneity is also significant for Output and that both machine learning methods suggest that there is significant heterogeneity in the treatment effect for these variables, with the exception for Output. The authors of the original paper suggest that microfinance availability has heterogeneous impacts on business-related outcomes that do not directly affect the business owner's standard of living measured through consumption. Crépon et al. (2015) give as a possible explanation that this is due to a reduction in labor supply when microcredit is available and leads to higher profits.

The GATES offer a more comprehensive understanding of the heterogeneity. In Table 3 I present the results of the GATES for the four different outcome variables, comparing the most and least affected groups. Furthermore, Figures 9-12 in the Appendix present visually the estimated GATES coefficients $\gamma_1 - \gamma_5$ along with joint confidence bands. Analyzing the results in the table, I note that similar to the original analysis in Chernozhukov et al. (2017), I find that the difference in GATES between the most and least affected groups is significantly different from zero at least at the 10% level for the Amount of Loans and Profit according to the Random Forest proxy, and significant at 10% for the Elastic Net proxy for Output, whereas I fail to reject the hypothesis that this difference is zero at conventional levels for all proxies of consumption. These differences in significance level could again be down to the differences in machine learning choices. Furthermore, I also find no significant negative impact on Profit, implying that the availability of microcredit generally at least not negatively impacts the profit of the households, and insignificant negative Consumption for the least affected group. For this fact I present the same explanation the authors give that households decrease their consumption to increase their investment.

In their paper, Chernozhukov et al. (2017) find that village pair fixed effects have a much stronger predictive power for treatment effect heterogeneity compared to baseline household characteristics. Specifically, the R-squares from their regressions indicate that village pair fixed effects account for a substantial portion of the variation in the treatment effects across different outcomes (Amount of Loans, Output, and Profit). For instance, the R-squares for village pair fixed effects are significantly higher (ranging from 0.72 to 0.98 across different methods and outcomes) compared to the baseline household covariates (ranging from 0.08 to 0.35). This suggests that the heterogeneity in treatment effects is more strongly driven by village-level factors and the dynamics of the branch managers rather than by individual household characteristics. For the sake of time, I do not replicate this analysis, but do make mention of this interesting finding and refer to the original paper for further details.

Table 3: GATES of 20% Most and Least Affected Groups

| | Elastic Net | | | Random Forest | | |
|---|---|---|---|---|---|---|
| | 20% Most ($\gamma_5$) | 20% Least ($\gamma_1$) | Difference ($\gamma_5 - \gamma_1$) | 20% Most ($\gamma_5$) | 20% Least ($\gamma_1$) | Difference ($\gamma_5 - \gamma_1$) |
| Amount of Loans | 2021.5 | 425.2 | 1646.7 | 2804.8 | -169.4 | 3002.5 |
| | (101.2, 3966) | (-1446.5, 2356) | (-1219.9, 4458) | (759.9, 4912) | (-2548.5, 2090) | (-21.3, 6085) |
| | [0.0388] | [0.5756] | [0.2517] | [0.0088] | [0.9195] | [0.0513] |
| Output | 22161 | -1155 | 22086 | 18692.3 | 583.5 | 18018.6 |
| | (2900, 42018) | (-13459, 10708) | (-1074, 46080) | (77.1, 39058) | (-15794.6, 15580) | (-7523.6, 43616) |
| | [0.0243] | [0.7966] | [0.0609] | [0.0485] | [0.8872] | [0.1707] |
| Profit | 5826.1 | 56.2 | 5853.5 | 11183.9 | -975.3 | 11637.8 |
| | (-3068.9, 14997) | (-8072.4, 9593) | (-6554.4, 18862) | (-538.8, 21851) | (-10687.4, 6945) | (-1669.2, 25621) |
| | [0.204] | [0.967] | [0.338] | [0.0594] | [0.8052] | [0.0796] |
| Consumption | 17.1 | -309.5 | 303.6 | -7.0 | -250.3 | 213.3 |
| | (-288.9, 300.3) | (-745.0, 129.5) | (-298.9, 849.7) | (-377.7, 363.7) | (-649.1, 137.3) | (-387.2, 799.9) |
| | [0.834] | [0.175] | [0.293] | [0.974] | [0.200] | [0.474] |

Notes: Medians over 100 splits. 90% confidence interval in parenthesis.
P-values for the hypothesis that the parameter is equal to zero in brackets.

Lastly, I examine what drives this heterogeneity in the data using CLAN. I only consider the variables for which I have found significant heterogeneity. As mentioned before, it is found that these baseline characteristics do not have as much predictive power as the village pair fixed effects. Still, as their contribution to the heterogeneity is not zero, I consider them in my analysis. The results of the CLAN are presented in Table 4. The CLAN for the 20% least and most affected groups defined by the quantiles of the CATE proxy $S(Z)$, as well as the difference between the two are shown. I show the three baseline covariates per outcome variable that are most correlated with the CATE. I find that households with younger heads, fewer members, and an agricultural related self employment activity are more likely to borrow. As Chernozhukov et al. (2017) I also find that households with fewer adult members are more likely to borrow, but this is omitted for brevity. For profit, I find that households with a non animal husbandry self-employment activity and fewer households members are more likely to have a higher profit, noting that the difference in head age is not significant by any of the two learners. This is similar to the findings of Crépon et al. (2015), which indicate that access to microcredit resulted in a notable rise in profit for self-employment activities, primarily animal husbandry and agriculture. One possible explanation for these findings is that households with fewer members may have a higher dependency on individual contributions to the household income, making them more likely to take risks such as borrowing to invest in self-employment activities. Additionally, younger household heads might be more inclined to innovate. The increase in profit related to above self-employment activities could be because for these households, for instance in agricultural activities, access to credit might allow for the purchase of better seeds, fertilizers, or equipment, leading to higher yields and, consequently, increased profits.

Table 4: CLAN of Microfinance Availability

| | Elastic Net | | | Random Forest | | |
|---|---|---|---|---|---|---|
| | 20% Most ($\delta_5$) | 20% Least ($\delta_1$) | Difference ($\delta_5 - \delta_1$) | 20% Most ($\delta_5$) | 20% Least ($\delta_1$) | Difference ($\delta_5$ - $\delta_1$) |
| **Amount of Loans** | | | | | | |
| Head Age | 33.655 | 39.502 | -6.428 | 25.520 | 36.570 | -11.490 |
| | (31.446, 35.863) | (37.498, 41.506) | (-9.443, -3.465) | (23.200, 27.778) | (34.600, 38.470) | (-14.560, -8.437) |
| | - | - | [0.000] | - | - | [0.000] |
| Number of Household Members | 3.583 | 4.021 | -0.478 | 2.773 | 4.170 | -1.351 |
| | (3.313, 3.853) | (3.756, 4.294) | (-0.855, -0.091) | (2.472, 3.073) | (3.916, 4.425) | (-1.743, -0.960) |
| | - | - | [0.016] | - | - | [0.000] |
| Non-agricultural self-emp. | 0.039 | 0.148 | -0.127 | 0.027 | 0.096 | -0.072 |
| | (0.023, 0.055) | (0.118, 0.178) | (-0.160, -0.091) | (0.013, 0.041) | (0.070, 0.119) | (-0.099, -0.044) |
| | - | - | [0.000] | - | - | [0.000] |
| **Profit** | | | | | | |
| Animal Husbandry self-emp. | 0.381 | 0.447 | -0.072 | 0.379 | 0.473 | -0.095 |
| | (0.339, 0.421) | (0.401, 0.489) | (-0.139, -0.014) | (0.338, 0.419) | (0.431, 0.515) | (-0.153, -0.036) |
| | - | - | [0.0152] | - | - | [0.001] |
| Head Age | 35.495 | 36.248 | -0.808 | 31.105 | 33.727 | -2.098 |
| | (33.468, 37.610) | (34.053, 38.410) | (-3.857, 2.210) | (28.905, 33.294) | (31.317, 36.137) | (-5.341, 1.107) |
| | - | - | [0.596] | - | - | [0.202] |
| Number of Household Members | 3.822 | 3.836 | -0.049 | 3.289 | 3.702 | -0.445 |
| | (3.555, 4.071) | (3.546, 4.120) | (-0.424, 0.334) | (2.998, 3.577) | (3.415, 3.990) | (-0.859, -0.040) |
| | - | - | [0.803] | - | - | [0.031] |

Notes: Medians over 100 splits. 90% confidence interval in parenthesis.
P-values for the hypothesis that the parameter is equal to zero in brackets.

## 4.2 Bosnia and Herzegovina

Next, I consider the results of the analysis of the Bosnia and Herzegovina dataset. The results are divided into five subsections corresponding to the grouped outcome variables.

### 4.2.1 Credit Outstanding at Endline

The scores for the Credit Outstanding at Endline variables are presented in Table 5. The definition of the outcome variables is given in the Appendix. The low $\Lambda$ scores for are explained by the outcome variables being binary, considering the low variation in these variables. Amount of Loans also exhibits low variation (variation of 1).

The results of the BLP for the CATE using machine learning proxies for the Credit Outstanding at Endline variables are presented in Table 6. For all outcome variables except Outstanding loan bank, the ATE is significant. The higher number of outstanding loans for the treatment group indicates improved access to liquidity. This improvement likely stems from the treatment group's ability to secure more funding and/or loans with longer maturities. Therefore, as Augsburg et al. (2015), I conclude that the treatment group had significantly better access to liquidity compared to the control group. I find no significant heterogeneity in any of the outcome variables.

Table 5: Comparison of ML Methods: Credit Outstanding at Endline

| | **Elastic Net** | **Boosting** | **SVM** | **Random Forest** |
|---|---|---|---|---|
| **Loan** | | | | |
| Best BLP ($\Lambda$) | 0.675 | 0.650 | 1.427 | 1.984 |
| Best GATES ($\bar{\Lambda}$) | 43.460 | 43.040 | 42.330 | 42.220 |
| **Amount of Loans** | | | | |
| Best BLP ($\Lambda$) | 3.015 | 4.026 | 2.467 | 3.928 |
| Best GATES ($\bar{\Lambda}$) | 223.500 | 222.600 | 217.700 | 228.100 |
| **Outstanding loan MFI** | | | | |
| Best BLP ($\Lambda$) | 0.554 | 1.115 | 2.121 | 2.185 |
| Best GATES ($\bar{\Lambda}$) | 202.000 | 204.000 | 204.500 | 204.300 |
| **Outstanding loan bank** | | | | |
| Best BLP ($\Lambda$) | 0.188 | 0.206 | 0.168 | 0.179 |
| Best GATES ($\bar{\Lambda}$) | 5.106 | 5.016 | 4.969 | 4.727 |

Notes: Medians over 100 splits.

Actual values are $\times 10^{-3}$.

