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

Reducing Intimate Partner Violence: heterogeneous treatment effects in Ethiopia

Ruben Baijense (489402)

Supervisor:	Welz, M.F.O.
Second assessor:	Franses, P.H.B.F.
Date final version:	1st July 2024

Abstract

The Unite for a Better Life program was series of gender-transformative education sessions aimed at reducing Intimate Partner Violence in Ethiopia (IPV). While previous research did not find significant average treatment effects on women's reported experience of (IPV), it did not examine heterogeneity in these effects. This paper evaluates whether heterogeneous effects were present and, if so, identifies possible drivers of heterogeneity. The data consists of a baseline survey on household characteristics ($n = 6770$) and an endline survey on women's experience of IPV ($n = 6045$). We employ Machine Learning methods to approximate Conditional Average Treatment Effects and use these proxies to construct measures of heterogeneity. We find significant evidence of heterogeneity in reported sexual and emotional IPV. Findings suggest that the treatment was most effective among young, poor, Islamic couples, and may have had an adverse effect among older, richer, Orthodox couples. This indicates that programs such as Unite for a Better Life could target the most receptive demographics in order to increase their effectiveness.

1 Introduction

The Randomised Controlled Trial (RCT) is one of the most widely used experimental designs in the field of data analysis. It is used to assess the effectiveness of treatments in many fields, ranging from economics to medicine to sociology. Estimating average treatment effects over the full sample is a staple of this analysis, but with the availability of larger data sets, researchers are more and more interested in establishing relations between treatment effects and observation characteristics, as this may allow for more insight into the underlying mechanics of the problem. However, performing subgroup analysis poses a dilemma. On the one hand, choosing which subgroups to analyse ex post risks overfitting, P-Hacking and hypothesising after the results are known. On the other hand, although ex ante subgroup selection is academically more responsible, it limits the amount of relations which can be inferred from the data set, possibly forcing researchers to overlook valuable information.

In recent years, Machine Learning (ML) methods have gained much traction due to their ability to identify patterns, especially in high-dimensional data sets, where they often outperform classical methods. This makes ML a viable tool for our subgroup selection problem when there are many covariates, but few whose relevance to the problem is already established.

One such field where heterogeneous treatment effects may be present, is that of reduction in Intimate Partner Violence (IPV). Globally, 30% of women are estimated to be the victim of physical and/or sexual violence perpetrated by their partner at least once in their life (Devries et al., 2013). Victims and their families are more likely to suffer from a range of both physical and mental health issues (Campbell, 2002). Its prevalence in Sub-Saharan Africa is especially high (Garcia-Moreno, Jansen, Ellsberg, Heise & Watts, 2006), which has motivated various organisations to conduct interventions throughout the region.

One of these interventions was the Unite for a Better Life (UBL) program, which was conducted in 64 villages in Ethiopia between 2014 and 2018. These villages were randomly divided into 3 treatment arms, one in which women would participate in gender-transformative education sessions, one in which men would participate, one in which couples would, and a control arm.

In each village, approximately 106 cohabiting couples were sampled to be included in the study, resulting in a total of 6770 households. 24 months after the study, 90% of the participating women completed a survey on their past year experience of physical, sexual, and emotional IPV. After its conclusion, its efficacy was reviewed in Sharma, Leight, Verani, Tewolde and Deyessa (2020), who found that when compared to the control group, the treatment did not result in a significant change in women's reporting of past-year experience of physical IPV, sexual IPV, and emotional IPV.

The research question of this paper is as follows: "Was there heterogeneity in the treatment effect of the UBL program?" In the original study, the significance of the treatment effect is evaluated using two odds ratios: one unadjusted and one adjusted for a set of covariates collected in a baseline survey, including household demographics and socioeconomic variables. This paper examines whether these covariates are a source of heterogeneity in the effectiveness of the program.

To do so we use a method developed in Chernozhukov, Demirer, Duflo and Fernández-Val (2023). In general, a key problem with ML methods is that in high-dimensional settings, strong assumptions are required to validly infer heterogeneous effects. The Conditional Average Treatment Effect (CATE), which is the average difference in outcome between a treated population and an untreated population conditional on covariates, may not even be estimated consistently by such methods. Chernozhukov et al. (2023) proposes a methodology to avoid some of the inherent problems with inference on the CATE with ML techniques. It does so by using ML, not to infer results on the CATE directly, but to make proxy predictors of the CATE and subsequently using them to infer results on three related measures: the Best Linear Predictor (BLP), the Sorted Group Average Treatment Effects (GATES), and the Classification Analysis (CLAN). To avoid overfitting, the sample is randomly split in an auxiliary sample to train the ML proxies, and a main sample which the proxies use to estimate the measures. To reduce the dependency of these estimates on the split, this process is done repeatedly and its results are aggregated.

We find that across all treatment arms, the included covariates do not capture significant heterogeneity in effect of the UBL program on the reported experience of physical IPV ($p > 0.145$). They do however suggest significant heterogeneity in the effect on sexual IPV ($p < 0.001$) and in the effect on emotional IPV ($p < 0.001$). For sexual IPV, the heterogeneity is driven by both a significant decrease (17.9% to 23.6%) in reporting by the women in the most positively affected quintile, as well as a significant increase (14.1% to 19.5%) in reporting by women in the most negatively affected quintile, with percentages dependent on the treatment arm. Heterogeneity in emotional IPV is driven across all arms by a significant reduction (23.1% to 28.0%) in reporting in the most positively affected quintile, while a significant increase (15.2%) in reporting is only found in the most negatively affected quintile of the men's treatment arm. Examining the covariates for the men's UBL, we find that on average the couples in the most positively affected quintile were had been married for shorter amount of time (Sex.: $p < 0.001$, Emo.: $p < 0.001$), had fewer children (Sex.: $p = 0.013$, Emo.: $p < 0.001$), had fewer assets (Sex.: $p < 0.001$, Emo.:

$p = 0.031$), and consisted of a younger man (Sex.: $p < 0.001$, Emo.: $p < 0.001$) and a younger woman (Sex.: $p < 0.001$, Emo.: $p < 0.001$), than the couples in the most negatively affected quintile. For couples' UBL, we find that couples that have fewer assets (Sex.: $p = 0.007$, Emo.: $p < 0.001$), are less wealthy (Sex.: $p < 0.001$, Emo.: $p < 0.001$), consist of a lower educated man (Sex.: $p < 0.001$, Emo.: $p = 0.035$), are Muslim (Sex.: $p < 0.001$, Emo.: $p < 0.001$), and are not Orthodox (Sex.: $p < 0.001$, Emo.: $p < 0.001$) are significantly more represented in the most positively affected quintile than in the most negatively affected quintile. These findings show that the UBL program did actually benefit certain demographics, and suggest that similar programs can reduce IPV, when applied to the appropriate people.

The rest of this paper will be structured as follows. Section 2 will provide a theoretical overview of heterogeneity targeting in RCTs and a review of current research in IPV. Section 3 describes the intervention and data collected before and afterward. Section 4 describes our methodology. Section 5 provides the results of our examination of heterogeneity. Section 6 summarises our findings and provides a short discussion on the limitations of this research.

2 Literature Review

Machine Learning refers to a broad array of adaptive estimation methods including Penalised Regression, Random Forests, Boosted Trees and Neural Networks. We refer to Friedman, Hastie and Tibshirani (2001) for a comprehensive overview. These methods are often better at capturing non-linear relations in high-dimensional setting than classical methods are, and have found widespread popularity over the past decade. In the context of RCTs, it thus seems to be an excellent tool for estimating the CATE if there are many covariates present and if few of them have pre-established effects. Their are however some major caveats. Stone (1982) shows that without an assumption of some form of sparsity, there exists no consistent general estimator of the CATE. Additionally, general adaptive confidence bands do not exist even in low dimensional settings (Genovese & Wasserman, 2007). Finally, even if consistent ML methods which require few assumptions would be found, then it is still a far way to go from the theoretical existence of perfect tuning parameters, to practical guidelines on how to find them.

Researchers have since attempted to work around these problems. Some have focused their efforts on developing methods for settings where certain assumptions can reasonably be assumed. Belloni, Chernozhukov and Kato (2014) proposes confidence regions for coefficients in such settings with the additional assumption of homoskedastic error terms, while Dezeure, Bühlmann and Zhang (2016) proposes a bootstrap for linear models which allow for non-normal heteroskedasticity. Hansen, Kozbur and Misra (2018) provides a post-model selection procedure to infer confidence bounds, while Zhao, Small and Ertefaie (2021) uses both a generic ML method and LASSO to eliminate nuisance parameters. All these methods are however heavily reliant on assumptions of sparsity. Additionally, Giné and Nickl (2010) and Chernozhukov et al. (2023) propose estimation methods for adaptive confidence bands, although these require self-similarity conditions, as well as low-dimensional settings.

Other researchers have chosen to forgo a general solution for ML methods, and explore specific methods. Imai and Ratkovic (2013) applies an adapted Support Vector Machine classifier to identify heterogeneity in the effects of voter mobilisation strategies and job training programs. An especially promising paper by Athey and Imbens (2016) proposes a method with very few assumptions using causal trees, and inferring results on the CATE based on observation membership in tree leaves.

Finally, several recent studies have chosen not to estimate the CATE directly, but to infer results on it by targeting similar measures. Semenova and Chernozhukov (2020) uses ML to evaluate the partial CATE by estimating a transformation of the outcome and projecting it on a pre-specified low-dimensional subset of the covariates. Semenova, Goldman, Chernozhukov and Taddy (2022) also targets the partial CATE, but through residualised outcomes and treatment. Critically though, these methods still partly rely on the consistency of ML in high-dimensional settings.

This paper follows the methodology of Chernozhukov et al. (2023). This method avoids making inference on CATE directly and uses ML methods to provide proxies of the CATE, and subsequently uses these proxies to estimate three measures: The BLP of the CATE on the ML proxy, the GATES, which are the average treatment effect by heterogeneity groups induced by the ML proxy, and the CLAN, which are the average characteristics of the members of the most and least affected groups induced by the ML proxy. Its approach requires few assumptions, is valid in high-dimensional settings, and is applicable regardless of the ML method. It relies on repeated data splitting and estimation, then aggregating the results, which allows it to avoid overfitting. The paper illustrates the method's practical use by applying it to an immunization trial in India.

The methodology is well-suited to evaluate the presence of heterogeneity in the effect of IPV reduction policies in Ethiopia. A study by Garcia-Moreno et al. (2006) found that lifetime physical or sexual IPV rates among Ethiopian women was 70.9%, the highest of all countries in the study. Prevalence is especially high among women who are rural, young, divorced or poor (Chernet & Cherie, 2020). A prior study of an intervention aimed at reducing IPV in rural Ethiopia found no significant changes in women's reporting of past year experience of IPV (Sharma et al., 2020). Consequences of IPV include increased rates of depression, anxiety (Ellsberg, Jansen, Heise, Watts & Garcia-Moreno, 2008), illness, injury and death (Campbell, 2002). Prior research to the effectiveness of programs to reduce IPV is extensive, with at least 95 such studies conducted in the 2010s (Picon et al., 2017). However, much of this research focuses on what factors make the treatments effective, while very few ask what factors make for a receptive participant, although there are some established findings. Jewkes, Flood and Lang (2014) finds that male engagement in preventive programs is critical. Extending these results, Doyle et al. (2023) finds that men who are younger, wage-employed and educated were more likely to drop out of one such program in Rwanda. Another study conducted in the Democratic Republic of the Congo found that the women who experienced the most severe physical or sexual IPV saw the largest decrease in both prevalence and severity after the intervention (Gurbuz Cuneo, 2023). Subgroup analysis for an intervention in rural Côte d'Ivoire found that women who were married as child brides were significantly more likely to report physical and/or sexual violence after the program than women who married in adulthood (Falb et al., 2015).

3 Data

For our empirical application, we use the dataset collected for Sharma et al. (2020). The purpose of this study was to evaluate the Unite for a Better Life (UBL) trial, which consisted of set of gender-transformative education sessions held in the context of a traditional coffee ceremony. The primary purpose of the trial was to reduce physical and sexual IPV, as well as to reduce the spread of HIV, with the promotion of more equal relations among genders as a secondary objective. The experimental design is as follows. 64 villages in 4 districts in the Garuge zone of Ethiopia were randomly selected and subsequently allocated in one of four equal-sized treatment arms: one in which only the men would participate in the sessions, one in which only the woman would, one in which the couple would, and a control arm in which no UBL sessions would take place. This allocation was stratified at the district level. In each village, households containing a married or cohabitating woman of 18 to 49 years old were randomly selected and screened for participation. This process would continue until each village had 106 participating households. In polygamous households, one woman was chosen randomly. In each village, 80% of households were randomly selected to actually participate in the sessions, in order to examine intra-village spillover effects.

Between December 2014 and March 2015, a baseline survey was conducted with a single member of the 6770 participating households. Baseline variables include whether the household was polygamous, how long the couple had been married, their amount of living children, an index of the total value of their assets, which wealth quintile they were in, the age and highest attained level of education for both the respondent and his/her spouse, and the religion of the respondent. The UBL program was conducted between April 2015 and October 2015. It consisted of 14 participatory sessions on, among others, gender roles, sexuality, violence, and conflict resolution. For a detailed overview of the contents of each session, we refer to Sharma et al. (2020). Approximately 24 months after the end of the intervention, an endline survey was complete by both the with the participants and their partners, containing questions on their past year experience or perpetration of IPV, attitudes towards gender roles, and their practice of safe sex. The final available obtained data set includes $n = 6045$ observations.

We conduct our analysis by examining three different types of IPV as our outcome variable Y_i: physical, sexual or emotional. In each case, this variable is a binary indicator on whether the woman in the household reported having experienced this type of violence in the 12 months prior to the endline survey. Table 2 in Sharma et al. (2020) provides an overview of what types of behaviour qualified as these forms of violence. The treatment D_i is a binary variable on whether the observation was part of a treatment arm, or a control arm. In our analysis, each treatment arm is compared separately against the control arm. All observations i where either Y_i or D_i is missing are dropped from the analysis. The covariates Z_i are vectors of the aforementioned variables from the baseline survey. The non-categorical variables among these are aggregated into ordinal categories, Table 3 in Sharma et al. (2020) gives an overview on how the categories are coded, as well as the distribution of observations over the treatment arms. For missing values, we follow the example of Chernozhukov et al. (2023) and add control indicators to our covariates, while setting the missing values to 0. This may introduce some bias in our findings, but since there are very few occurrences, this likely is negligible. Finally, Z_i also contains a village specific indicator for account for possible local effects.

4 Methodology

4.1 Model

We derive the following setup from Rosenbaum and Rubin (1983). Let our data set be Data = $(Y_i, D_i, Z_i)_{i=1}^N$, where Y_i denotes our outcome variable, D_i is a binary variable denoting whether observation i has received treatment, and let Z_i be a vector of covariates. Furthermore, assume that all observations are independent and identically distributed, and that the treatment assignment is only dependent on a (sub)vector of Z. We denote this propensity score as

$$p(Z) := \mathbb{P}[D = 1|Z]. \tag{1}$$

If we define $Y(0)$ as the set of outcome variables of untreated observations, and $Y(1)$ as the set of outcomes of treated observations, we then define the Baseline Conditional Average (BCA) as

$$b_0(Z) := \mathbb{E}[Y(0)|Z] \tag{2}$$

and the Conditional Average Treatment Effect (CATE) as

$$s_0(Z) := \mathbb{E}[Y(1)|Z] - \mathbb{E}[Y(0)|Z]. \tag{3}$$

Then, the BCA and CATE can be identified through the regression

$$Y = b_0(Z) + Ds_0(Z) + U \tag{4}$$

with $\mathbb{E}[U|Z] = 0$.

4.2 Estimation Targets

Since the functional forms of b_0 and s_0 can be quite complex and ML techniques are not guaranteed to produce consistent estimators unless further assumptions are made on the functional forms, (Chernozhukov et al., 2023) propose targeting alternative measures which can be used for inference and policy recommendation.

4.2.1 Best Linear Predictor

The Best Linear Predictor of CATE $s_0(Z)$ by proxy $S(Z)$ is given by the solution to

$$\min_{b_1,b_2} \mathbb{E}[s_0(Z) - b_1 - b_2 S(Z)]^2, \tag{5}$$

which, if it exist, is

$$\text{BLP}[s_0(Z)|S(Z)] = \beta_1 - \beta_2(S_Z - \mathbb{E}[S(Z)]) \tag{6}$$

with $\beta_1 = \mathbb{E}[s_0(Z)]$ and $\beta_2 = \text{Cov}[s_0(Z), S(Z)]/\text{Var}[S(Z)]$. Estimating BLP has two main benefits. First, by design it should be a better estimator of $s_0(Z)$ than $S(Z)$, in the sense that it improves the mean squared error. Second, it does this approximation by way of splitting the CATE into an average treatment effect (ATE) β_1, and a heterogeneous loading parameter (HET) β_2. Specifically, it can be shown that if $S(Z)$ is uncorrelated with $s_0(Z)$, then $\beta_2 = 0$, and also that if $s_0(Z)$ is constant, that is that there is no heterogeneity in the treatment effect, then $\beta_2 = 0$. Thus testing if $\beta_2 = 0$ can tell us whether there is heterogeneity in the treatment effect and whether $S(Z)$ is a relevant predictor of $s_0(Z)$. Identification of the BLP can be done through Weighted Residual BLP:

$$Y = \alpha' X_1 + \beta_1(D - p(Z)) + \beta_2(D - p(Z))(S(Z) - \mathbb{E}[S(Z)]) + \varepsilon, \tag{7}$$

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

$$\begin{aligned} w(Z) :=& p(Z)(1 - p(Z))^{-1} \\ X :=& (X_1', X_2')' \\ X_1 =& [1, B(Z), p(Z), p(Z)S(Z)]' \\ X_2 =& [D - p(Z), (D - p(Z))(S(Z) - \mathbb{E}[S(Z)])]'. \end{aligned}$$

Under some weak assumptions this formula correctly identifies β_1 and β_2. Estimation can then be done via the empirical alternative

$$Y_i = \hat{\alpha}' X_{1i} + \hat{\beta}_1(D_i - p(Z_i)) + \hat{\beta}_2(D_i - p(Z_i))(S(Z_i) - \mathbb{E}_{N,M}[S(Z_i)]) + \hat{\varepsilon}_i, \quad i \in M, \tag{8}$$

with $\mathbb{E}_{N,M}[w(Z_i)\hat{\varepsilon}_i X_i] = 0$, where $\mathbb{E}_{N,M}[\cdot]$ is the empirical expectation with respect to M, which denotes the set of observations included in the main sample.

4.2.2 Sorted Grouped Average Treatment Effects

The Sorted Group Average Treatment Effects are defined as the expected treatment effects within groups G_k for $k = 1, ..., K$, where K denotes the amount of groups, or algebraically,

$$\gamma_k := \mathbb{E}[s_0(Z)|G_k], \quad \text{for } k = 1, ..., K. \tag{9}$$

To explore whether there is heterogeneity in the treatment effects, these groups can be constructed by sorting our observations by their ML proxies values of the CATE, and subsequently dividing them into quantiles based on these values. We thus choose $G_k = \{S(Z) \in I_k\}$ where $I_k := [l_{k-1}, l_k)$, with $-\infty = l_0 < l_1 < ... < l_K = \infty$. Under some weak assumptions, we can identify the GATES parameters by weighed residual GATES:

$$Y = \alpha' X_1 + \sum_{k=1}^{K} \gamma_k (D - p(Z))\mathbb{1}(G_k) + \varepsilon, \tag{10}$$

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

$$\begin{aligned} W :=& (X_1', W_1')' \\ X_1 :=& (B(Z), p(Z)\{\mathbb{1}(G_k)\}_{k=1}^K)' \\ W_2 :=& ((D - p(Z))\{\mathbb{1}(G_k)\}_{k=1}^K)'. \end{aligned}$$

We can then estimate GATES through its empirical version

$$Y_i = \hat{\alpha}_0' X_{1i} + \hat{\gamma} W_{2i} + \hat{\varepsilon}_i \tag{11}$$

where $\hat{\gamma} = (\hat{\gamma}_1, .., \hat{\gamma_K})$.

4.2.3 Classification Analysis

Let G_1 and G_K be the least and most affected groups by the treatment respectively. It may be useful to analyse the characteristics of the members of these groups, as significant differences in these characteristics could hint at the underlying forces driving the heterogeneity in the treatment effect. Let $g(Y, D, Z)$ be some vector of the observed data. Then the Classification Analysis (CLAN) parameters are

$$\delta_1 := \mathbb{E}[g(Y, D, Z)|G_1] \text{ and } \delta_K := \mathbb{E}[g(Y, D, Z)|G_K]. \tag{12}$$

Obtaining the empirical equivalent involves simply taking the group means of the observed data, or

$$\hat{\delta}_1 = \frac{\mathbb{E}_{N,M}[g(Y, D, Z)G_{1,i}]}{\mathbb{E}_{N,M}[G_{1,i}]} \text{ and } \hat{\delta}_K = \frac{\mathbb{E}_{N,M}[g(Y, D, Z)G_{K,i}]}{\mathbb{E}_{N,M}[G_{K,i}]}, \tag{13}$$

where $\mathbb{E}_{N,M}[\cdot]$ is the empirical expectation with respect to M, and $G_{k,i} = \mathbb{1}\{S(Z) \in I_k\}$, where $I_k = (l_{k-1}, l_k)$, where l_k is the (k/K)-th quantile of $\{S(Z_i)\}_{i \in M}$.