Table 6: BLP of Credit Outstanding at Endline

| | SVM | | Random Forest | |
|---|---|---|---|---|
| | ATE ($\beta_1$) | HET ($\beta_2$) | ATE ($\beta_1$) | HET ($\beta_2$) |
| Loan | 0.191 | 0.185 | 0.191 | 0.264 |
| | (0.122, 0.262) | (-0.177, 0.518) | (0.120, 0.263) | (-0.165, 0.692) |
| | [0.000] | [0.303] | [0.000] | [0.206] |
| Amount of Loans | 0.431 | -0.061 | 0.434 | 0.094 |
| | (0.251, 0.614) | (-0.323, 0.216) | (0.256, 0.616) | (-0.357, 0.554) |
| | [0.000] | [0.690] | [0.000] | [0.668] |
| Outstanding loan MFI | 0.444 | 0.166 | 0.444 | 0.236 |
| | (0.365, 0.523) | (-0.131, 0.461) | (0.365, 0.523) | (-0.181, 0.680) |
| | [0.000] | [0.262] | [0.000] | [0.249] |
| Outstanding loan bank | -0.057 | 0.035 | -0.057 | -0.065 |
| | (-0.101, -0.012) | (-0.724, 0.766) | (-0.100, -0.013) | (-0.567, 0.419) |
| | [0.013] | [0.910] | [0.012] | [0.803] |

Notes: Medians over 100 splits. 90% confidence interval in parenthesis.

P-values for the hypothesis that the parameter is equal to zero in brackets.

Finally, I show the GATES results in Table 7. Here the results of the best learner per variable is shown. The results show that the difference in GATES between the most and least affected groups is not significantly different from zero. I only show the CLAN for the Amount of Loans and Outstanding loan MFI in the Appendix because of the lack of significant heterogeneity in the treatment effects. This is done for the top 3 most correlated variables with the CATE. The rest of this paper follows this approach of presentation too. I note that households with more employed members, less retired members, and less members aged 16-19 are significantly more likely to have a higher number of loans - possibly daring to take on more risk.

Table 7: GATES of Credit Outstanding at Endline

| | 20% Most ($\gamma_5$) | 20% Least ($\gamma_1$) | Difference ($\gamma_5 - \gamma_1$) |
|---|---|---|---|
| | Elastic Net | | |
| Loan | 0.186 | 0.192 | -0.005 |
| | (0.030, 0.341) | (0.035, 0.351) | (-0.224, 0.219) |
| | [0.018] | [0.016] | [0.965] |
| | Random Forest | | |
| Amount of Loans | 0.443 | 0.357 | 0.077 |
| | (0.068, 0.812) | (-0.048, 0.779) | (-0.504, 0.647) |
| | [0.020] | [0.082] | [0.769] |
| | SVM | | |
| Outstanding loan MFI | 0.483 | 0.397 | 0.085 |
| | (0.309, 0.658) | (0.215, 0.577) | (-0.169, 0.332) |
| | [0.000] | [0.000] | [0.497] |
| | Elastic Net | | |
| Outstanding loan bank | -0.057 | -0.052 | 0.007 |
| | (-0.156, 0.037) | (-0.151, 0.044) | (-0.126, 0.140) |
| | [0.230] | [0.284] | [0.885] |

Notes: Medians over 100 splits. 90% confidence interval in parenthesis.
P-values for the hypothesis that the parameter is equal to zero in brackets.

#### 4.2.2 Self-Employment Activities

Table 14 in the Appendix shows the $\Lambda$ scores for the Self-Employment Activities. I note again low scores for the binary outcome variables. I take a closer look at the BLP in Table 15 in the Appendix. For these outcome variables, I find no significant ATE and HET, except a significant ATE at the 10% level for Inventory. Meaning the treated group is more likely to hold inventory compared to the control group. I note that Asset Value decreases for the treatment group, this is consistent with the findings of Augsburg et al. (2015) better presented in their online Appendix, but contrasting to Crépon et al. (2015), who find that asset value increases for treated households. This could be down to how the authors have defined the asset value, the latter including the stock of livestock as well, but this is not made clear enough to draw a conclusion. Interestingly, similar to Augsburg et al. (2015), no significant ATE is discovered for the business related variables. The authors discuss whether this 14-month period is too short to see any significant impact here, but conclude that that is not the case. One possible explanation for this low significance is the low number of observations in the dataset, Profit as an example containing only 329 observations, split into 196 treated and 133 control observations. This could lead to a lack of statistical power to detect significant treatment effects. Unfortunately, I do not have access to more observations, so I cannot draw any further conclusions. The authors do perform a trimming for the top 1% of profits and then find significance, therefore I present GATES in Table 16 in the Appendix and show the GATES for Asset Value and Business Profit in Figures 1 and 2. Interestingly enough, I do see that the most affected quantile does in fact have the highest profit, but that the group before the most affected would have negative profit impact. The least affected group also does contain the lowest profit, but due to the low amount of observations, different from the original authors, I cannot safely say that there has been a positive impact on profit by access to microcredit. Furthermore, I note that the most affected group had less Business Expenses than the other groups, and that Asset Value actually increased for the least affected group, decreasing as we reach more affected groups. This could give a better explanation as to the different findings by both previously mentioned studies about the increment of asset value for the treated group.

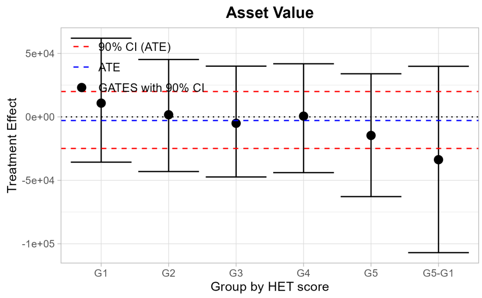


Figure 1: GATES of Asset Value

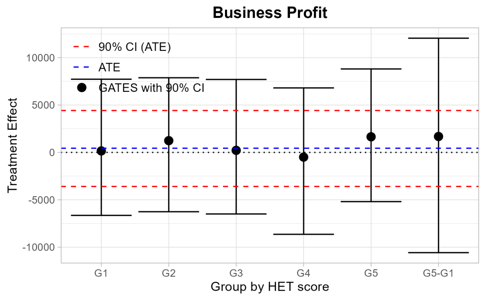


Figure 2: GATES of Business Profit

#### 4.2.3 Income

The $\Lambda$ scores for the Income outcome variables are presented in Table 17 in the Appendix. Looking at the BLP results also in the Appendix, Table 18 I find no significant ATE or HET for any of the outcome variables. Here I find the only difference with the analysis by the authors. They find that the likelihood of receiving income from wages is significant at the 5% level, indicating a shift in activity towards self-employment and away from wage employment, similar to the findings of Crépon et al. (2015). However, because of the lack of significance I cannot draw the same conclusion.

Concerning GATES, presented in the Appendix, I don't find any significant differences or interesting patterns between the groups, but do find a small decline in the income received from wages between the least and most affected shown in Figure 4. Also included in Figure 3 is the Wages Likelihood. While not significant, this could be an indication that the availability of microcredit has a negative impact on the income received from wages, due to the fact that the availability of microcredit allows individuals to start their own business, or work more on their self-employment activities, which would be consistent with the findings mentioned above. Lastly, from CLAN (Table 20 in the Appendix) I find that for the amount of remittance received, households who are more likely to own a dwelling and a business, and of which the respondent is less likely to be a female receive more remittance. Family members abroad may prefer to send money to households where they feel is more collateral present.

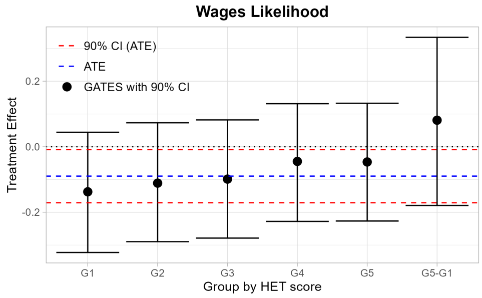


Figure 3: GATES of Wages Likelihood

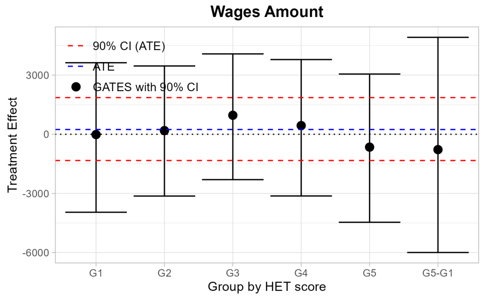


Figure 4: GATES of Wages Amount

#### 4.2.4 Hours Worked

As mentioned in Section 2, for the outcome variables related to the number of hours worked, I only focus on the hours worked for all adults and teens (aged 16 to 64). The $\Lambda$ scores for the Hours Worked are presented in Table 21 in the Appendix. The results of the BLP are also there, presented in Table 22. Similar to Augsburg et al. (2015), I find no significant ATE for any of the outcome variables. While not significant I do note that the hours worked in the business increased, while the hours spent on other activities decreased, perhaps because starting a business replaced other work activities. This same finding is present in Augsburg et al. (2015). Furthermore, as is evident from the table there is no significant heterogeneity in the treatment effects for any of the outcome variables. Lastly, GATES is presented in the Appendix, Table 23, I note that the least affected group spends considerably more hours on business, and secondly that the most affected group works a sizeable amount less than the least affected group (about 18 hours) and also about 8 hours less than the second most affected group. Both visualized in Figures 5 and 6. It seems that those most effected by the availability of microcredit reduce their labor supply as is consistent with other studies, Crépon et al. (2015). Interestingly, for the total hours worked, see Table 24 in the Appendix, the group that is most affected generally have younger children and are older. This age difference is most notable for the number of hours worked on other activities. This makes sense if you consider that younger children take more supervision and care, and older people on average might have more and different priorities than younger people.

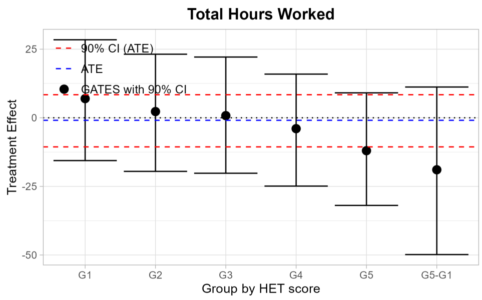


Figure 5: GATES of Total Hours Worked

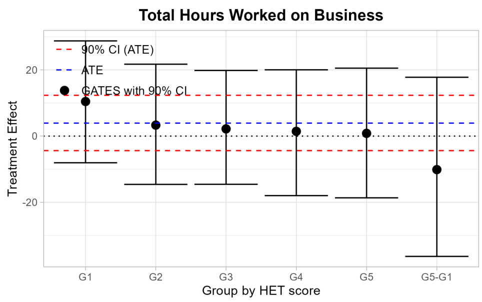


Figure 6: GATES of Total Hours Worked on Business

### 4.2.5 Consumption and Savings

Finally, I consider the heterogeneity in the impact on Consumption and Savings. Again, I present the BLP scores in Table 25 in the Appendix. Concerning the ATE and HET for the variables it is shown in the Appendix in Table 26 that the ATE is significant at the no more than the 10% significance level for the Home durable good index (HDGI) and the Savings. The HDGI is an index created by the authors for a list of ownership of 18 home durable goods where each asset is given a weight using the coefficients of the first factor of a principal component analysis. More information can be found in the variable definitions in the Appendix, Table 10. Similar to Crépon et al. (2015) and other studies, there is a negative impact on Consumption.