4.2.4 Goodness of Fit Measures

Since different ML methods usually result in different ML proxies, it is useful to construct measures to compare their effectiveness. Chernozhukov et al. (2023) propose to two measures, one based on BLP

$$\Lambda := \beta_2^2 \text{Var}(S(Z)), \tag{14}$$

and one based on GATES,

$$\bar{\Lambda} := \sum_{k=1}^{K} \gamma_k^2 \mathbb{P}(S(Z) \in I_k). \tag{15}$$

Both measures quantify the ability of $S(Z)$ to explain variation in $s_0(Z)$: maximising Λ is equivalent to maximising the R^2 of a regression of $s_0(Z)$ on $S(Z)$, while maximising $\bar{\Lambda}$ is equivalent to maximising the R^2 of $s_0(Z)$ on $\bar{S}(Z)$, where $\bar{S}(Z) = \sum_{k=1}^{K} \gamma_k \mathbb{1}(S_Z \in I_k)$. Their empirical analogs are given by

$$\hat{\Lambda} = \hat{\beta}_2^2 \mathbb{E}_{N,M}[S(Z_i) - \mathbb{E}_{N,M}[S(Z_i)]]^2 \text{ and } \hat{\bar{\Lambda}} = \sum_{k=1}^{K} \hat{\gamma}_k^2 \mathbb{E}_{N,M} \mathbb{1}(S(Z_i) \in I_k). \tag{16}$$

4.3 Aggregation

Since the ML proxies are trained on an auxiliary sample and then used for estimation on the main sample, it follows that all point estimates, p-values and confidence intervals for the parameters described above are dependent on this random split. To reduce this dependence, we perform this random splitting process N_S times, after which we aggregate the results. We will do so by taking the median values of these random variables. The validity of the inferences made from these medians is dependent on quite a few assumptions and regularity conditions, but these are outside of the scope of this paper. For a detailed discussion we refer to Chernozhukov et al. (2023).

4.4 Algorithm

Our methodology can now be summarized in the following algorithm:

0. Given the data, fix a number of splits N_S, a significance score α and a set of ML methods.
1. Generate N_S random splits of the data into a main sample $\text{Data}_M := \{(Y_i, D_i, Z_i)\}$ and an auxiliary sample A. Over each split perform the following steps:
   a. Train each ML-method on A to get proxies $B(\cdot)$ and $S(\cdot)$ of $b_0(\cdot)$ and $s_0(\cdot)$ respectively. For each $i \in M$, calculate $B(Z_i)$ and $S(Z_i)$.
   b. Estimate the Best Linear Predictor (BLP) parameters in M
   c. Estimate the Sorted Group Average Treatment Effects (GATES) parameters in M
   d. Estimate the Classification Analysis (CLAN) parameters in M
   e. Compute goodness-of-fit values for each ML method
2. Aggregate goodness-of-fit measures to choose an optimal method
3. Compute and aggregate point estimates, p-values and confidence intervals.

5 Results

We set our number of splits $N_S = 100$ and significance $\alpha = 0.05$. For our GATES and CLAN parameters, we group our observations in quintiles ($K = 5$). As our ML-methods we choose Elastic Net, Gradient Boosted Trees and Random Forests. Additionally, to split our sample into a main and an auxiliary sample, we use stratification, such that half the households in each village end up in each sample.

Table 1 shows the goodness-of-fit measures for each combination of treatment and outcome variable

	Women's UBL			Men's UBL			Couples' UBL		
	EN	GBDT	RF	EN	GBDT	RF	EN	GBDT	RF
Physical									
Best BLP	**0.00103**	0.00029	0.00029	0.00016	0.00025	**0.00029**	**0.00059**	0.00023	0.00043
Best GATES	**0.00345**	0.00227	0.00274	0.00202	0.00224	**0.00239**	0.00226	0.00183	**0.00258**
Sexual									
Best BLP	**0.01818**	0.00124	0.00590	**0.01866**	0.00076	0.00333	**0.01959**	0.00158	0.00347
Best GATES	**0.01818**	0.00539	0.00899	**0.01865**	0.00452	0.00635	**0.02292**	0.00682	0.00730
Emotional									
Best BLP	**0.03112**	0.00845	0.01069	**0.02077**	0.00314	0.00752	**0.04128**	0.01068	0.01346
Best GATES	**0.02073**	0.00918	0.01226	**0.02107**	0.00829	0.01153	**0.02229**	0.01059	0.01453

Table 1: Estimated performance measures of ML proxies

	Elastic Net		Random Forest	
	ATE (β_1)	HTE (β_2)	ATE (β_1)	HTE (β_2)
Physical				
Estimate	0.018	0.028	0.021	-0.092
CI	[-0.025, 0.063]	[-0.788, 0.842]	[-0.023, 0.065]	[-0.526, 0.354]
p-value	0.396	0.936	0.338	0.673
Sexual				
Estimate	-0.041	1.122	-0.028	0.428
CI	[-0.091, 0.012]	[0.669, 1.574]	[-0.079, 0.022]	[0.030, 0.843]
p-value	0.130	<0.001	0.274	0.035
Emotional				
Estimate	-0.051	1.010	-0.046	0.556
CI	[-0.105, 0.001]	[0.517, 1.504]	[-0.098, 0.005]	[0.161, 0.936]
p-value	0.055	<0.001	0.077	0.005

Table 2: Best Linear Predictor parameter estimates men's UBL.

Elastic Net is best able to capture the heterogeneity in the data in all but three situations, in which Random Forests works better. We will continue analysis with these two methods to see if our results are robust, but, excluding in the cases of perpetration of physical violence after men's or couples' UBL, the results incurred by Elastic Net should be considered more trustworthy than those by Random Forest.

Table 2 displays BLP estimates for the men's UBL treatment arm. Tables 11 and 12 for the women's and couples' arms are in appendix B.

Across all treatment arms we find the same pattern. The ATE is at the 5%-significance level not different from 0 for any of the types of IPV. When looking at 10% significance, we robustly find that the reporting of emotional IPV declined for the spouses in the men's UBL treatment arm. Additionally, reported sexual IPV declined in the couples' UBL arm, although only when Elastic Net was used for estimation. This is somewhat in line with the findings in Sharma et al. (2020), who neither found these effects to be significant at the 5%-level, although at the 10% they found that sexual IPV declined in the men's UBL treatment arm. When examining heterogeneity, we find the HET not to be statistically significant for physical IPV, indicating that the

	Elastic Net			Random Forest		
	γ_1	γ_5	$\gamma_5 - \gamma_1$	γ_1	γ_5	$\gamma_5 - \gamma_1$
Physical						
Estimate	0.004	0.026	0.018	0.024	0.006	-0.021
CI	[-0.096, 0.107]	[-0.074, 0.123]	[-0.137, 0.168]	[-0.076, 0.126]	[-0.089, 0.103]	[-0.163, 0.118]
p-value	0.917	0.597	0.779	0.618	0.867	0.771
Sexual						
Estimate	-0.236	0.141	0.388	-0.114	0.059	0.165
CI	[-0.349, -0.125]	[0.002, 0.277]	[0.204, 0.573]	[-0.229, 0.003]	[-0.055, 0.173]	[0.000, 0.328]
p-value	**0.000**	**0.042**	**0.000**	0.056	0.305	**0.047**
Emotional						
Estimate	-0.231	0.152	0.361	-0.122	0.099	0.217
CI	[-0.346, -0.118]	[0.009, 0.291]	[0.172, 0.555]	[-0.234, -0.007]	[-0.023, 0.220]	[0.053, 0.384]
p-value	**0.000**	**0.035**	**0.000**	**0.038**	0.112	**0.010**

Table 3: GATES parameters for Men's UBL.

treatment effect is fairly homogeneous across the included covariates. For sexual and emotional IPV we robustly find that the used ML-proxies capture a significant amount heterogeneity in the treatment effect across all arms.

Table 3 displays the GATES parameters for the men's UBL. Women's and couples' UBL are shown in Tables 13 and 14 in the Appendix. Before we analyse these numbers, a quick word on terminology. CLAN in the original Chernozhukov et al. (2023) paper tends to name G_1 and G_5 the least and most affected groups. As we see in this table, in many cases the GATES parameter is significant with opposite signs for these groups. Since the intended purpose of the intervention was to reduce IPV, we will hereafter refer to these groups as the most positively affected group G_1 and the most negatively affected group G_5. The findings are reasonable consistent across treatment arms and are in line with the BLP parameters. Physical violence is not significantly affected by the UBL-program in any quintile. When looking at sexual IPV, just as when examining BLP, we robustly find significant treatment effects in the most positively and the most negatively affected quintiles. The underlying parameters show that there may be some problems with the UBL-program. When using Elastic Net, women's UBL, men's UBL and couples' UBL are estimated to have caused a 18%, 24% and 21% reduction of sexual IPV for the households in G_1 respectively, corresponding to respective increases in the effectiveness of the treatment by 20, 20, and 17 percentage-points when compared to the ATE. However, for the members of G_5, the intervention is estimated to have actually significantly increased sexual IPV. Except for the significance of each $\gamma_5 - \gamma_1$ and the significance of γ_5 for women's UBL, these findings are not present if Random Forest is used as ML-proxy. They should nevertheless be taken seriously, as Elastic Net is the method with larger goodness of fit measures in all these cases. When examining emotional IPV, UBL seems more effective. The treatment significantly reduced reported emotional IPV in the most positively affected group, and although the most negatively affected groups all reported an increase, in each case except for men's UBL proxied by Elastic Net, these increases are not significant at the 5%-level.

Before we continue with CLAN, we shortly examine the relative importance of the covariates

	Women's UBL Household	Village	Men's UBL Household	Village	Couples' UBL Household	Village
Sexual						
Best BLP	0.0005	0.0458	0.0003	0.0453	0.0010	0.0423
Best GATES	0.0040	0.0332	0.0034	0.0313	0.0056	0.0358
Emotional						
Best BLP	0.0029	0.0606	0.0011	0.0508	0.0002	0.0747
Best GATES	0.0063	0.0246	0.0044	0.0281	0.0044	0.0310

Table 4: Comparison of ML proxy performance between household and village characteristics.

obtained in the baseline survey with the importance of local effects. We will do this by comparing the goodness of fit values, using the same methodology as laid out in the Moroccan MFI application. In subsequent analysis, physical IPV is dropped from the outcome variables, as neither BLP nor GATES shows signs of captured heterogeneity in the treatment effects. Additionally, only the results incurred by the usage of Elastic net will be examined, as it outperforms Random Forest in each of the remaining cases. Table 4 shows the goodness of fit measures for the household and village effects.

We find that across all situations, the local village effects are better at capturing heterogeneity than the household demographics. It should be mentioned here that household sampling in each village was not done throughout the entire village: within each village first a subvillage was sampled, from which all households for the village were subsequently sampled. This would result in most participants living in the same community and, since the UBL-program consists of communal sessions with active participation, individual behavioral effects induced by the intervention could very well largely be shaped through collective factors, such as the effectiveness of the session facilitators or the overall attitude of the participants towards the program.

Table 5 shows the CLAN-estimates for the set of covariates where difference-estimate $\delta_5 - \delta_1$ is either significantly positive for both sexual and emotional IPV, or significantly negative for both. The full CLAN-estimates are in Tables 15-17 in the Appendix.

Before analysing these numbers, it would be good to have a quick disclaimer that these effect should not be interpreted causally, as the ML-proxies may very well target other variable correlated with the outcomes. First, we see that, in order to have a policy which significantly reduces both sexual and emotional IPV, it is not sufficient to just focus on women: men have to participate as well, either alone or as a couple. These treatment arms show some differences. The program with only men participating seems best to target a certain age cohort, with it having the best result on young, shortly married couples with few children. The effect of the couples' sessions looks to be more dependent on the socioeconomic status of the couple, with poorer couples with a less educated man benefiting most. Religion also plays a significant role here, with Muslim couples forming 58% of the most positively affected group and 40% of the most negatively affected group, while Ethiopian Orthodox Christians form 35% of the most positively and 52% of the most negatively affected groups. The full CLAN Tables show some additional finding. The women from polygamous households which followed the sessions, either with or without their partners, make up a larger part of the group seeing a reduction in reported sexual

	Sexual IPV			Emotional IPV		
	δ_1	δ_5	$\delta_5 - \delta_1$	δ_1	δ_5	$\delta_5 - \delta_1$
Men's UBL						
Cat. Marriage Length						
Estimate	2.639	3.507	0.920	2.669	3.421	0.706
CI	[2.503, 2.771]	[3.344, 3.672]	[0.713, 1.131]	[2.535, 2.815]	[3.261, 3.581]	[0.490, 0.925]
p-value			0.000			0.000
Living children						
Estimate	3.967	4.442	0.483	3.411	4.812	1.404
CI	[3.702, 4.236]	[4.183, 4.736]	[0.094, 0.875]	[3.161, 3.679]	[4.532, 5.084]	[1.009, 1.799]
p-value			0.013			0.000
Asset Index						
Estimate	3.364	4.019	0.673	3.519	3.956	0.415
CI	[3.102, 3.620]	[3.738, 4.300]	[0.294, 1.054]	[3.227, 3.806]	[3.669, 4.238]	[0.014, 0.808]
p-value			0.000			0.031
Age Woman						
Estimate	1.656	1.941	0.286	1.556	2.006	0.456
CI	[1.573, 1.742]	[1.856, 2.026]	[0.164, 0.409]	[1.474, 1.638]	[1.920, 2.091]	[0.334, 0.576]
p-value			0.000			0.000
Age Man						
Estimate	2.264	2.483	0.212	2.137	2.537	0.388
CI	[2.179, 2.350]	[2.408, 2.559]	[0.100, 0.324]	[2.052, 2.226]	[2.465, 2.609]	[0.274, 0.505]
p-value			0.000			0.000
Couples' UBL						
Asset Index						
Estimate	3.450	3.996	0.533	3.415	4.250	0.841
CI	[3.178, 3.726]	[3.697, 4.295]	[0.129, 0.933]	[3.155, 3.686]	[3.954, 4.539]	[0.434, 1.234]
p-value			0.007			0.000
Wealth Index						
Estimate	-0.191	0.261	0.420	-0.098	0.287	0.397
CI	[-0.294, -0.089]	[0.091, 0.437]	[0.213, 0.623]	[-0.210, 0.024]	[0.119, 0.462]	[0.185, 0.605]
p-value			0.000			0.000
Schooling Man						
Estimate	1.561	1.756	0.196	1.612	1.730	0.110
CI	[1.492, 1.631]	[1.679, 1.833]	[0.092, 0.295]	[1.538, 1.688]	[1.654, 1.806]	[0.007, 0.213]
p-value			0.000			0.035
Muslim						
Estimate	0.582	0.402	-0.193	0.654	0.268	-0.393
CI	[0.523, 0.640]	[0.342, 0.459]	[-0.275, -0.110]	[0.598, 0.711]	[0.216, 0.321]	[-0.469, -0.316]
p-value			0.000			0.000
Orthodox						
Estimate	0.357	0.519	0.174	0.301	0.632	0.316
CI	[0.298, 0.413]	[0.459, 0.578]	[0.091, 0.256]	[0.247, 0.356]	[0.575, 0.690]	[0.316, 0.396]
p-value			0.000			0.000

Table 5: CLAN parameters of covariates with significant difference and equal-signed effects on sexual and emotional IPV.

IPV than they do in the group where this increased. Female education seems to be a double-edged sword here, as for emotional IPV more highly-educated women are underrepresented in the most negatively affected groups, but are overrepresented in these groups for reporting sexual IPV.

6 Conclusion

The goal of this paper is to examine the possibility of heterogeneity in the effects of the Unite for a Better Life trial on women's reported experience physical, sexual and emotional Intimate Partner Violence in Ethiopia, and identify possible drivers of this heterogeneity. It find that the treatment effect on both sexual and emotional IPV exhibits heterogeneity, while this is not found for physical IPV. Quintile analysis shows that for sexual IPV, the heterogeneity is driven by both a reduction in reported violence in the most positively affected quintile, and an increase in the most negatively affected quintile. Emotional IPV finds similar positive effects across all treatment arms, while only finding significant negative effects in the men's treatment arm. In the men's treatment arm, couples which were married for a shorter time, with fewer living children, fewer assets, and who consisted of a younger man and woman, were significantly more represented in the most positively affected quintile than in the most negatively affected quintile for both sexual and emotional IPV. In the couple's treatment arm, this effect was found for households which had fewer assets, less wealth, contained a lower educated man, and adhered to Islam. In the women's treatment arm, none of the covariates produced such a relation. All this showcases the effectiveness of our methodology in inferring heterogeneous relations in a way which is both extensive, as well as academically responsible. Where the original analysis of the UBL program did not find significant effects of the intervention on the prevalence of IPV, this paper finds that certain demographics were more likely to benefit from the treatment, while in others the treatment had the opposite intended effect. Future interventions can be planned with this in mind: resources for such programs are often slim, especially in the impoverished regions where IPV is most common. Knowing which groups to target and which to avoid could make progress more efficient.

This study comes with two main limitations. The first is inherent to research in the field of domestic violence, namely that the outcome of interest is self-reported. Fear of reporting and victim blaming are but a few of the reasons why reported statistics may not accurately represent the true numbers. Although the endline survey included thorough definitions, most of the problems on data collection remain relevant. The results in this paper should therefore not be interpreted as definitive measures on IPV, but only on its reporting.

The second limitation comes with the loss of power associated with ML methods. Since there are few pre-established driving factors of heterogeneity in the literature, this loss was necessary trade-off which allowed for the identification of new possible driving factors. Subsequent research could parameterise these factors, either to see if they still hold up as significant, or as to reduce the loss of power in future analysis.

References

Angelucci, M., Karlan, D. & Zinman, J. (2015, January). Microcredit impacts: Evidence from a randomized microcredit program placement experiment by compartamos banco. *American Economic Journal: Applied Economics*, *7*(1), 151–82. Retrieved from `https://www.aeaweb.org/articles?id=10.1257/app.20130537` doi: 10.1257/app.20130537

Athey, S. & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, *113*(27), 7353–7360. Retrieved from `http://dx.doi.org/10.1073/pnas.1510489113` doi: 10.1073/pnas.1510489113

Banerjee, A., Duflo, E., Glennerster, R. & Kinnan, C. (2015, January). The miracle of microfinance? evidence from a randomized evaluation. *American Economic Journal: Applied Eco-*

nomics, *7*(1), 22–53. Retrieved from `https://www.aeaweb.org/articles?id=10.1257/app.20130533` doi: 10.1257/app.20130533

Belek, A. & Jean Marie, A. (2021). The impact of microfinance services on the efficiency of family farms in sub-saharan africa: the case of cameroon. *Agricultural Finance Review*, *81*(1), 21-38. doi: 10.1108/AFR-02-2019-0021

Belloni, A., Chernozhukov, V. & Kato, K. (2014). Uniform post-selection inference for least absolute deviation regression and other z-estimation problems. *Biometrika*, *102*(1), 77–94. Retrieved from `http://dx.doi.org/10.1093/biomet/asu056` doi: 10.1093/biomet/asu056

Campbell, J. (2002). Health consequences of intimate partner violence. *Lancet*, *359*, 1331-1336. doi: https://doi.org/10.1016/S0140-6736(02)08336-8

Chernet, A. G. & Cherie, K. T. (2020). Prevalence of intimate partner violence against women and associated factors in ethiopia. *BMC Women's Health*, *10*.

Chernozhukov, V., Demirer, M., Duflo, E. & Fernández-Val, I. (2023). *Generic machine learning inference on heterogenous treatment effects in randomized experiments, with an application to immunization in india.*

Crépon, B., Devoto, F., Duflo, E. & 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), 123–150. Retrieved 2024-06-12, from `http://www.jstor.org/stable/43189515`

Devries, K. M., Mak, J. Y. T., García-Moreno, C., Petzold, M., Child, J. C., Falder, G., ... Watts, C. H. (2013). The global prevalence of intimate partner violence against women. *Science*, *340*(6140), 1527-1528. Retrieved from `https://www.science.org/doi/abs/10.1126/science.1240937` doi: 10.1126/science.1240937

Dezeure, R., Bühlmann, P. & Zhang, C.-H. (2016). *High-dimensional simultaneous inference with the bootstrap.* Retrieved from `https://arxiv.org/abs/1606.03940`

Diamoutene, A. & Jatoe, J. (2021). Access to credit and maize productivity in mali. *Agricultural Finance Review*, *81*(3), 458-477. doi: 10.1108/AFR-05-2020-0066

Doyle, K., Levtov, R., Karamage, E., Rakshit, D., Kazimbaya, S., Sayinzoga, F., ... Barker, G. (2023, 10). Long-term impacts of the bandebereho programme on violence against women and children, maternal health-seeking, and couple relations in rwanda: a six-year follow-up of a randomised controlled trial. *eClinicalMedicine*, *64*, 102233. doi: 10.1016/j.eclinm.2023.102233

Ellsberg, M., Jansen, H., Heise, L., Watts, C. & Garcia-Moreno, C. (2008, 05). Who multi-country study, women's health and domestic violence against women study team. intimate partner violence and women's physical and mental health in the who multi-country study on women's health and domestic violence: an observational study. *Lancet*, *371*, 1165-72. doi: 10.1016/S0140-6736(08)60522-X

Falb, K. L., Annan, J., Kpebo, D., Cole, H., Willie, T., Xuan, Z., ... Gupta, J. (2015). Differential impacts of an intimate partner violence prevention program based on child marriage status in rural côte d'ivoire. *Journal of Adolescent Health*, *57*(5), 553-558. Retrieved from `https://www.sciencedirect.com/science/article/pii/S1054139X15003018` doi:

https://doi.org/10.1016/j.jadohealth.2015.08.001

Friedman, J., Hastie, T. & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1). Springer.

Garcia-Moreno, C., Jansen, H. A., Ellsberg, M., Heise, L. & Watts, C. H. (2006). Prevalence of intimate partner violence: findings from the who multi-country study on women's health and domestic violence. *The Lancet*, *368*(9543), 1260-1269. Retrieved from https://www.sciencedirect.com/science/article/pii/S0140673606695238 doi: https://doi.org/10.1016/S0140-6736(06)69523-8

Genovese, C. R. & Wasserman, L. (2007). *Adaptive confidence bands.* Retrieved from https://arxiv.org/abs/math/0701513

Giné, E. & Nickl, R. (2010). Confidence bands in density estimation. *The Annals of Statistics*, *38*(2). Retrieved from http://dx.doi.org/10.1214/09-AOS738 doi: 10.1214/09-aos738

Gurbuz Cuneo, V. J. K. E. P. R. S. F. K. . K. R., A. (2023). Prevention, cessation, or harm reduction: Heterogeneous effects of an intimate partner violence prevention program in eastern democratic republic of the congo. *PloS One*, *18*.