Furthermore, unlike the authors of Augsburg et al. (2015) who find no significance in any of the outcome variables, I do for the HDGI and Savings. For the latter the authors give as a possible explanation that when a loan becomes accessible, it may enable a profitable investment by combining the loan with household savings. So therefore, the impact on Savings could, like Consumption, be negative. I do find significance at the 10% level to support this. This difference in significance level could very well be down to the fact there might be heterogeneity present, as savings is influenced by a variety of household-specific factors such as income variability, risk preferences, and financial literacy. This could be supported by the fact that the $\beta_2$ parameter is close to significant at the 10% level. The same theory could possibly be said for the HDGI. In fact, looking at the GATES for the variable Savings in Figure 7 and Table 27 in the Appendix, I do note stark differences between the least and most affected group, the same goes for HDGI presented in Figure 8. Therefore, I look further into these variables and present the CLAN in Table 28. Also, by GATES I see a decrease in Durables consumption for the more affected group. Finally, the authors find that for individual commodities food consumption declined among the lower educated differing from that of the higher educated (p-value of 0.02). I however find, while the education level is most correlated with the CATE, no significant differences between the groups.

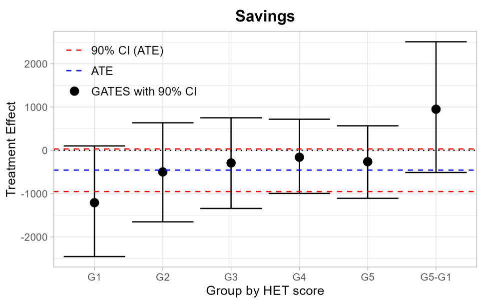


Figure 7: GATES of Savings

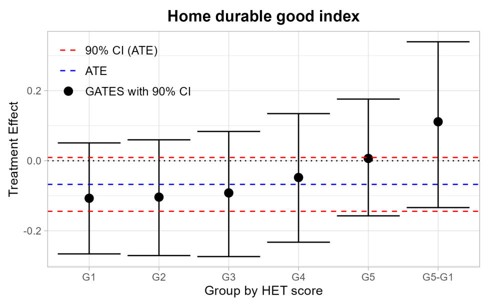


Figure 8: GATES of HDGI

In Table 28 I show the top 3 most correlated variables with the CATE for Savings and HDGI. For both outcome variables, the differences are significant for all three variables. I find that households where predominately the respondent was female, fewer number of employed household members and fewer household members attending school were more likely to spend savings. This could be because woman were less able to retrieve as large of loans as men, which even in the US is still a common issue[6]. Additionally, with fewer sources of income, households may need to dip into savings to cover basic expenses or households may prioritize spending their savings on expanding the business rather than having members attend education. For the HDGI, I note that households who own their house rather than rent, have fewer female members and fewer members attending school are more likely to have a lower HDGI. Households that own their homes might rather allocate resources towards maintenance, or in general be less dependent on durable goods.

Table 8: CLAN of Asset Value and Business in Services

| | 20% Most ($\delta_5$) | 20% Least ($\delta_1$) | Difference ($\delta_5 - \delta_1$) |
|---|---|---|---|
| | | XGBoost | |
| **Asset Value** | | | |
| Number of hh members attending school | 0.690 | 0.980 | -0.280 |
| | (0.509, 0.882) | (0.775, 1.188) | (-0.537, 0.002) |
| | - | - | [0.052] |
| Number of female hh members | 1.655 | 1.860 | -0.190 |
| | (1.474, 1.853) | (1.659, 2.068) | (-0.473, 0.094) |
| | - | - | [0.190] |
| Respondent age | 37.100 | 39.450 | -2.530 |
| | (34.460, 39.692) | (37.030, 41.922) | (-6.040, 1.057) |
| | - | - | [0.166] |
| | | XGBoost | |
| **Business in Services** | | | |
| Highest respondent grade is a university level | 0.075 | 0.025 | 0.055 |
| | (0.020, 0.120) | (-0.008, 0.048) | (-0.001, 0.120) |
| | - | - | [0.051] |
| Dwelling ownership | 0.825 | 0.885 | -0.065 |
| | (0.744, 0.896) | (0.816, 0.944) | (-0.156, 0.033) |
| | - | - | [0.208] |
| House ownership | 0.795 | 0.845 | -0.055 |
| | (0.710, 0.870) | (0.768, 0.912) | (-0.158, 0.051) |
| | - | - | [0.353] |

Notes: Medians over 100 splits. 90% confidence interval in parenthesis.
P-values for the hypothesis that the parameter is equal to zero in brackets.

#### 4.2.6 Discussion of Significance

In general, for the Bosnian and Herzegovina dataset I do not find significant heterogeneity. It could very well be possible that there is none, but a few other reasons could also attribute to this lack of significance. Firstly, a lot of the outcome variables are binary and exhibit low variation, which could impact the performance and reliability of ML methods. Furthermore, when there is little variation within a variable, it can be challenging to identify significant differences between treatment and control groups or within subgroups. Additionally, the dataset has a limited number of observations for certain variables, such as Profit, which contains only 329 observations split between the treated and control groups. This small sample size reduces the statistical power to detect significant treatment effects and heterogeneity. In addition to

[6]https://attorney-newyork.com/business-loan-statistics/

this, some variables like Profit contain some outliers, - see the Appendix, Figures 13 and 14 for a histogram and box plot - which could affect the significance. Lastly, while the authors of Augsburg et al. (2015) address this for the variable Profit, the study duration of 14 months may be insufficient to observe significant impacts for some other certain business-related variables. Some effects, especially in business and economic activities, might require a longer period to manifest significantly.

# 5 Conclusion

This study investigated the heterogeneous effects of microcredit on various economic outcomes by analyzing data from randomized control trials (RCTs) conducted in Morocco and Bosnia and Herzegovina. Using machine learning based techniques developed by Chernozhukov et al. (2020), I sought to understand the differential impacts of microcredit on households, focusing on metrics such as loan amounts, profits, consumption, and savings.

The analysis of the Moroccan dataset reveals that microcredit increases the amount of loans and output, though the impact on profits and consumption remains modest. The presence of significant heterogeneity suggests that these effects vary among different demographic groups. Notably, the Random Forest method identified significant heterogeneity in the treatment effect on the amount of loans and profits, emphasizing the nuanced impact of microfinance initiatives. It is shown however, that this heterogeneity is not namely caused by individual baseline characteristics, but rather by village pair fixed effects. In Bosnia and Herzegovina, the impact of microcredit was similarly varied. While I observed a significant positive effect in the treatment for numerous outcome variables, the presence of heterogeneity is not significant. A possible explanation for the lack of heterogeneity in these outcome variables could be due to the fact that the microcredit availability in Bosnia and Herzegovina was targeted at a marginal segment of the population, individuals who would typically be rejected for loans but might be considered if slightly more risk was accepted. This could mean that the treatment effect is more homogeneous across the population. The uniform effect indicates that these households shared similar economic conditions or reacted similarly to the availability of credit, suggesting that microfinance programs designed for specific, narrowly-defined groups might result in more uniform outcomes. This is analogous with Angelucci et al. (2015) who also find no significant heterogeneity at the community level from a group lending expansion, where only for revenues, profits, and household decision-making power stronger effects at the upper end of the distribution are found.

Still despite this, our study underscores the importance of considering heterogeneity when evaluating microfinance programs. The differential impacts observed suggest that microcredit can be a valuable tool for economic empowerment, particularly when targeted effectively. However, the moderate effects on for example consumption and the lack of significant impact on some business-related outcomes indicate that microcredit alone may not be sufficient to drive substantial economic transformation for all recipients.

Future research should explore the more long-term effects of microcredit and investigate additional variables that may influence its effectiveness. Moreover, policymakers should account for potential heterogeneous effects when designing and implementing microfinance programs.

## References


Alemu, A. and Ganewo, Z. (2023). Impact analysis of formal microcredit on income of borrowers in rural areas of sidama region, ethiopia: A propensity score matching approach. *Journal of the Knowledge Economy*, 14(1):65–85.

Angelucci, M., Karlan, D., and Zinman, J. (2015). Microcredit impacts: Evidence from a randomized microcredit program placement experiment by compartamos banco. *American Economic Journal: Applied Economics*, 7(1):151–182.

Armendáriz de Aghion, B. and Morduch, J. (2005). Microfinance: A critical literature survey.

Aslam, M., Kumar, S., and Sorooshian, S. (2020). Impact of microfinance on poverty: Qualitative analysis for grameen bank borrowers. *International Journal of Financial Research*, 11(1):49–59.

Attanasio, O., Augsburg, B., De Haas, R., Fitzsimons, E., and Harmgart, H. (2015). The impacts of microfinance: Evidence from joint-liability lending in mongolia. *American Economic Journal: Applied Economics*, 7(1):90–122.

Augsburg, B., De Haas, R., Harmgart, H., and Meghir, C. (2015). The impacts of microcredit: Evidence from bosnia and herzegovina. *American Economic Journal: Applied Economics*, 7(1):183–203.

Banerjee, A., Duflo, E., Glennerster, R., and Kinnan, C. (2015). The miracle of microfinance? evidence from a randomized evaluation. *American Economic Journal: Applied Economics*, 7(1):22–53.

Chernozhukov, V., Demirer, M., Duflo, E., and Fernández-Val, I. (2017). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in india.

Chernozhukov, V., Demirer, M., Duflo, E., and Fernández-Val, I. (2020). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in india.

Chernozhukov, V., Fernandez-Val, I., and Luo, Y. (2018). The sorted effects method: Discovering heterogeneous effects beyond their averages.

Christensen, R., Bours, M. J., and Nielsen, S. M. (2021). Effect modifiers and statistical tests for interaction in randomized trials. *Journal of Clinical Epidemiology*, 134:174–177.

Crépon, B., Devoto, F., Duflo, E., and Parienté, W. (2015). 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):123–50.

Dhungana, B. R., Timlasina, A., Thapa, B., and Tiwari, D. R. (2022). Effects of microfinance services on economic status improvement: A case of skbbl, pokhara, nepal. *Quest Journal of Management and Social Sciences*, 4(2):200–209.

Kolstad, I., Garcia Pires, A. J., and Wiig, A. (2017). Within-group heterogeneity and group dynamics: Analyzing exit of microcredit groups in angola. *Oxford Development Studies*, 45(3):338–351.