Hansen, C., Kozbur, D. & Misra, S. (2018). *Targeted undersmoothing.* Retrieved from https://arxiv.org/abs/1706.07328

Imai, K. & Ratkovic, M. (2013). Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics*, *7*(1). Retrieved from http://dx.doi.org/10.1214/12-AOAS593 doi: 10.1214/12-aoas593

Jewkes, R., Flood, M. & Lang, J. (2014, 11). From work with men and boys to changes of social norms and reduction of inequities in gender relations: A conceptual shift in prevention of violence against women and girls. *Lancet*, *385*. doi: 10.1016/S0140-6736(14)61683-4

Nguimkeu, P. (2014). A structural econometric analysis of the informal sector heterogeneity. *Journal of Development Economics*, *107*, 175-191. Retrieved from https://www.sciencedirect.com/science/article/pii/S0304387813001739 doi: https://doi.org/10.1016/j.jdeveco.2013.12.001

Nordjo, R. & Adjasi, C. (2020). The impact of credit on productivity of smallholder farmers in ghana. *Agricultural Finance Review*, *80*(1), 91-109. doi: 10.1108/AFR-10-2018-0096

Olanrewaju, O. R., O. & Fasakin, J. (2021). The anchor borrowers programme and youth rice farmers in northern nigeria. *Agricultural Finance Review*, *81*(2), 222-236. doi: 10.1108/AFR-03-2020-0039

Picon, M., Rankin, K., Ludwig, J., Delaney, A., Holst, A. & Sabet, S. (2017, 06). *Intimate partner violence prevention: An evidence gap map.*

Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. Retrieved 2024-06-05, from http://www.jstor.org/stable/2335942

Sagbo, N. & Kusunose, Y. (2021). Does experience with agricultural loans improve farmers' well-being? evidence from benin. *Agricultural Finance Review*, *81*(4), 503-519. doi: 10.1108/AFR-06-2020-0082

Semenova, V. & Chernozhukov, V. (2020, 08). Debiased machine learning of conditional average treatment effects and other causal functions. *The Econometrics Journal*, *24*(2), 264-289.

Retrieved from https://doi.org/10.1093/ectj/utaa027 doi: 10.1093/ectj/utaa027

Semenova, V., Goldman, M., Chernozhukov, V. & Taddy, M. (2022). *Estimation and inference on heterogeneous treatment effects in high-dimensional dynamic panels under weak dependence.* Retrieved from https://arxiv.org/abs/1712.09988

Sharma, V., Leight, J., Verani, F., Tewolde, S. & Deyessa, N. (2020). Effectiveness of a culturally appropriate intervention to prevent intimate partner violence and hiv transmission among men, women, and couples in rural ethiopia: Findings from a cluster-randomized controlled trial. *PLoS medicine*.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics*, *10*, 1040–1053.

Tarozzi, A., Desai, J. & Johnson, K. (2015, January). The impacts of microcredit: Evidence from ethiopia. *American Economic Journal: Applied Economics*, *7*(1), 54–89. Retrieved from https://www.aeaweb.org/articles?id=10.1257/app.20130475 doi: 10.1257/app.20130475

Teye, E. S. & Quarshie, P. T. (2022). Impact of agricultural finance on technology adoption, agricultural productivity and rural household economic wellbeing in ghana: a case study of rice farmers in shai-osudoku district. *South African Geographical Journal*, *104*(2), 231–250. Retrieved from https://doi.org/10.1080/03736245.2021.1962395 doi: 10.1080/03736245.2021.1962395

Weber, R. & Musshoff, O. (2012). Is agricultural microcredit really more risky? evidence from tanzania. *Agricultural Finance Review*, *72*(3), 416-435. doi: 10.1108/00021461211277268

Zhao, Q., Small, D. S. & Ertefaie, A. (2021). *Selective inference for effect modification via the lasso.* Retrieved from https://arxiv.org/abs/1705.08020

Appendices

A Availability of microcredit in rural Morocco

This appendix features an additional application of our methodology. It is a replication of the analysis performed in version 4 of Chernozhukov et al. (2023).

A.1 Background

Throughout the developing world, various experiments have been conducted to evaluate the effects of microcredit availability on economic activity. Overall, its effectiveness seems limited: various studies find that this access increases total borrowing (Angelucci, Karlan & Zinman, 2015; Banerjee, Duflo, Glennerster & Kinnan, 2015; Tarozzi, Desai & Johnson, 2015), but that long term effects on profit, consumption and income are insignificant (Angelucci et al., 2015; Banerjee et al., 2015; Crépon, Devoto, Duflo & Parienté, 2015). However, recent studies conducted on specific subgroups sketch a more nuanced picture. Households whose main economic activity involves agriculture are for instance less likely to receive loans than those that are not employed in this sector (Teye & Quarshie, 2022; Weber & Musshoff, 2012). Additionally, access to credit is found to significantly increase agricultural production (Teye & Quarshie, 2022; Nordjo & Adjasi, 2020; Diamoutene & Jatoe, 2021), especially among younger farmers (Olanrewaju & Fasakin, 2021; Belek & Jean Marie, 2021), as well as income (Sagbo & Kusunose, 2021). Additionally, Nguimkeu (2014) finds that access to microcredit for the informal sector reduces the percentage of the population affected by poverty and Banerjee et al. (2015) finds that prior business experience is a significant driver of both production and income. All this suggests that microcredit availability could very well have significant effects on specific subgroups.

We will attempt to identify these subgroups in the setting of a previous study constructed by Crépon et al. (2015). For the experiment, 162 villages in rural Morocco were chosen in which Al Amana, a microfinance institution (MFI), was interested in starting operations. These villages were subsequently divided in 81 pairs, based on similarities in number of households, accessibility to the center of the community, existing infrastructure, type of activities carried out by the households, and type of agriculture activities. One of the villages in each pair was randomly chosen to have Al Amara start operating between 2006 and 2007. Al Amara services in the rural areas in which this experiment was conducted mainly consisted of group liability loans ranging from 1000 MAD to 15000 MAD (US$124, US$1855). Two years after the start of the intervention, the endline survey was conducted on 5551 households. Before and during the intervention, no other MFIs were available to any of the included villages.

For our analysis we will evaluate the effect of access to MFIs on economic activity by looking at four different outcome variables Y_i: the amount of money borrowed, the output from self-employment activities, profit from self-employment activities, and monthly consumption. D_i is a binary variable indicating whether a household lives in a village where an MFI opened. Z_i includes vectors of various demographic and economic indicators, as well as 81 village pair fixed effects and indicators for missing observations.

We set our number of splits $N_S = 100$ and significance $\alpha = 0.05$. For our GATES and CLAN parameters, we group our observations in quintiles ($K = 5$). As our ML-methods we choose Elastic Net, Gradient Boosted Trees and Random Forests. Additionally, to split our sample into a main and an auxiliary sample, we use stratification, such that half the households in each village pair end up in each sample. Finally, some observations on household consumption were missing, so those observations were removed, retaining a sample of 5513 households.

A.2 Results

Table 6 displays the goodness of fit metrics for each method on each dependent variable.

	Elastic Net	GBDT	Ranfom Forest
Amount of loans			
Best BLP (Λ)	683019	374236	**1068682**
Best GATES ($\bar{\Lambda}$)	**2738043**	2141998	2652319
Output			
Best BLP (Λ)	**61028036**	9986077	34500649
Best GATES ($\bar{\Lambda}$)	**135656852**	87831789	129444078
Profit			
Best BLP (Λ)	20961681	4372081	**24015937**
Best GATES ($\bar{\Lambda}$)	35181167	18047699	**39948929**
Consumption			
Best BLP (Λ)	**14012**	12709.07	12543
Best GATES ($\bar{\Lambda}$)	**41847**	33561	35441

Table 6: Estimated performance measures of ML proxies

We see that, depending on the output variable and the chosen metric, either Elastic Net or Random Forests perform best, thus we will continue our analysis with these methods.

Table 7 shows the BLP of the CATE through the selected ML proxies. There are a couple of main results to infer from this table. First, it is encouraging to see that both our ML-methods agree on the significance of the parameters in each situation. Second, when looking at the Average Treatment Effect (ATE), we find that the opening of microfinance institutions significantly increases the amount of loans household take, but we cannot reject at the 5% level that the effect on total output, profits, or consumption is more than 0. REASONS? When inspecting the presence of heterogeneity in the effect of the access to MFIs, we can however see that not all groups are affected equally by this access: we can reject at the 5% significance level that there is no presence of heterogeneity for all researched output variables except consumption.

Table 8 shows the GATES parameters for the least and most affected groups, as well as their difference.

The GATES parameters expand on the story told by the BLP: the opening of MFIs only significantly increases the amount of loans, output, or profits of the most affected quintile, while the other households do not experience significant increases. When examining Figure ... we see

	Elastic Net		Random Forest	
	ATE (β_1)	HET (β_2)	ATE (β_1)	HET (β_2)
Amount of Loans				
Estimate	1079	0.213	1117	0.232
CI	[301, 1840]	[0.009, 0.403]	[348, 1875]	[0.063, 0.405]
p-value	0.005	0.039	0.003	0.008
Output				
Estimate	4750	0.220	5003	0.157
CI	[-1714, 11160]	[0.045, 0.397]	[-1561, 11280]	[-0.021, 0.341]
p-value	0.146	0.013	0.130	0.081
Profit				
Estimate	1725	0.248	1828	0.248
CI	[-1813, 5290]	[0.056, 0.436]	[-1782, 5289]	[0.061, 0.415]
p-value	0.323	0.009	0.310	0.007
Consumption				
Estimate	-57.8	0.161	-76.4	0.145
CI	[-197.4, 86.2]	[-0.043, 0.378]	[-219.5, 71.9]	[-0.050, 0.346]
p-value	0.442	0.116	0.310	0.137

Table 7: Best Linear Predictor parameter estimates

	Elastic Net			Random Forest		
	γ_1	γ_5	$\gamma_5 - \gamma_1$	γ_1	γ_5	$\gamma_5 - \gamma_1$
Amount of Loans						
Estimate	-291.1	2637.2	2763.3	532.6	2788.9	2355.1
CI	[-1978.7, 1486.2]	[897.5, 4314.7]	[288, 5100.1]	[-1204, 2012.2]	[1087.6, 4581.9]	[107.8 4595.8]
p-value	0.756	0.003	0.022	0.528	0.001	0.039
Output						
Estimate	-1730	20532	22464	258.2	21912.4	20697.8
CI	[-15772, 12051]	[6173, 34671]	[1376, 42322]	[-14456.4, 15170.3]	[7869.8, 36020.4]	[1227.4, 41002.9]
p-value	0.792	0.005	0.034	0.973	0.002	0.036
Profit						
Estimate	-1350	11262.5	11984	-1331	12334.1	13433.2
CI	[-9084.4, 6410.9]	[3419.2, 19215.3]	[1107.5, 23172.8]	[-9401.1, 6303.1]	[4390.3, 19928.9]	[2392.2, 24273.8]
p-value	0.732	0.005	0.03	0.733	0.002	0.015
Consumption						
Estimate	-364.93	67.01	374.13	-309.23	-22.11	286.35
CI	[-683.61, -24.56]	[-254.72, 384.92]	[-105.59, 814.37]	[-622.44, 14.83]	[-335.81, 304.96]	[-184.34, 729.17]
p-value	0.035	0.644	0.112	0.062	0.899	0.228

Table 8: GATES estimates for the least and most affected quintiles

	Elastic Net		Random Forest	
	Household	Village	Household	Village
Amount of loans				
Best BLP (Λ)	32827	963209	21680	930974
Best GATES ($\bar{\Lambda}$)	1691802	3327410	1719767	3223021

Table 9: Comparison of ML proxy performance between household and village characteristics on predicting the CATE of income.

that in multiple cases the point estimates for the least affected quintiles can even be negative, although these results are never significant. The main exception here is consumption. Here, the consumption of the least affected group is significantly lower after the intervention, while in none of the other quintiles a significantly increase occurs. This can possibly be explained in two ways. If a loan is taken, it may not cover the complete upfront cost of the investment the client wants to make, thus the client may reduce consumption to make up for the rest. Alternatively, if the client uses the loan to make an unproductive investment, he may need to reduce his consumption in order to pay it back. It should be noted that these explanations are not mutually exclusive, nor are they likely to be the only ones possible.

Before conducting CLAN, it is useful to remember that our covariates come in two groups. There are the household variables, which are individually present in our data set, and there is the categorical variable indicating in which village pair the household is located. Table 9 displays the goodness of fit measures if only one of either set of covariates is included in Z_i. The amount of loans is chosen as dependent variable. Table ... in Appendix B displays the full table containing all dependent variables.

Although these values do not have a readily available interpretation, it is noteworthy that the ML-proxies trained only on the village pair are better able to capture CATE than the ones trained on the household variables. It should therefore be kept in mind that when deciding where to open an MFI, if the purpose is to enhance economic activity overall, it is likely more useful to look at the village level characteristics such as infrastructure and the type of existing economic activities, rather than individual household characteristics. The large effect of the village-pair variable may also be explained through a more sociological lens. For instance, some MFIs may have a more competent staff than others, leading to an uptick in the amount of customers they serve, or alternatively, if many people in a village apply for a loan, this may motivate others in the village to do the same, creating a snowballing effect.

We continue with performing Classification Analysis. Since BLP and GATES do not provide significant evidence for the presence of heterogeneity when looking at consumption, we will not conduct this analysis for this dependent variable. Table ... shows the CLAN parameters of the covariates for which the difference between the most negatively and most positively affected groups ($\delta_5 - \delta_1$) is found to be significantly different than 0 by both the Elastic Net and Random Forest proxies. Tables ... contain the CLAN-parameters for all combinations of dependent variables and covariates (excluding village-pair fixed effects).

	Elastic Net			Random Forest		
	δ_1	δ_5	$\delta_5 - \delta_1$	δ_1	δ_5	$\delta_5 - \delta_1$
Amount of loans						
Number of hh. members						
Estimate	4.486	3.311	-1.223	4.261	2.692	-1.540
CI	[4.231, 4.760]	[3.041, 3.582]	[-1.611, -0.839]	[4.001, 4.533]	[2.393, 2.991]	[-1.945, -1.136]
p-value			<0.001			<0.001
Age head of hh.						
Estimate	39.344	30.464	-8.381	36.155	24.621	-11.174
CI	[37.304, 41.346]	[28.250, 32.620]	[-11.388, -5.374]	[34.127, 38.251]	[22.294, 26.928]	[-14.354, -8.024]
p-value			<0.001			<0.001
Self-emp. in animal husb.						
Estimate	0.467	0.340	-0.132	0.409	0.309	-0.100
CI	[0.426, 0.509]	[0.299, 0.378]	[-0.190, -0.074]	[0.368, 0.450]	[0.269, 0.347]	[-0.156, -0.043]
p-value			<0.001			0.001
Has borrowed money						
Estimate	0.226	0.157	-0.068	0.273	0.136	-0.136
CI	[0.192, 0.261]	[0.127, 0.188]	[-0.114, -0.023]	[0.235, 0.309]	[0.107, 0.164]	[-0.185, -0.089]
p-value			0.003			<0.001
Spouse self-emp.						
Estimate	0.128	0.033	-0.096	0.091	0.027	-0.065
CI	[0.099, 0.155]	[0.018, 0.047]	[-0.129, -0.064]	[0.067, 0.115]	[0.014, 0.041]	[-0.092, -0.039]
p-value			<0.001			<0.001
Output						
Number of memb. over 16						
Estimate	2.362	2.882	0.517	2.230	2.605	0.322
CI	[2.193, 2.518]	[2.665, 3.097]	[0.242, 0.797]	[2.048, 2.413]	[2.363, 2.835]	[0.015, 0.616]
p-value			<0.001			0.030
Non-agricultural self-emp.						
Estimate	0.055	0.274	0.218	0.124	0.241	0.111
CI	[0.036, 0.074]	[0.237, 0.311]	[0.176, 0.263]	[0.096, 0.151]	[0.205, 0.276]	[0.066, 0.155]
p-value			<0.001			<0.001
Spouse self-emp.						
Estimate	0.111	0.038	-0.071	0.078	0.040	-0.038
CI	[0.083, 0.135]	[0.022, 0.054]	[-0.101, -0.039]	[0.055, 0.100]	[0.023, 0.056]	[-0.066, -0.011]
p-value			<0.001			0.006
Profit						
Age head of hh.						
Estimate	39.703	33.924	-5.928	34.236	30.672	-3.433
CI	[37.680, 41.728]	[31.826, 36.036]	[-8.786, -3.069]	[31.962, 36.671]	[28.467, 32.859]	[-6.683, -0.221]
p-value			<0.001			0.035
Self-emp. in animal husb.						
Estimate	0.564	0.307	-0.259	0.472	0.368	-0.097
CI	[0.521, 0.604]	[0.268, 0.345]	[-0.316, -0.203]	[0.429, 0.513]	[0.327, 0.408]	[-0.155, -0.040]
p-value			<0.001			0.001
Has borrowed money						
Estimate	0.257	0.168	-0.092	0.191	0.148	-0.045
CI	[0.221, 0.294]	[0.137, 0.199]	[-0.139, -0.045]	[0.157, 0.223]	[0.118, 0.177]	[-0.089, 0.000]
p-value			<0.001			0.049

Table 10: Selected CLAN parameters with significant differences

If we compare these findings to those in the pre-existing literature, we find that our methodology is able to infer some of the results found in similar trials. It finds that self employment in animal husbandry, which is a form of agriculture, indeed lowers the total amount of loans received, and loans and profit seem to have a disproportionate effect on young households. An other previously observed result, increase in production for households working in agriculture, is seemingly contradicted in our findings. There may be two explanations for this. First, all studies discussed before involve on small family owned farms in Sub-Saharan Africa. It may simply be that the socioeconomic dynamic in Morocco are different than those in countries south of the Sahara. Alternatively, it could be that agricultural production actually does increase, but that the likelihood of being in the most affected group is diminished by the fact that those not employed in agriculture benefit significantly more. Finally we should take note that those not included in the group "Non-agricultural self-employed", includes every household where the head is not self-employed, thus the impact of microcredit on self-employed farmers is not very well assessed by this covariate anyhow.

Aside from confirming most results found in previous studies, our methodology finds a number of additional relations. Smaller households with a younger heads are most likely to take up a loan. Furthermore, those where the spouse is self-employed, and/or those that had already borrowed money were less likely to do so. Considering output, the opening of an MFI had the most positive influence on the output of households with a higher number of adults and, as mentioned before, on those where the head of the household owned a non-agricultural business prior to the intervention Oddly enough though, the introduction of availability of microcredit seems to provide disproportionally little benefit for household productivity if the spouse is self-employed. Looking at profits, households with a younger head are likely to benefit most from the intervention. Combined with the propensity of this age group to take up loans, this could result in a positive feedback loop: if a young household takes up a loan and profits, then acquaintances in their age group may be especially motivated to borrow as well. We also see that households self-employed in animal husbandry and those which had borrowed money prior to the intervention saw comparatively lower profits. We already saw that these households are less likely to take up a loan and that their output is not robustly affected by the intervention, suggesting that these groups may be vulnerable to get left behind by such a policy. This could be something to consider for local governments. About a third of Moroccans are employed in agriculture (SOURCE), of which a significant part is in animal husbandry and, although Morocco has an industrial livestock sector, rural Moroccan animal husbandry is largely still comprised of nomadic herding. Economic disenfranchisement of (semi-)nomadic peoples is a large driving force in much of the current political instability in the wider Saharan region and, although a comparatively small intervention such is this one is unlikely to cause much trouble, similar policies on a large scale could thus result in security issues.

B Additional tables and figures

B.1 IPV in Ethiopia

	Elastic Net		Random Forest	
	ATE (β_1)	HTE (β_2)	ATE (β_1)	HTE (β_2)
Physical				
Estimate	0.020	0.527	0.023	0.089
CI	[-0.024, 0.063]	[-0.185, 1.275]	[-0.021, 0.068]	[-0.301, 0.487]
p-value	0.362	0.146	0.303	0.662
Sexual				
Estimate	0.018	1.079	0.032	0.515
CI	[-0.035, 0.070]	[0.592, 1.542]	[-0.020, 0.084]	[0.157, 0.896]
p-value	0.506	<0.001	0.231	0.005
Emotional				
Estimate	-0.032	1.148	-0.035	0.622
CI	[-0.085, 0.019]	[0.658, 1.602]	[-0.087, 0.016]	[0.299, 0.965]
p-value	0.221	<0.001	0.18	<0.001

Table 11: Best Linear Predictor parameter estimates women's UBL.

	Elastic Net		Random Forest	
	ATE (β_1)	HTE (β_2)	ATE (β_1)	HTE (β_2)
Physical				
Estimate	0.000	0.453	0.004	0.126
CI	[-0.044, 0.043]	[-0.321, 1.234]	[-0.040, 0.047]	[-0.259, 0.530]
p-value	0.991	0.248	0.846	0.532
Sexual				
Estimate	-0.046	1.127	-0.038	0.419
CI	[-0.098, 0.006]	[0.638, 1.584]	[-0.089, 0.012]	[0.043, 0.791]
p-value	0.081	<0.001	0.139	0.028
Emotional				
Estimate	-0.028	1.330	-0.041	0.737
CI	[-0.080, 0.025]	[0.890, 1.781]	[-0.093, 0.011]	[0.387, 1.098]
p-value	0.305	<0.001	0.123	<0.001

Table 12: Best Linear Predictor parameter estimates couples' UBL.