# A Variable Description

Table 9: Description of Variables: Morocco

| | Variable | Description |
|---|---|---|
| **Baseline Covariates** | ccm_resp_activ | =1 if spouse of head responded to self-employment section |
| | other_resp_activ | =1 if other member responded to self-employment section |
| | borrowed_total_bl | =1 if borrowed from any source |
| | act_livestock_bl | =1 if declared animal husbandry self-employment activity |
| | act_business_bl | =1 if declared non-agricultural self-employment activity |
| | members_resid_bl | Number of hh members |
| | nadults_resid_bl | Number of members 16 years old or older |
| | head_age_bl | Head age |
| | head_age_d_bl | =1 if head_age missing |
| | members_resid_d_bl | =1 if missing obs at baseline |
| | nadults_resid_d_bl | =1 if missing obs at baseline |
| | act_livestock_d_bl | =1 if missing obs at baseline |
| | act_business_d_bl | =1 if missing obs at baseline |
| | borrowed_total_d_bl | =1 if missing obs at baseline |
| | ccm_resp_activ_d | =1 if ccm_resp_activ missing |
| | other_resp_activ_d | =1 if other_resp_activ missing |
| **Outcome Variables** | Amount of Loans | Total Amount of Loans |
| | Output | Total output from self-employment activities (past 12 months) |
| | Profit | Total profit from self-employment activities (past 12 months) |
| | Consumption | Total monthly consumption |

Table 10: Description of Variables: Bosnia and Herzegovina

| | Variable | Description |
|---|---|---|
| **Baseline Covariates** | b_resp_female | =1 if respondent is female, 0 male |
| | b_resp_age | Age of respondent |
| | b_resp_ms | Marital status of respondent |
| | b_resp_es | Economic status of respondent |
| | b_resp_ss | =1 if highest respondent grade is a secondary school grade (incl vocational) |
| | b_resp_ul | =1 if highest respondent grade is a university level |
| | b_resp_school | Respondent currently attending school, 0/1 |
| | b_hhmem_female | Number of female household members |
| | b_kids_05 | Number of kids 5yrs or younger |
| | b_kids_610 | Number of kids older than 5 and younger than 11 |
| | b_kids_1116 | Number of kids older than 10 and younger than 17 |
| | b_kids_1619 | Number of kids age 16-19 |
| | b_hhmem_school | Number of household members attending school |
| | b_hhmem_employed | Number of employed household members |
| | b_hhmem_retired | Number of retired household members |
| | b_dw_house | =1 if primary dwelling is a house |
| | b_dw_own | =1 if primary dwelling is owned |
| | b_bm_own | =1 if respondent owns a business |
| **Outcome variables** | Loan | =1 if respondent has any loan outstanding |
| | Amount of Loans | Number of loans outstanding |
| | Outstanding loan MFI | =1 if household has any loan outstanding from an MFI at endline |
| | Outstanding loan bank | =1 if household has any loan outstanding from a bank at endline |
| | Asset value | Current market value (at time of analysis) of the household's assets (in BAM) |
| | Ownership of Inventory | =1 if the household owns inventory |
| | Any Self-employment Income | =1 if the respondent gets income from self-employment |
| | Business Ownership | =1 if the respondent owns a business |
| | Business in Services | =1 if the respondent owns a business in services |
| | Business in Agriculture | =1 if the respondent owns a business in agriculture |
| | Has Started a Business in Last 14 Months | =1 if the household started a business since the baseline survey |
| | Has Closed a Business in Last 14 Months | =1 if the household closed a business since the baseline survey |
| | Business Profit | Amount of profit (BAM) from the respondent's main business |
| | Business Revenue | Amount of revenue (BAM) from the respondent's business |
| | Business Expenses | Amount of expenses (BAM) made by the respondent's main business |
| | Self-Employment | =1 if at least one household member is self-employed |
| | Self-Employment Amount | Amount of profit (BAM) from self-employment for the entire household |
| | Wages | =1 if at least one household member receives income from salaried work |
| | Wages Amount | Total salary (BAM) received from wages |
| | Remittance | =1 if at least one household member receives remittance |
| | Remittance Amount | Total (BAM) received from remittances |
| | Government Benefits | =1 if at least one household member receives benefits from the government (e.g. welfare) |
| | Government Benefits Amount | Total (BAM) received from benefits |
| | Total Hours Worked | Number of hours worked by all household members in the last week aged 16-64 |
| | Total Hours Worked on Business | Number of hours worked on the business by all household members in the last week aged 16-64 |
| | Total Hours Worked on Other Activities | Number of hours worked on other activities by all household members in the last week aged 16-64 |
| | Total consumption per capita | Total yearly expenditures (BAM) of the household per household member |
| | Durables | Expenditures (BAM) on durable items in the last 12 months |
| | Nondurable | Expenditures (BAM) on non-durable items in the last month |
| | Food | Amount (BAM) spent on food (inside and outside the house) by the household in the last week |
| | Education | Expenditures (BAM) on education in the last month |
| | Cigarettes and alcohol | Amount (BAM) spent on cigarettes and alcohol by the household in the last week |
| | Recreation | Expenditures (BAM) on recreation in the last month |
| | Home durable good index | Index calculated for a list of 18 home durable goods (stock, not flow variable). Each asset is given a weight using the coefficients of the first factor of a principal component analysis. The index, for a household i, is calculated as the weighted sum of standardized dummies equal to 1 if the household owns the durable good. |
| | Savings | Total savings of the household. Savings data was collected in ranges and to calculate average savings allocated the midpoint of indicated ranges to the households. |

# B Summary Statistics

Table 11: Summary Statistics: Morocco

| | Observations | All | Treated | Control |
|---|---|---|---|---|
| **Baseline Covariates** | | | | |
| ccm_resp_activ | 5513 | 0.067 | 0.074 | 0.061 |
| other_resp_activ | 5513 | 0.044 | 0.047 | 0.041 |
| borrowed_total_bl | 5513 | 0.211 | 0.225 | 0.197 |
| act_livestock_bl | 5513 | 0.415 | 0.426 | 0.404 |
| act_business_bl | 5513 | 0.146 | 0.128 | 0.163 |
| members_resid_bl | 5513 | 3.879 | 3.870 | 3.887 |
| nadults_resid_bl | 5513 | 2.604 | 2.600 | 2.608 |
| head_age_bl | 5513 | 35.950 | 35.900 | 36.010 |
| head_age_d_bl | 5513 | 0.257 | 0.266 | 0.248 |
| members_resid_d_bl | 5513 | 0.257 | 0.266 | 0.248 |
| nadults_resid_d_bl | 5513 | 0.257 | 0.266 | 0.248 |
| act_livestock_d_bl | 5513 | 0.257 | 0.266 | 0.248 |
| act_business_d_bl | 5513 | 0.257 | 0.266 | 0.248 |
| borrowed_total_d_bl | 5513 | 0.257 | 0.266 | 0.248 |
| ccm_resp_activ_d | 5513 | 0.134 | 0.131 | 0.136 |
| other_resp_activ_d | 5513 | 0.134 | 0.131 | 0.136 |
| **Outcome variables** | | | | |
| Amount of Loans | 5513 | 2364 | 2934 | 1807 |
| Output | 5513 | 32559 | 35198.9 | 29976 |
| Profit | 5513 | 10120 | 11050.4 | 9210 |
| Consumption | 5513 | 3011.9 | 2996 | 3027 |

All monetary variables are expressed in Moroccan Dirhams, or MAD. The exchange rate at baseline was US$1 to MAD 8.06.

Table 12: Summary Statistics: Bosnia and Herzegovina

| | Observations | All | Treated | Control |
|---|---|---|---|---|
| **Baseline Covariates** | | | | |
| b_resp_female | 994 | 0.410 | 0.414 | 0.404 |
| b_resp_age | 994 | 37.810 | 38.390 | 37.100 |
| b_resp_ms | 994 | 2.146 | 2.172 | 2.113 |
| b_resp_es | 994 | 1.826 | 1.842 | 1.806 |
| b_resp_ss | 994 | 0.624 | 0.615 | 0.634 |
| b_resp_ul | 994 | 0.045 | 0.042 | 0.050 |
| b_resp_school | 994 | 0.028 | 0.033 | 0.023 |
| b_hhmem_female | 994 | 1.717 | 1.721 | 1.713 |
| b_kids_05 | 994 | 0.285 | 0.269 | 0.305 |
| b_kids_610 | 994 | 0.287 | 0.290 | 0.282 |
| b_kids_1116 | 994 | 0.355 | 0.407 | 0.291 |
| b_kids_1619 | 994 | 0.263 | 0.283 | 0.237 |
| b_hhmem_school | 994 | 0.805 | 0.869 | 0.725 |
| b_hhmem_employed | 994 | 1.138 | 1.169 | 1.099 |
| b_hhmem_retired | 994 | 0.312 | 0.312 | 0.312 |
| b_dw_house | 994 | 0.838 | 0.835 | 0.842 |
| b_dw_own | 994 | 0.877 | 0.893 | 0.858 |
| b_bm_own | 994 | 0.623 | 0.632 | 0.612 |
| **Outcome Variables** | | | | |
| Loan | 994 | 0.805 | 0.893 | 0.695 |
| Amount of Loans | 994 | 1.311 | 1.504 | 1.070 |
| Outstanding loan MFI | 994 | 0.571 | 0.768 | 0.325 |
| Outstanding loan bank | 994 | 0.064 | 0.040 | 0.095 |
| Asset value | 994 | 109041 | 107282 | 111229 |
| Ownership of Inventory | 994 | 0.120 | 0.142 | 0.093 |
| Any Self-employment Income | 994 | 0.701 | 0.728 | 0.668 |
| Business Ownership | 994 | 0.542 | 0.572 | 0.506 |
| Business in Services | 994 | 0.188 | 0.203 | 0.169 |
| Business in Agriculture | 994 | 0.259 | 0.276 | 0.237 |
| Has Started a Business in Last 14 Months | 994 | 0.136 | 0.145 | 0.124 |
| Has Closed a Business in Last 14 Months | 994 | 0.216 | 0.205 | 0.230 |
| Business Profit | 329 | 9887 | 10035 | 9669 |
| Business Revenue | 326 | 15644 | 16239 | 14770 |
| Business Expenses | 324 | 6053 | 6334 | 5639 |
| Self-Employment | 994 | 0.701 | 0.728 | 0.728 |
| Self-Employment Amount | 994 | 6152 | 6177 | 6177 |
| Wages | 994 | 0.651 | 0.615 | 0.615 |
| Wages Amount | 994 | 7037 | 7149 | 7149 |
| Remittance | 994 | 0.212 | 0.202 | 0.202 |
| Remittance Amount | 994 | 583 | 575 | 575 |
| Government Benefits | 994 | 0.302 | 0.281 | 0.281 |
| Government Benefits Amount | 994 | 546 | 489 | 489 |
| Total Hours Worked | 994 | 77.780 | 77.590 | 78.000 |
| Total Hours Worked on Business | 994 | 40.800 | 42.580 | 38.600 |
| Total Hours Worked on Other Activities | 994 | 36.970 | 35.010 | 39.400 |
| Total consumption per capita | 994 | 3788 | 3483 | 4167 |
| Durables | 994 | 2212 | 2204 | 2221 |
| Nondurable | 993 | 188.400 | 182 | 196.400 |
| Food | 994 | 115.600 | 114.100 | 117.500 |
| Education | 994 | 416.900 | 390.800 | 449.400 |
| Cigarettes and alcohol | 994 | 13.780 | 12.620 | 15.230 |
| Recreation | 994 | 43.970 | 40.230 | 48.620 |
| Home durable good index | 994 | 0.453 | 0.421 | 0.492 |
| Savings | 994 | 943.700 | 743.200 | 1193 |