	Elastic Net			Random Forest		
	γ_1	γ_5	$\gamma_5 - \gamma_1$	γ_1	γ_5	$\gamma_5 - \gamma_1$
Physical						
Estimate	-0.029	0.066	0.096	0.012	0.046	0.032
CI	[-0.133, 0.071]	[-0.032, 0.167]	[-0.049, 0.246]	[-0.091, 0.115]	[-0.051, 0.141]	[-0.109, 0.174]
p-value	0.579	0.184	0.188	0.817	0.348	0.639
Sexual						
Estimate	-0.179	0.195	0.376	-0.060	0.148	0.212
CI	[-0.292, -0.065]	[0.052, 0.336]	[0.186, 0.558]	[-0.177, 0.057]	[0.030, 0.265]	[0.042, 0.379]
p-value	**0.002**	**0.007**	**<0.001**	0.314	**0.013**	**0.013**
Emotional						
Estimate	-0.265	0.131	0.389	-0.185	0.089	0.272
CI	[-0.379, -0.149]	[-0.010, 0.275]	[0.197, 0.582]	[-0.299, -0.071]	[-0.034, 0.210]	[0.103, 0.442]
p-value	**<0.001**	0.068	**<0.001**	**0.002**	0.146	**0.002**

Table 13: GATES parameters for women's UBL.

	Elastic Net			Random Forest		
	γ_1	γ_5	$\gamma_5 - \gamma_1$	γ_1	γ_5	$\gamma_5 - \gamma_1$
Physical						
Estimate	-0.023	0.050	0.069	0.008	0.054	0.048
CI	[-0.125, 0.079]	[-0.049, 0.148]	[-0.075, 0.212]	[-0.095, 0.110]	[-0.042, 0.149]	[-0.093, 0.187]
p-value	0.658	0.304	0.340	0.882	0.260	0.484
Sexual						
Estimate	-0.213	0.180	0.400	-0.103	0.059	0.159
CI	[-0.324, -0.102]	[0.048, 0.316]	[0.212, 0.580]	[-0.219, 0.014]	[-0.058, 0.175]	[-0.005, 0.328]
p-value	**<0.001**	**0.007**	**<0.001**	0.085	0.323	0.058
Emotional						
Estimate	-0.280	0.100	0.372	-0.211	0.096	0.308
CI	[-0.393, -0.168]	[-0.048, 0.251]	[0.176, 0.570]	[-0.323, -0.100]	[-0.027, 0.220]	[0.140, 0.477]
p-value	**<0.001**	0.183	**<0.001**	**<0.001**	0.124	**<0.001**

Table 14: GATES parameters for couples' UBL.

	Sexual IPV			Emotional IPV		
	δ_1	δ_5	$\delta_5 - \delta_1$	δ_1	δ_5	$\delta_5 - \delta_1$
Polygamous						
Estimate	0.135	0.047	-0.088	0.099	0.076	-0.024
CI	[0.094, 0.177]	[0.020, 0.070]	[-0.137, -0.041]	[0.062, 0.133]	[0.043, 0.106]	[-0.073, 0.024]
p-value			0.000			0.346
Cat. Marriage Length						
Estimate	3.083	3.113	0.017	2.653	3.556	0.899
CI	[2.924, 3.242]	[2.953, 3.273]	[-0.189, 0.263]	[2.524, 2.793]	[3.396, 3.716]	[0.690, 1.108]
p-value			0.714			0.000
Living children						
Estimate	4.464	4.090	-0.383	3.735	4.558	0.821
CI	[4.186, 4.762]	[3.819, 4.361]	[-0.766, 0.045]	[3.448, 4.020]	[4.288, 4.842]	[0.434, 1.214]
p-value			0.080			0.000
Asset Index						
Estimate	3.876	3.797	-0.109	3.552	4.198	0.683
CI	[3.593, 4.160]	[3.507, 4.087]	[-0.518, 0.287]	[3.266, 3.830]	[3.908, 4.506]	[0.273, 1.101]
p-value			0.602			0.001
Wealth Index						
Estimate	0.095	-0.077	-0.189	0.081	0.051	-0.054
CI	[-0.052, 0.236]	[-0.204, 0.040]	[-0.375, 0.002]	[-0.063, 0.244]	[-0.082, 0.180]	[-0.247, 0.143]
p-value			0.052			0.591
Age Woman						
Estimate	1.932	1.731	-0.203	1.640	1.948	0.306
CI	[1.844, 2.021]	[1.648, 1.819]	[-0.323, -0.083]	[1.555, 1.725]	[1.864, 2.032]	[0.184, 0.428]
p-value			0.001			0.000
Schooling Woman						
Estimate	1.149	1.329	0.188	1.306	1.223	-0.090
CI	[1.107, 1.191]	[1.270, 1.391]	[0.110, 0.266]	[1.249, 1.363]	[1.169, 1.277]	[-0.165, -0.014]
p-value			0.000			0.021
Age Man						
Estimate	2.468	2.333	-0.132	2.239	2.487	0.261
CI	[2.393, 2.548]	[2.253, 2.415]	[-0.245, -0.020]	[2.154, 2.323]	[2.415, 2.562]	[0.144, 0.374]
p-value			0.020			0.000
Schooling Man						
Estimate	1.607	1.718	0.107	1.720	1.677	-0.034
CI	[1.535, 1.680]	[1.644, 1.791]	[0.010, 0.208]	[1.645, 1.794]	[1.609, 1.749]	[-0.135, 0.068]
p-value			0.029			0.492
Muslim						
Estimate	0.352	0.436	0.090	0.597	0.156	-0.440
CI	[0.292, 0.407]	[0.376, 0.496]	[0.008, 0.173]	[0.538, 0.656]	[0.113, 0.200]	[-0.514, -0.368]
p-value			0.027			0.000
Orthodox						
Estimate	0.568	0.487	-0.083	0.368	0.668	0.295
CI	[0.508, 0.627]	[0.425, 0.545]	[-0.167, 0.001]	[0.309, 0.425]	[0.611, 0.724]	[0.214, 0.376]
p-value			0.053			0.000
Other Religion						
Estimate	0.062	0.070	0.011	0.024	0.136	0.110
CI	[0.032, 0.089]	[0.037, 0.098]	[-0.029, 0.054]	[0.005, 0.040]	[0.093, 0.175]	[0.067, 0.157]
p-value			0.491			0.000

Table 15: CLAN parameters women's UBL

	Sexual IPV			Emotional IPV		
	δ_1	δ_5	$\delta_5 - \delta_1$	δ_1	δ_5	$\delta_5 - \delta_1$
Polygamous						
Estimate	0.076	0.108	0.033	0.054	0.103	0.044
CI	[0.043, 0.106]	[0.071, 0.145]	[-0.013, 0.083]	[0.025, 0.078]	[0.067, 0.140]	[-0.003, 0.091]
p-value			0.137			0.055
Cat. Marriage Length						
Estimate	2.639	3.507	0.920	2.669	3.421	0.706
CI	[2.503, 2.771]	[3.344, 3.672]	[0.713, 1.131]	[2.535, 2.815]	[3.261, 3.581]	[0.490, 0.925]
p-value			0.000			0.000
Living children						
Estimate	3.967	4.442	0.483	3.411	4.812	1.404
CI	[3.702, 4.236]	[4.183, 4.736]	[0.094, 0.875]	[3.161, 3.679]	[4.532, 5.084]	[1.009, 1.799]
p-value			0.013			0.000
Asset Index						
Estimate	3.364	4.019	0.673	3.519	3.956	0.415
CI	[3.102, 3.620]	[3.738, 4.300]	[0.294, 1.054]	[3.227, 3.806]	[3.669, 4.238]	[0.014, 0.808]
p-value			0.000			0.031
Wealth Index						
Estimate	-0.064	0.078	0.123	0.009	0.085	0.058
CI	[-0.182, 0.054]	[-0.081, 0.226]	[-0.074, 0.322]	[-0.133, 0.129]	[-0.063, 0.234]	[-0.132, 0.258]
p-value			0.182			0.419
Age Woman						
Estimate	1.656	1.941	0.286	1.556	2.006	0.456
CI	[1.573, 1.742]	[1.856, 2.026]	[0.164, 0.409]	[1.474, 1.638]	[1.920, 2.091]	[0.334, 0.576]
p-value			0.000			0.000
Schooling Woman						
Estimate	1.152	1.351	0.205	1.344	1.242	-0.101
CI	[1.105, 1.201]	[1.289, 1.417]	[0.123, 0.286]	[1.286, 1.410]	[1.187, 1.300]	[-0.184, -0.015]
p-value			0.000			0.023
Age Mea						
Estimate	2.264	2.483	0.212	2.137	2.537	0.388
CI	[2.179, 2.350]	[2.408, 2.559]	[0.100, 0.324]	[2.052, 2.226]	[2.465, 2.609]	[0.274, 0.505]
p-value			0.000			0.000
Schooling Man						
Estimate	1.524	1.744	0.212	1.680	1.686	0.020
CI	[1.459, 1.593]	[1.670, 1.821]	[0.112, 0.312]	[1.605, 1.751]	[1.611, 1.760]	[-0.090, 0.124]
p-value			0.000			0.723
Muslim						
Estimate	0.424	0.383	-0.041	0.644	0.283	-0.375
CI	[0.365, 0.483]	[0.321, 0.437]	[-0.125, 0.043]	[0.587, 0.702]	[0.230, 0.337]	[-0.454, -0.297]
p-value			0.340			0.000
Orthodox						
Estimate	0.494	0.522	0.045	0.289	0.609	0.335
CI	[0.434, 0.554]	[0.461, 0.580]	[-0.040, 0.129]	[0.235, 0.343]	[0.551, 0.667]	[0.257, 0.413]
p-value			0.263			0.000
Other Religion						
Estimate	0.075	0.076	0.004	0.041	0.098	0.061
CI	[0.043, 0.107]	[0.043, 0.106]	[-0.040, 0.047]	[0.017, 0.064]	[0.061, 0.131]	[0.016, 0.106]
p-value			0.853			0.007

Table 16: CLAN parameters men's UBL

	Sexual IPV			Emotional IPV		
	δ_1	δ_5	$\delta_5 - \delta_1$	δ_1	δ_5	$\delta_5 - \delta_1$
Polygamous						
Estimate	0.161	0.069	-0.096	0.118	0.079	-0.046
CI	[0.116, 0.203]	[0.037, 0.096]	[-0.149, -0.044]	[0.079, 0.156]	[0.045, 0.109]	[-0.095, 0.005]
p-value			0.000			0.080
Cat. Marriage Length						
Estimate	3.415	2.987	-0.426	3.035	3.272	0.257
CI	[3.258, 3.571]	[2.834, 3.143]	[-0.647, -0.204]	[2.882, 3.188]	[3.113, 3.430]	[0.039, 0.482]
p-value			0.000			0.021
Living children						
Estimate	4.504	4.211	-0.189	4.283	4.316	0.026
CI	[4.222, 4.782]	[3.944, 4.484]	[-0.574, 0.211]	[3.999, 4.583]	[4.050, 4.582]	[-0.380, 0.415]
p-value			0.344			0.875
Asset Index						
Estimate	3.450	3.996	0.533	3.415	4.250	0.841
CI	[3.178, 3.726]	[3.697, 4.295]	[0.129, 0.933]	[3.155, 3.686]	[3.954, 4.539]	[0.434, 1.234]
p-value			0.007			0.000
Wealth Index						
Estimate	-0.191	0.261	0.420	-0.098	0.287	0.397
CI	[-0.294, -0.089]	[0.091, 0.437]	[0.213, 0.623]	[-0.210, 0.024]	[0.119, 0.462]	[0.185, 0.605]
p-value			0.000			0.000
Age Woman						
Estimate	1.848	1.828	-0.030	1.743	1.892	0.151
CI	[1.759, 1.937]	[1.741, 1.912]	[-0.154, 0.094]	[1.656, 1.829]	[1.803, 1.979]	[0.028, 0.274]
p-value			0.638			0.015
Schooling Woman						
Estimate	1.198	1.319	0.122	1.276	1.285	0.022
CI	[1.149, 1.246]	[1.259, 1.381]	[0.042, 0.205]	[1.221, 1.332]	[1.227, 1.346]	[-0.064, 0.106]
p-value			0.002			0.452
Age Man						
Estimate	2.433	2.363	-0.085	2.338	2.425	0.081
CI	[2.353, 2.514]	[2.282, 2.446]	[-0.194, 0.024]	[2.259, 2.421]	[2.346, 2.505]	[-0.031, 0.192]
p-value			0.125			0.140
Schooling Man						
Estimate	1.561	1.756	0.196	1.612	1.730	0.110
CI	[1.492, 1.631]	[1.679, 1.833]	[0.092, 0.295]	[1.538, 1.688]	[1.654, 1.806]	[0.007, 0.213]
p-value			0.000			0.035
Muslim						
Estimate	0.582	0.402	-0.193	0.654	0.268	-0.393
CI	[0.523, 0.640]	[0.342, 0.459]	[-0.275, -0.110]	[0.598, 0.711]	[0.216, 0.321]	[-0.469, -0.316]
p-value			0.000			0.000
Orthodox						
Estimate	0.357	0.519	0.174	0.301	0.632	0.316
CI	[0.298, 0.413]	[0.459, 0.578]	[0.091, 0.256]	[0.247, 0.356]	[0.575, 0.690]	[0.316, 0.396]
p-value			0.000			0.000
Other Religion						
Estimate	0.039	0.061	0.015	0.017	0.086	0.070
CI	[0.015, 0.060]	[0.031, 0.087]	[-0.020, 0.053]	[0.000, 0.029]	[0.051, 0.118]	[0.035, 0.108]
p-value			0.383			0.000

Table 17: CLAN parameters couple's UBL

B.2 MFI in Morocco

	Elastic Net		Random Forest	
	Household	Village	Household	Village
Amount of loans				
Best BLP (Λ)	32827	963209	21680	930974
Best GATES ($\bar{\Lambda}$)	1691802	3327410	1719767	3223021
Output				
Best BLP (Λ)	10957333	89784168	8824837	91818058
Best GATES ($\bar{\Lambda}$)	64757516	175362050	68620780	173850321
Profit				
Best BLP (Λ)	1009426	21090494	8056840	20721090
Best GATES ($\bar{\Lambda}$)	14309376	42844786	21369126	41191085
Consumption				
Best BLP (Λ)	2456	20477	6185	19468
Best GATES ($\bar{\Lambda}$)	23310	64322	21174	61755

Table 18: Comparison of ML proxy performance between household and village characteristics.

	Elastic Net			Random Forest		
	δ_1	δ_5	$\delta_5 - \delta_1$	δ_1	δ_5	$\delta_5 - \delta_1$
Number of hh. members						
Estimate	4.486	3.311	-1.223	4.261	2.692	-1.540
CI	[4.231, 4.760]	[3.041, 3.582]	[-1.611, -0.839]	[4.001, 4.533]	[2.393, 2.991]	[-1.945, -1.136]
p-value			<0.0005			<0.0005
Number of memb. over 16						
Estimate	2.654	2.370	-0.266	2.779	1.915	-0.822
CI	[2.494, 2.831]	[2.148, 2.606]	[-0.547, 0.013]	[2.577, 2.975]	[1.689, 2.129]	[-1.117, -0.523]
p-value			0.062			<0.0005
Age head of hh.						
Estimate	39.344	30.464	-8.381	36.155	24.621	-11.174
CI	[37.304, 41.346]	[28.250, 32.620]	[-11.388, -5.374]	[34.127, 38.251]	[22.294, 26.928]	[-14.354, -8.024]
p-value			<0.0005			<0.0005
Self-emp. in animal husb.						
Estimate	0.467	0.340	-0.132	0.409	0.309	-0.100
CI	[0.426, 0.509]	[0.299, 0.378]	[-0.190, -0.074]	[0.368, 0.450]	[0.269, 0.347]	[-0.156, -0.043]
p-value			<0.0005			0.001
Non-agricultural self-emp.						
Estimate	0.062	0.212	0.155	0.119	0.138	0.021
CI	[0.042, 0.082]	[0.178, 0.246]	[0.113, 0.193]	[0.091, 0.145]	[0.109, 0.166]	[-0.019, 0.062]
p-value			<0.0005			0.238
Has borrowed money						
Estimate	0.226	0.157	-0.068	0.273	0.136	-0.136
CI	[0.192, 0.261]	[0.127, 0.188]	[-0.114, -0.023]	[0.235, 0.309]	[0.107, 0.164]	[-0.185, -0.089]
p-value			0.003			<0.0005
Spouse self-emp.						
Estimate	0.128	0.033	-0.096	0.091	0.027	-0.065
CI	[0.099, 0.155]	[0.018, 0.047]	[-0.129, -0.064]	[0.067, 0.115]	[0.014, 0.041]	[-0.092, -0.039]
p-value			<0.0005			<0.0005
Other hh. memb. self-emp.						
Estimate	0.050	0.040	-0.016	0.058	0.069	0.010
CI	[0.031, 0.067]	[0.024, 0.056]	[-0.042, 0.010]	[0.038, 0.077]	[0.048, 0.090]	[-0.018, 0.039]
p-value			0.221			0.465

Table 19: CLAN estimates income

	Elastic Net			Random Forest		
	δ_1	δ_5	$\delta_5 - \delta_1$	δ_1	δ_5	$\delta_5 - \delta_1$
Number of hh. members						
Estimate	3.741	4.060	0.359	3.448	3.629	0.140
CI	[3.487, 3.979]	[3.761, 4.343]	[-0.038, 0.756]	[3.166, 3.729]	[3.298, 3.960]	[-0.279, 0.558]
p-value			0.055			0.432
Number of memb. over 16						
Estimate	2.362	2.882	0.517	2.230	2.605	0.322
CI	[2.193, 2.518]	[2.665, 3.097]	[0.242, 0.797]	[2.048, 2.413]	[2.363, 2.835]	[0.015, 0.616]
p-value			<0.0005			0.030
Age head of hh.						
Estimate	36.634	35.793	-1.697	31.664	33.060	0.897
CI	[34.646, 38.756]	[33.524, 37.812]	[-4.717, 1.322]	[29.471, 33.820]	[30.715, 35.405]	[-2.235, 4.178]
p-value			0.271			0.478
Self-emp. in animal husb.						
Estimate	0.451	0.402	-0.056	0.400	0.410	0.014
CI	[0.410, 0.493]	[0.361, 0.443]	[-0.113, 0.002]	[0.359, 0.441]	[0.369, 0.450]	[-0.045, 0.072]
p-value			0.059			0.619
Non-agricultural self-emp.						
Estimate	0.055	0.274	0.218	0.124	0.241	0.111
CI	[0.036, 0.074]	[0.237, 0.311]	[0.176, 0.263]	[0.096, 0.151]	[0.205, 0.276]	[0.066, 0.155]
p-value			<0.0005			<0.0005
Has borrowed money						
Estimate	0.240	0.180	-0.064	0.208	0.184	-0.027
CI	[0.202, 0.273]	[0.147, 0.211]	[-0.111, -0.016]	[0.174, 0.242]	[0.151, 0.215]	[-0.075, 0.019]
p-value			0.008			0.244
Spouse self-emp.						
Estimate	0.111	0.038	-0.071	0.078	0.040	-0.038
CI	[0.083, 0.135]	[0.022, 0.054]	[-0.101, -0.039]	[0.055, 0.100]	[0.023, 0.056]	[-0.066, -0.011]
p-value			<0.0005			0.006
Other hh. memb. self-emp.						
Estimate	0.058	0.038	-0.022	0.055	0.063	0.000
CI	[0.038, 0.077]	[0.022, 0.053]	[-0.049, 0.004]	[0.035, 0.073]	[0.043, 0.084]	[-0.025, 0.028]
p-value			0.092			0.890

Table 20: CLAN estimates output

	Elastic Net			Random Forest		
	δ_1	δ_5	$\delta_5 - \delta_1$	δ_1	δ_5	$\delta_5 - \delta_1$
Number of hh. members						
Estimate	4.027	3.728	-0.248	3.737	3.236	-0.542
CI	[3.772, 4.288]	[3.460, 4.001]	[-0.623, 0.126]	[3.442, 4.036]	[2.957, 3.521]	[-0.962, -0.119]
p-value			0.193			0.012
Number of memb. over 16						
Estimate	2.728	2.461	-0.214	2.550	2.249	-0.292
CI	[2.552, 2.912]	[2.265, 2.657]	[-0.483, 0.050]	[2.336, 2.765]	[2.048, 2.448]	[-0.594, 0.011]
p-value			0.112			0.059
Age head of hh.						
Estimate	39.703	33.924	-5.928	34.236	30.672	-3.433
CI	[37.680, 41.728]	[31.826, 36.036]	[-8.786, -3.069]	[31.962, 36.671]	[28.467, 32.859]	[-6.683, -0.221]
p-value			<0.0005			0.035
Self-emp. in animal husb.						
Estimate	0.564	0.307	-0.259	0.472	0.368	-0.097
CI	[0.521, 0.604]	[0.268, 0.345]	[-0.316, -0.203]	[0.429, 0.513]	[0.327, 0.408]	[-0.155, -0.040]
p-value			<0.0005			0.001
Non-agricultural self-emp.						
Estimate	0.107	0.186	0.080	0.122	0.168	0.038
CI	[0.081, 0.133]	[0.152, 0.217]	[0.037, 0.122]	[0.094, 0.149]	[0.136, 0.200]	[-0.003, 0.077]
p-value			<0.0005			0.058
Has borrowed money						
Estimate	0.257	0.168	-0.092	0.191	0.148	-0.045
CI	[0.221, 0.294]	[0.137, 0.199]	[-0.139, -0.045]	[0.157, 0.223]	[0.118, 0.177]	[-0.089, 0.000]
p-value			<0.0005			0.049
Spouse self-emp.						
Estimate	0.077	0.055	-0.024	0.066	0.054	-0.014
CI	[0.054, 0.098]	[0.035, 0.073]	[-0.052, 0.006]	[0.045, 0.086]	[0.035, 0.072]	[-0.042, 0.015]
p-value			0.116			0.342
Other hh. memb. self-emp.						
Estimate	0.078	0.032	-0.049	0.068	0.056	-0.018
CI	[0.056, 0.100]	[0.017, 0.047]	[-0.077, -0.020]	[0.047, 0.089]	[0.037, 0.075]	[-0.043, 0.011]
p-value			<0.0005			0.229

Table 21: CLAN estimates production