# C Credit Outstanding at Endline

Table 13: CLAN of Amount of Loans and Outstanding Loan MFI

| | 20% Most ($\delta_5$) | 20% Least ($\delta_1$) | Difference ($\delta_5 - \delta_1$) |
|---|---|---|---|
| | | Random Forest | |
| **Amount of Loans** | | | |
| Number of employed hh members | 1.430 | 0.755 | 0.640 |
| | (1.262, 1.604) | (0.592, 0.934) | (0.410, 0.872) |
| | - | - | [0.000] |
| Number of retired hh members | 0.135 | 0.430 | -0.270 |
| | (0.056, 0.207) | (0.327, 0.553) | (-0.416, -0.135) |
| | - | - | [0.000] |
| Number of hh members aged 16-19 | 0.120 | 0.430 | -0.315 |
| | (0.048, 0.190) | (0.305, 0.555) | (-0.460, -0.173) |
| | - | - | [0.000] |
| | | SVM | |
| **Outstanding Loan MFI** | | | |
| Highest respondent grade is a university level | 0.035 | 0.125 | -0.090 |
| | (0.000, 0.064) | (0.056, 0.184) | (-0.163, -0.015) |
| | - | - | [0.021] |
| Number of retired hh members | 0.225 | 0.495 | -0.275 |
| | (0.134, 0.317) | (0.364, 0.623) | (-0.432, -0.108) |
| | - | - | [0.001] |
| Economic status of respondent | 1.760 | 2.260 | -0.530 |
| | (1.532, 1.987) | (1.963, 2.560) | (-0.907, -0.153) |
| | - | - | [0.005] |

Notes: Medians over 100 splits. 90% confidence interval in parenthesis.

P-values for the hypothesis that the parameter is equal to zero in brackets.

# D Self-Employment Activities

Table 14: Comparison of ML Methods: Self-Employment Activities

| | **Model** | **Best BLP ($\lambda$)** | **Best GATES ($\bar{\lambda}$)** | **Observations** |
|---|---|---|---|---|
| **Asset Value (BAM)** | Random Forest | 211,070,886 | 584,555,496 | 994 |
| | Elastic Net | 103,318,069 | 590,827,826 | |
| | SVM | 71,300,562 | 550,175,821 | |
| | Boosting | 98,520,074 | 609,229,523 | |
| **Ownership of Inventory [Yes = 1]*** | Random Forest | 0.288 | 5.074 | 994 |
| | Elastic Net | 0.399 | 5.298 | |
| | SVM | 0.299 | 4.777 | |
| | Boosting | 0.247 | 5.090 | |
| **Any Self-employment Income (HH) [Yes = 1]*** | Random Forest | 0.593 | 8.952 | 994 |
| | Elastic Net | 0.484 | 9.609 | |
| | SVM | 0.750 | 10.094 | |
| | Boosting | 0.645 | 9.746 | |
| **Business Ownership [Yes = 1]*** | Random Forest | 0.583 | 9.435 | 994 |
| | Elastic Net | 1.191 | 9.977 | |
| | SVM | 1.664 | 14.275 | |
| | Boosting | 0.824 | 10.394 | |
| **Business in Services [Yes = 1]*** | Random Forest | 0.830 | 5.401 | 994 |
| | Elastic Net | 0.548 | 5.670 | |
| | SVM | 0.959 | 5.294 | |
| | Boosting | 0.654 | 6.192 | |
| **Business in Agriculture [Yes = 1]*** | Random Forest | 0.421 | 6.117 | 994 |
| | Elastic Net | 0.677 | 7.070 | |
| | SVM | 0.784 | 6.730 | |
| | Boosting | 0.733 | 6.932 | |
| **Has Started a Business in Last 14 Months*** | Random Forest | 1.153 | 4.419 | 994 |
| | Elastic Net | 1.581 | 3.831 | |
| | SVM | 0.796 | 3.762 | |
| | Boosting | 0.675 | 3.777 | |
| **Has Closed a Business in Last 14 Months*** | Random Forest | 0.361 | 4.896 | 994 |
| | Elastic Net | 0.515 | 6.030 | |
| | SVM | 0.525 | 6.292 | |
| | Boosting | 0.828 | 5.168 | |
| **Business Expenses (BAM)** | Random Forest | 783,241 | 22,736,441 | 324 |
| | Elastic Net | 2109007 | 24,097,763 | |
| | SVM | 1,649,447 | 24,625,438 | |
| | Boosting | 2,013,669 | 23,707,611 | |
| **Business Profit (BAM)** | Random Forest | 646,261 | 14,717,100 | 329 |
| | Elastic Net | 1,486,803 | 14,934,754 | |
| | SVM | 1,396,432 | 15,118,520 | |
| | Boosting | 1,843,465 | 19,193,780 | |
| **Business Revenue (BAM)** | Random Forest | 2,279,012 | 55,120,555 | 326 |
| | Elastic Net | 7,589,317 | 48,289,716 | |
| | SVM | 3,580,122 | 54,205,159 | |
| | Boosting | 3,636,313 | 52,564,872 | |

Notes: Medians over 100 splits.

BAM: Bosnia and Herzegovina convertible mark. The exchange rate at baseline was US$1 to BAM 1.634.

*Actual values are times $\times 10^{-3}$

Table 15: BLP of Self-Employment Activities

| | Best Learner | | Second Best Learner | |
|---|---|---|---|---|
| | ATE ($\beta_1$) | HET ($\beta_2$) | ATE ($\beta_1$) | HET ($\beta_2$) |
| | Random Forest | | Elastic Net | |
| Asset Value (BAM) | -3190.000 | -0.292 | -2143.000 | -0.122 |
| | (-25990.000, 19263.519) | (-0.760, 0.216) | (-25200.000, 20467.360) | (-0.904, 0.735) |
| | [0.780] | [0.258] | [0.851] | [0.793] |
| | Elastic Net | | SVM | |
| Ownership of Inventory [Yes = 1] | 44.548 | -40.162 | 46.087 | -159.983 |
| | (-9.459, 100.000) | (-771.216, 705.000) | (-9.231, 101.000) | (-718.668, 369.000) |
| | [0.101] | [0.920] | [0.099] | [0.568] |
| | SVM | | Boosting | |
| Any Self-employment Income (HH) [Yes = 1] | 63.630 | -83.330 | 56.980 | 11.150 |
| | (-17.080, 144.000) | (-396.600, 248.000) | (-22.590, 137.000) | (-179.250, 204.000) |
| | [0.116] | [0.623] | [0.157] | [0.910] |
| | SVM | | Elastic Net | |
| Business Ownership [Yes = 1] | 38.760 | -89.330 | 37.800 | 37.980 |
| | (-38.140, 115.000) | (-423.120, 272.000) | (-37.300, 114.000) | (-699.230, 800.000) |
| | [0.302] | [0.653] | [0.309] | [0.903] |
| | SVM | | Random Forest | |
| Business in Services [Yes = 1] | 31.110 | 169.190 | 32.560 | -176.580 |
| | (-35.860, 100.000) | (-223.950, 559.000) | (-35.190, 98.000) | (-657.170, 311.000) |
| | [0.357] | [0.388] | [0.339] | [0.446] |
| | SVM | | Boosting | |
| Business in Agriculture [Yes = 1] | 38.760 | -89.330 | 35.600 | -56.660 |
| | (-38.140, 115.000) | (-423.120, 272.000) | (-39.410, 112.000) | (-238.570, 127.000) |
| | [0.302] | [0.653] | [0.351] | [0.558] |
| | Elastic Net | | Random Forest | |
| Has Started a Business in Last 14 Months | 31.660 | 562.050 | 25.460 | 255.040 |
| | (-18.990, 82.000) | (-703.770, 1741.000) | (-26.150, 77.000) | (-410.600, 938.000) |
| | [0.213] | [0.392] | [0.320] | [0.431] |
| | Elastic Net | | SVM | |
| Has Closed a Business in Last 14 Months | -31.290 | 209.140 | -32.760 | -95.130 |
| | (-97.790, 35.000) | (-860.960, 1336.000) | (-105.400, 39.000) | (-511.870, 301.000) |
| | [0.355] | [0.652] | [0.374] | [0.611] |
| | Elastic Net | | Boosting | |
| Business Expenses (BAM) | 731.700 | 48.560 | 750.800 | 0.219 |
| | (-3927.000, 5325.535) | (-1.113, 1.088) | (-3784.000, 5378.432) | (-0.190, 0.221) |
| | [0.695] | [0.879] | [0.665] | [0.959] |
| | Boosting | | Elastic Net | |
| Business Profit (BAM) | 442.900 | 44.140 | 451.900 | 51.230 |
| | (-3591.000, 4421.957) | (-0.139, 0.201) | (-3538.000, 4328.143) | (-1.326, 1.265) |
| | [0.811] | [0.631] | [0.812] | [0.937] |
| | Elastic Net | | Boosting | |
| Business Revenue (BAM) | 1661.000 | 8.826 | 1678.000 | 9.662 |
| | (-5590.000, 8739.771) | (-1.213, 1.268) | (-5604.000, 8964.463) | (-0.204, 0.214) |
| | [0.595] | [0.972] | [0.637] | [0.915] |

Notes: Medians over 100 splits. 90% confidence interval in parenthesis.

P-values for the hypothesis that the parameter is equal to zero in brackets.

Table 16: GATES of Self-Employment Activities

| | Best Learner | | |
|---|---|---|---|
| | 20% Most ($\gamma_5$) | 20% Least ($\gamma_1$) | Difference ($\gamma_5$ - $\gamma_1$) |
| | | XGBoost | |
| Asset Value (BAM) | -14626.5 | 10837.7 | -33729.1 |
| | (-62817.1, 33981.0) | (-35674.9, 62007.0) | (-106985.5, 39870.0) |
| | [0.559] | [0.601] | [0.327] |
| | | Elastic Net | |
| Ownership of Inventory [Yes = 1] | 0.048 | 0.051 | -0.006 |
| | (-0.073, 0.176) | (-0.071, 0.174) | (-0.182, 0.164) |
| | [0.400] | [0.416] | [0.949] |
| | | SVM | |
| Any Self-employment Income (HH) [Yes = 1] | 0.046 | 0.109 | -0.074 |
| | (-0.144, 0.231) | (-0.079, 0.297) | (-0.334, 0.191) |
| | [0.603] | [0.255] | [0.583] |
| | | SVM | |
| Business Ownership [Yes = 1] | 0.012 | 0.161 | -0.143 |
| | (-0.181, 0.208) | (-0.034, 0.356) | (-0.418, 0.132) |
| | [0.871] | [0.104] | [0.306] |
| | | XGBoost | |
| Business in Services [Yes = 1] | 0.014 | 0.071 | -0.064 |
| | (-0.150, 0.170) | (-0.084, 0.221) | (-0.295, 0.153) |
| | [0.845] | [0.351] | [0.573] |
| | | Elastic Net | |
| Business in Agriculture [Yes = 1] | 0.031 | 0.035 | 0.011 |
| | (-0.144, 0.201) | (-0.124, 0.193) | (-0.225, 0.244) |
| | [0.702] | [0.648] | [0.918] |
| | | Random Forest | |
| Has Started a Business in Last 14 Months | 0.105 | 0.008 | 0.073 |
| | (-0.070, 0.289) | (-0.126, 0.154) | (-0.157, 0.305) |
| | [0.215] | [0.906] | [0.529] |
| | | SVM | |
| Has Closed a Business in Last 14 Months | -0.054 | 0.004 | -0.073 |
| | (-0.218, 0.114) | (-0.150, 0.167) | (-0.303, 0.163) |
| | [0.533] | [0.958] | [0.540] |
| | | SVM | |
| Business Expenses (BAM) | 85.830 | 1358.730 | -2156.790 |
| | (-7498.260, 7295.000) | (-4903.830, 7678.000) | (-12880.770, 10278.000) |
| | [0.941] | [0.594] | [0.788] |
| | | XGBoost | |
| Business Profit (BAM) | 1648.600 | 157.500 | 1681.700 |
| | (-5191.400, 8807.000) | (-6643.000, 7724.000) | (-10574.800, 12051.000) |
| | [0.663] | [0.945] | [0.773] |
| | | Random Forest | |
| Business Revenue (BAM) | 3731.000 | 4068.000 | 1066.000 |
| | (-9122.000, 15763.000) | (-7995.000, 16888.000) | (-19644.000, 20184.000) |
| | [0.523] | [0.573] | [0.909] |

Notes: Medians over 100 splits. 90% confidence interval in parenthesis.

P-values for the hypothesis that the parameter is equal to zero in brackets.

# E Income

Table 17: Comparison of ML Methods: Income

| | Model | Best BLP ($\Lambda$) | Best GATES ($\bar{\Lambda}$) | Observations |
|---|---|---|---|---|
| **Self-employment Likelihood*** | Random Forest | 0.5929 | 8.952 | 994 |
| | Elastic Net | 0.4841 | 9.609 | |
| | SVM | 0.7495 | 10.094 | |
| | Boosting | 0.6447 | 9.746 | |
| **Self-employment Amount (BAM)** | Random Forest | 215308 | 2335121 | 994 |
| | Elastic Net | 235749 | 2825106 | |
| | SVM | 134352 | 2553007 | |
| | Boosting | 275719 | 2891504 | |
| **Wages Likelihood*** | Random Forest | 0.9931 | 13.71 | 994 |
| | Elastic Net | 0.6528 | 13.49 | |
| | SVM | 0.4237 | 14.31 | |
| | Boosting | 1.1889 | 15.07 | |
| **Wages Amount (BAM)** | Random Forest | 232099 | 2042054 | 994 |
| | Elastic Net | 313385 | 2477025 | |
| | SVM | 216850 | 2788649 | |
| | Boosting | 242469 | 2238628 | |
| **Remittances Likelihood*** | Random Forest | 0.3332 | 5.613 | 994 |
| | Elastic Net | 0.5308 | 5.707 | |
| | SVM | 0.4284 | 5.493 | |
| | Boosting | 0.5666 | 5.847 | |
| **Remittances Amount (BAM)** | Random Forest | 27753 | 171649 | 994 |
| | Elastic Net | 24261 | 164098 | |
| | SVM | 30308 | 145685 | |
| | Boosting | 12290 | 163481 | |
| **Government Benefits Likelihood*** | Random Forest | 0.5975 | 8.704 | 994 |
| | Elastic Net | 0.7204 | 9.677 | |
| | SVM | 0.5207 | 7.116 | |
| | Boosting | 0.4855 | 7.472 | |
| **Government Benefits Amount (BAM)** | Random Forest | 5573 | 73969 | 994 |
| | Elastic Net | 6403 | 66417 | |
| | SVM | 11726 | 62092 | |
| | Boosting | 5190 | 61859 | |

Notes: Medians over 100 splits.

BAM: Bosnia and Herzegovina convertible mark. The exchange rate at baseline was US$1 to BAM 1.634.

*Actual values are times $\times 10^{-3}$

Table 18: BLP of Income

| | Best Learner | | Second Best Learner | |
|---|---|---|---|---|
| | ATE ($\beta_1$) | HET ($\beta_2$) | ATE ($\beta_1$) | HET ($\beta_2$) |
| | SVM | | Boosting | |
| Self-employment Likelihood | 0.001 | 0.010 | 0.001 | 0.010 |
| | (-0.121, 0.041) | (-0.237, 0.335) | (-0.122, 0.038) | (-0.210, 0.189) |
| | [0.331] | [0.836] | [0.311] | [0.904] |
| | Boosting | | Elastic Net | |
| Self-employment Amount (BAM) | 275719 | 2891504 | 235749 | 2825106 |
| | (-1477.000, 1762.620) | (-0.338, 0.283) | (-1564.000, 1692.189) | (-0.781, 0.797) |
| | [0.832] | [0.917] | [0.892] | [0.942] |
| | Boosting | | Random Forest | |
| Wages Likelihood | 0.001 | 0.015 | 0.001 | 0.014 |
| | (-0.096, 0.051) | (-0.181, 0.175) | (-0.097, 0.048) | (-0.558, 0.421) |
| | [0.535] | [0.987] | [0.498] | [0.810] |
| | Elastic Net | | Boosting | |
| Wages Amount (BAM) | 313385 | 2477025 | 242469 | 2238628 |
| | (-1330.000, 1860.079) | (-0.352, 0.301) | (-1489.000, 1683.246) | (-0.235, 0.163) |
| | [0.769] | [0.813] | [0.869] | [0.802] |
| | Boosting | | Elastic Net | |
| Remittances Likelihood | 0.001 | 0.006 | 0.001 | 0.006 |
| | (-0.097, 0.048) | (-0.558, 0.421) | (-0.095, 0.048) | (-0.739, 0.856) |
| | [0.498] | [0.810] | [0.513] | [0.870] |
| | SVM | | Random Forest | |
| Remittances Amount (BAM) | 30.308 | 145.685 | 27.753 | 171.649 |
| | (-0.420, 358.636) | (-0.640, 0.586) | (-0.426, 362.847) | (-0.174, 0.095) |
| | [0.895] | [0.938] | [0.945] | [0.584] |
| | Elastic Net | | Random Forest | |
| Government Benefits Likelihood | 0.001 | 0.010 | 0.001 | 0.009 |
| | (-0.130, 0.029) | (-1.248, 0.984) | (-0.129, 0.028) | (-0.607, 0.376) |
| | [0.221] | [0.821] | [0.211] | [0.647] |
| | Elastic Net | | Random Forest | |
| Government Benefits Amount (BAM) | 6.403 | 66.417 | 5.573 | 73.969 |
| | (-0.361, 85.594) | (-0.809, 0.817) | (-0.377, 65.340) | (-0.434, 0.670) |
| | [0.233] | [0.943] | [0.166] | [0.706] |

Notes: Medians over 100 splits. 90% confidence interval in parenthesis.

P-values for the hypothesis that the parameter is equal to zero in brackets.

## Table 19: GATES of Income

| | 20% Most ($\gamma_5$) | 20% Least ($\gamma_1$) | Difference ($\gamma_5 - \gamma_1$) |
|---|---|---|---|
| | | Elastic Net | |
| Government Benefits Likelihood | -0.082 | -0.032 | -0.053 |
| | (-0.251, 0.086) | (-0.215, 0.149) | (-0.313, 0.208) |
| | [0.327] | [0.742] | [0.670] |
| | | XGBoost | |
| Remittances Likelihood | -0.039 | -0.021 | -0.022 |
| | (-0.201, 0.120) | (-0.183, 0.143) | (-0.238, 0.204) |
| | [0.610] | [0.785] | [0.849] |
| | | SVM | |
| Self-employment Likelihood | 0.046 | 0.109 | -0.074 |
| | (-0.144, 0.231) | (-0.079, 0.297) | (-0.334, 0.191) |
| | [0.603] | [0.255] | [0.583] |
| | | XGBoost | |
| Wages Likelihood | -0.047 | -0.138 | 0.081 |
| | (-0.227, 0.133) | (-0.323, 0.044) | (-0.179, 0.334) |
| | [0.614] | [0.138] | [0.536] |
| | | Random Forest | |
| Government Benefits Amount (BAM) | -201.170 | -330.090 | 98.470 |
| | (-570.630, 184.900) | (-922.690, 303.800) | (-609.500, 855.900) |
| | [0.300] | [0.302] | [0.750] |
| | | Random Forest | |
| Remittances Amount (BAM) | -302.110 | 244.390 | -641.940 |
| | (-1363.290, 657.000) | (-501.020, 975.500) | (-1952.510, 669.600) |
| | [0.468] | [0.520] | [0.362] |
| | | XGBoost | |
| Self-employment Amount (BAM) | -139.100 | 379.800 | -244.700 |
| | (-3896.900, 3686.000) | (-3574.800, 4259.000) | (-5523.000, 5289.000) |
| | [0.950] | [0.808] | [0.929] |
| | | SVM | |
| Wages Amount (BAM) | -651.630 | -13.820 | -777.960 |
| | (-4462.390, 3055.000) | (-3954.140, 3628.000) | (-5994.710, 4916.000) |
| | [0.717] | [0.987] | [0.807] |

Notes: Medians over 100 splits. 90% confidence interval in parenthesis.

P-values for the hypothesis that the parameter is equal to zero in brackets.

Table 20: CLAN of Remittances Amount and Government Benefits Likelihood

| | 20% Most ($\delta_5$) | 20% Least ($\delta_1$) | Difference ($\delta_5 - \delta_1$) |
|---|---|---|---|
| | Random Forest | | |
| **Remittances Amount** | | | |
| Dwelling ownership | 0.905 | 0.705 | 0.186 |
| | (0.841, 0.959) | (0.610, 0.790) | (0.083, 0.296) |
| | - | - | [0.000] |
| Ownnership of business | 0.745 | 0.495 | 0.245 |
| | (0.654, 0.826) | (0.392, 0.588) | (0.120, 0.378) |
| | - | - | [0.000] |
| Respondent is female | 0.255 | 0.525 | -0.255 |
| | (0.165, 0.335) | (0.426, 0.624) | (-0.384, -0.120) |
| | - | - | [0.000] |
| | Elastic Net | | |
| **Government Benefits Likelihood** | | | |
| Number of hh members attending school | 0.775 | 0.920 | -0.110 |
| | (0.598, 0.959) | (0.732, 1.111) | (-0.370, 0.150) |
| | - | - | [0.407] |
| Number of kids aged 6-10 | 0.245 | 0.360 | -0.075 |
| | (0.141, 0.348) | (0.229, 0.490) | (-0.234, 0.083) |
| | - | - | [0.371] |
| Number of kids aged 11-16 | 0.310 | 0.380 | -0.045 |
| | (0.198, 0.417) | (0.260, 0.513) | (-0.217, 0.117) |
| | - | - | [0.612] |

Notes: Medians over 100 splits. 90% confidence interval in parenthesis.

P-values for the hypothesis that the parameter is equal to zero in brackets.

# F Hours-Worked

Table 21: Comparison of ML Methods: Hours Worked

| | **Model** | **Best BLP ($\Lambda$)** | **Best GATES ($\bar{\Lambda}$)** | **Observations** |
|---|---|---|---|---|
| **Total Hours Worked** | Random Forest | 50.07 | 119.60 | 994 |
| | Elastic Net | 13.04 | 106.30 | |
| | SVM | 41.92 | 126.40 | |
| | Boosting | 17.39 | 99.00 | |
| **Total Hours Worked on Business** | Random Forest | 10.11 | 88.21 | 994 |
| | Elastic Net | 6.07 | 93.77 | |
| | SVM | 5.47 | 75.60 | |
| | Boosting | 10.15 | 99.58 | |
| **Total Hours Worked on Other Activities** | Random Forest | 7.69 | 70.69 | 994 |
| | Elastic Net | 3.86 | 71.62 | |
| | SVM | 10.09 | 71.45 | |
| | Boosting | 6.50 | 62.84 | |

Notes: Medians over 100 splits.

Table 22: BLP of Hours Worked

| | Best Learner | | Second Best Learner | |
|---|---|---|---|---|
| | ATE ($\beta_1$) | HET ($\beta_2$) | ATE ($\beta_1$) | HET ($\beta_2$) |
| | Random Forest | | SVM | |
| Total Hours Worked | -2.243 | -0.333 | -0.913 | -0.295 |
| | (-11.605, 7.519) | (-0.830, 0.144) | (-10.600, 8.405) | (-0.749, 0.155) |
| | [0.659] | [0.172] | [0.852] | [0.194] |
| | Elastic Net | | XGBoost | |
| Total Hours Worked on Business | 4.208 | -0.026 | 3.922 | -0.056 |
| | (-4.209, 12.750) | (-0.762, 0.880) | (-4.383, 12.319) | (-0.240, 0.131) |
| | [0.325] | [0.948] | [0.342] | [0.572] |
| | Elastic Net | | SVM | |
| Total Hours Worked on Other Activities | -5.342 | -0.258 | -4.241 | -0.166 |
| | (-12.063, 1.467) | (-2.048, 1.404) | (-11.255, 2.830) | (-0.588, 0.253) |
| | [0.124] | [0.712] | [0.245] | [0.449] |

Notes: Medians over 100 splits. 90% confidence interval in parenthesis.

P-values for the hypothesis that the parameter is equal to zero in brackets.

Table 23: GATES of Hours Worked

| | 20% Most ($\gamma_5$) | 20% Least ($\gamma_1$) | Difference ($\gamma_5 - \gamma_1$) |
|---|---|---|---|
| | | SVM | |
| Total Hours Worked | -11.989 | 6.985 | -18.940 |
| | (-31.938, 9.087) | (-15.598, 28.428) | (-49.847, 11.212) |
| | [0.246] | [0.509] | [0.233] |
| | | XGBoost | |
| Total Hours Worked on Business | 0.830 | 10.443 | -10.126 |
| | (-18.651, 20.520) | (-8.061, 28.760) | (-36.348, 17.770) |
| | [0.931] | [0.244] | [0.465] |
| | | Elastic Net | |
| Total Hours Worked on Other Activities | -7.913 | -5.125 | -3.487 |
| | (-21.933, 6.271) | (-21.853, 10.172) | (-26.367, 19.209) |
| | [0.269] | [0.475] | [0.779] |

Notes: Medians over 100 splits. 90% confidence interval in parenthesis.

P-values for the hypothesis that the parameter is equal to zero in brackets.

Table 24: CLAN of Total Hours Worked and Total Hours Worked on Other Activities

| | 20% Most ($\delta_5$) | 20% Least ($\delta_1$) | Difference ($\delta_5 - \delta_1$) |
|---|---|---|---|
| | | SVM | |
| **Total Hours Worked** | | | |
| Number of kids aged 0-5 | 0.385 | 0.145 | 0.240 |
| | (0.261, 0.524) | (0.074, 0.226) | (0.088, 0.401) |
| | - | - | [0.002] |
| Number of kids aged 11-16 | 0.255 | 0.495 | -0.220 |
| | (0.146, 0.366) | (0.365, 0.632) | (-0.400, -0.048) |
| | - | - | [0.012] |
| Respondent age | 39.810 | 35.755 | 5.040 |
| | (37.601, 42.071) | (33.324, 38.021) | (1.768, 8.282) |
| | - | - | [0.002] |
| | | Elastic Net | |
| **Total Hours Worked on Other Activities** | | | |
| Respondent age | 43.200 | 33.560 | 9.170 |
| | (40.781, 45.900) | (31.385, 35.810) | (5.854, 12.630) |
| | - | - | [0.000] |
| Number of employed hh members | 0.990 | 1.133 | -0.143 |
| | (0.792, 1.188) | (0.958, 1.307) | (-0.407, 0.121) |
| | - | - | [0.289] |
| Number of kids aged 16-19 | 0.210 | 0.350 | -0.122 |
| | (0.125, 0.300) | (0.237, 0.477) | (-0.275, 0.028) |
| | - | - | [0.112] |

Notes: Medians over 100 splits. 90% confidence interval in parenthesis.

P-values for the hypothesis that the parameter is equal to zero in brackets.

# G Consumption and Savings

Table 25: Comparison of ML Methods: Consumption and Savings

| | Model | Best BLP ($\Lambda$) | Best GATES ($\bar{\Lambda}$) | Observations |
|---|---|---|---|---|
| **Total consumption per capita (BAM)** | Random Forest | 78799 | 1330812 | 994 |
| | Elastic Net | 85600 | 1251915 | |
| | SVM | 147244 | 1426732 | |
| | Boosting | 106975 | 1300171 | |
| **Durables (BAM)** | Random Forest | 70808 | 1185298 | 994 |
| | Elastic Net | 86347 | 1054495 | |
| | SVM | 99363 | 1051619 | |
| | Boosting | 91651 | 1228525 | |
| **Nondurable (BAM)** | Random Forest | 266.7 | 1704 | 993 |
| | Elastic Net | 424.8 | 2174 | |
| | SVM | 141.6 | 2018 | |
| | Boosting | 211.5 | 2216 | |
| **Food (BAM)** | Random Forest | 27.05 | 247.7 | 994 |
| | Elastic Net | 25.26 | 291.3 | |
| | SVM | 17.60 | 264.7 | |
| | Boosting | 25.78 | 274.0 | |
| **Education (BAM)** | Random Forest | 6492 | 45497 | 994 |
| | Elastic Net | 6467 | 49283 | |
| | SVM | 8092 | 57420 | |
| | Boosting | 6298 | 53072 | |
| **Cigarettes and alcohol (BAM)** | Random Forest | 2.316 | 19.67 | 994 |
| | Elastic Net | 1.438 | 20.66 | |
| | SVM | 1.500 | 20.16 | |
| | Boosting | 2.198 | 19.92 | |
| **Recreation (BAM)** | Random Forest | 153.0 | 1637 | 994 |
| | Elastic Net | 217.0 | 1900 | |
| | SVM | 218.3 | 1474 | |
| | Boosting | 352.2 | 1687 | |
| **Home durable good index*** | Random Forest | 1.268 | 11.479 | 994 |
| | Elastic Net | 0.755 | 10.896 | |
| | SVM | 0.483 | 9.378 | |
| | Boosting | 0.303 | 10.091 | |
| **Savings (BAM)** | Random Forest | 56944 | 483652 | 994 |
| | Elastic Net | 40114 | 437093 | |
| | SVM | 150878 | 593797 | |
| | Boosting | 45062 | 499930 | |

Notes: Medians over 100 splits.

BAM: Bosnia and Herzegovina convertible mark. The exchange rate at baseline was US$1 to BAM 1.634.

*Actual values are times $\times 10^{-3}$

Table 26: BLP of Consumption and Savings

| | Best Learner | | Second Best Learner | |
|---|---|---|---|---|
| | ATE ($\beta_1$) | HET ($\beta_2$) | ATE ($\beta_1$) | HET ($\beta_2$) |
| | SVM | | Boosting | |
| Total consumption per capita (BAM) | -729.038 | 0.240 | -641.900 | -0.013 |
| | (-1625.438, 230.606) | (-0.409, 0.879) | (-1609.000, 265.839) | (-0.187, 0.162) |
| | [0.133] | [0.410] | [0.152] | [0.853] |
| | SVM | | Boosting | |
| Durables (BAM) | -26.994 | 0.148 | -45.180 | -0.035 |
| | (-1010.366, 987.238) | (-0.737, 0.935) | (-1009.000, 983.380) | (-0.211, 0.129) |
| | [0.947] | [0.741] | [0.908] | [0.653] |
| | Elastic Net | | Random Forest | |
| Nondurable (BAM) | -15.785 | 0.023 | -17.036 | -0.171 |
| | (-58.998, 25.908) | (-0.815, 0.752) | (-58.699, 24.929) | (-0.637, 0.289) |
| | [0.454] | [0.930] | [0.420] | [0.485] |
| | Random Forest | | Elastic Net | |
| Food (BAM) | -4.380 | 0.132 | -2.951 | -0.057 |
| | (-19.640, 11.534) | (-0.349, 0.613) | (-19.182, 12.554) | (-0.810, 0.758) |
| | [0.596] | [0.575] | [0.722] | [0.891] |
| | SVM | | Elastic Net | |
| Education (BAM) | -58.433 | 0.340 | -49.796 | -0.025 |
| | (-252.619, 150.928) | (-0.386, 1.049) | (-250.968, 167.283) | (-0.676, 0.599) |
| | [0.537] | [0.357] | [0.656] | [0.930] |
| | Random Forest | | Elastic Net | |
| Cigarettes and alcohol (BAM) | -2.475 | -0.159 | -2.571 | 0.019 |
| | (-5.912, 1.172) | (-0.623, 0.349) | (-5.998, 1.023) | (-0.720, 0.792) |
| | [0.182] | [0.527] | [0.167] | [0.931] |
| | Elastic Net | | SVM | |
| Recreation (BAM) | -7.523 | -0.013 | -8.197 | -0.397 |
| | (-46.548, 31.848) | (-0.730, 0.688) | (-47.449, 31.372) | (-2.091, 1.342) |
| | [0.725] | [0.965] | [0.669] | [0.666] |
| | Random Forest | | Elastic Net | |
| Home durable good index | -0.068 | 0.204 | -0.071 | -0.004 |
| | (-0.144, 0.009) | (-0.223, 0.628) | (-0.148, 0.011) | (-0.774, 0.783) |
| | [0.085] | [0.368] | [0.088] | [0.980] |
| | SVM | | Boosting | |
| Savings (BAM) | -456.470 | 0.735 | -461.203 | 0.031 |
| | (-952.182, 32.633) | (-0.222, 1.767) | (-945.529, 42.915) | (-0.133, 0.217) |
| | [0.069] | [0.125] | [0.074] | [0.642] |

Notes: Medians over 100 splits. 90% confidence interval in parenthesis.

P-values for the hypothesis that the parameter is equal to zero in brackets.

Table 27: GATES of Consumption and Savings

| | 20% Most ($\gamma_5$) | 20% Least ($\gamma_1$) | Difference ($\gamma_5 - \gamma_1$) |
|---|---|---|---|
| | | SVM | |
| Total consumption per capita (BAM) | -233.2 | -1212.5 | 929.6 |
| | (-1655.8, 1376.1) | (-2983.0, 568.5) | (-1333.9, 3537.2) |
| | [0.734] | [0.181] | [0.391] |
| | | XGBoost | |
| Durables (BAM) | -379.0 | 533.8 | -1167.0 |
| | (-2034.1, 1470.0) | (-1177.4, 2272.0) | (-3673.9, 1725.0) |
| | [0.662] | [0.510] | [0.385] |
| | | XGBoost | |
| Nondurable (BAM) | -52.472 | -12.078 | -35.454 |
| | (-139.271, 36.770) | (-109.406, 89.190) | (-165.980, 98.060) |
| | [0.252] | [0.867] | [0.636] |
| | | Elastic Net | |
| Food (BAM) | -1.994 | -3.833 | -1.631 |
| | (-35.392, 33.270) | (-36.767, 30.830) | (-50.586, 47.520) |
| | [0.908] | [0.837] | [0.948] |
| | | SVM | |
| Education (BAM) | -6.131 | -229.865 | 243.820 |
| | (-218.281, 268.000) | (-689.281, 214.800) | (-271.025, 787.200) |
| | [0.964] | [0.318] | [0.322] |
| | | Elastic Net | |
| Cigarettes and alcohol (BAM) | -1.386 | -2.491 | 1.059 |
| | (-8.518, 5.791) | (-9.530, 5.520) | (-10.087, 11.373) |
| | [0.661] | [0.491] | [0.836] |
| | | Elastic Net | |
| Recreation (BAM) | -11.132 | -8.159 | -5.306 |
| | (-92.655, 52.830) | (-77.078, 60.460) | (-115.419, 100.890) |
| | [0.787] | [0.797] | [0.922] |
| | | Random Forest | |
| Home durable good index | 0.007 | -0.107 | 0.111 |
| | (-0.157, 0.176) | (-0.266, 0.051) | (-0.134, 0.340) |
| | [0.929] | [0.174] | [0.359] |
| | | SVM | |
| Savings (BAM) | -261.7 | -1208.9 | 950.1 |
| | (-1108.2, 567.3) | (-2453.5, 102.3) | (-511.0, 2508.1) |
| | [0.515] | [0.072] | [0.187] |

Notes: Medians over 100 splits. 90% confidence interval in parenthesis.

P-values for the hypothesis that the parameter is equal to zero in brackets.

Table 28: CLAN of Savings and HDGI

| | 20% Most ($\delta_5$) | 20% Least ($\delta_1$) | Difference ($\delta_5 - \delta_1$) |
|---|---|---|---|
| | | SVM | |
| **Savings** | | | |
| Respondent is female | 0.565 | 0.240 | 0.310 |
| | (0.462, 0.658) | (0.156, 0.324) | (0.182, 0.438) |
| | - | - | [0.000] |
| Number of employed hh members | 1.045 | 1.395 | -0.360 |
| | (0.877, 1.229) | (1.196, 1.596) | (-0.623, -0.084) |
| | - | - | [0.010] |
| Number of hh members attending school | 0.675 | 1.075 | -0.410 |
| | (0.513, 0.846) | (0.875, 1.289) | (-0.677, -0.140) |
| | - | - | [0.002] |
| | | Random Forest | |
| **HDGI** | | | |
| Dwelling ownership | 0.975 | 0.645 | 0.315 |
| | (0.936, 0.999) | (0.545, 0.735) | (0.220, 0.417) |
| | - | - | [0.000] |
| Number of female hh members | 1.405 | 1.850 | -0.465 |
| | (1.236, 1.595) | (1.664, 2.066) | (-0.758, -0.176) |
| | - | - | [0.001] |
| Number of hh members attending school | 0.470 | 0.990 | -0.530 |
| | (0.316, 0.625) | (0.793, 1.205) | (-0.774, -0.268) |
| | - | - | [0.000] |

Notes: Medians over 100 splits. 90% confidence interval in parenthesis.
P-values for the hypothesis that the parameter is equal to zero in brackets.

# H Other

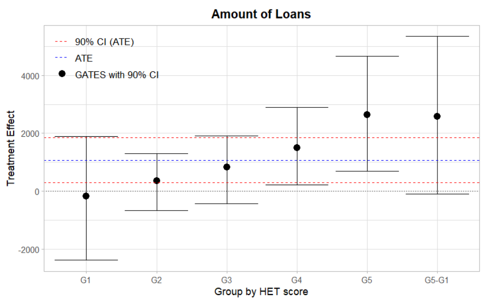


Figure 9: GATES of Moroccan Amount of Loans

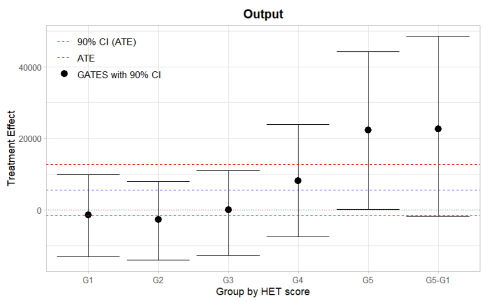


Figure 10: GATES of Moroccan Output

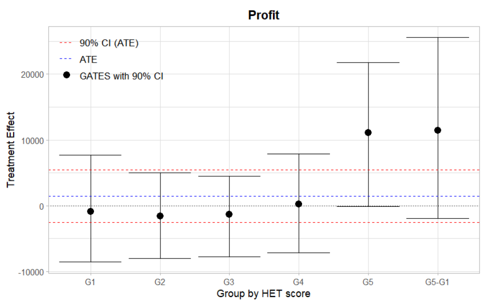


Figure 11: GATES of Moroccan Profit

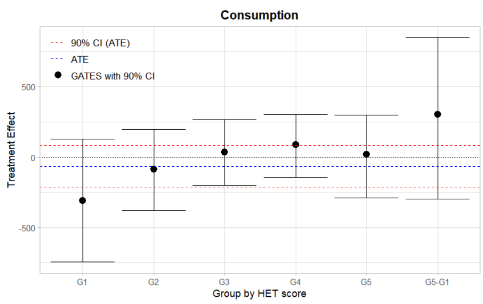


Figure 12: GATES of Moroccan Consumption

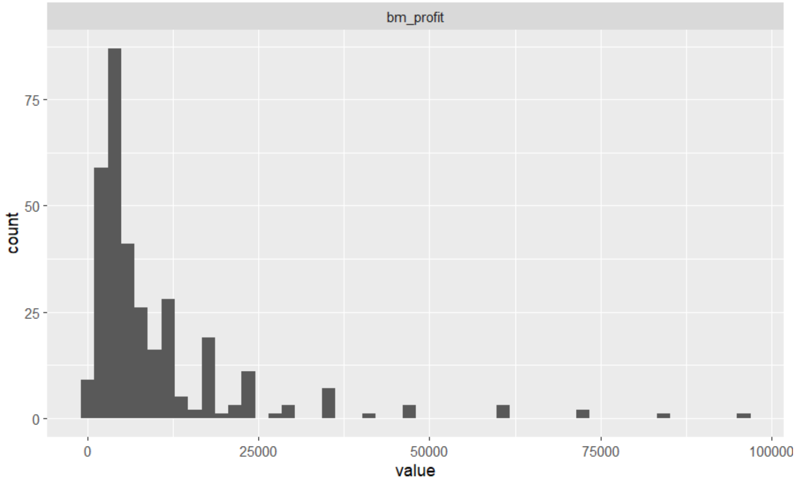


Figure 13: Histogram of Profit (Bosnia)

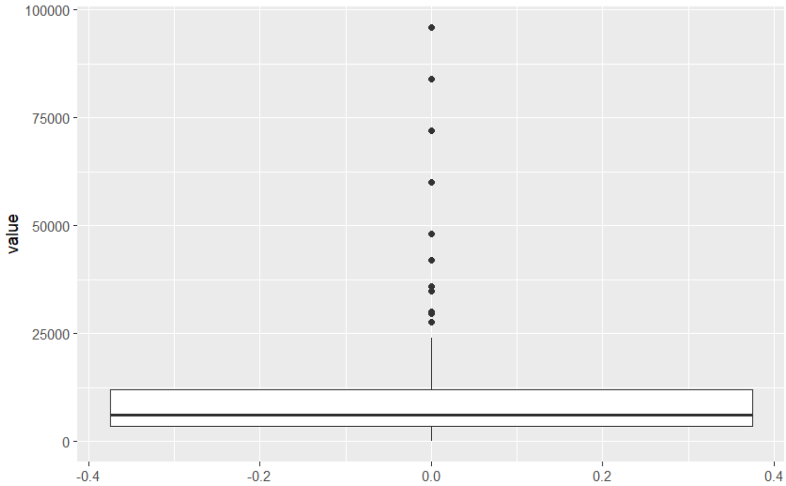


Figure 14: Box plot of Profit (Bosnia)