Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis BSc2 Econometrics and Economics

A Deep Dive into Microcredit: A Machine Learning Analysis of Heterogeneous Treatment Effects

Diego Saverys (530650)

Supervisor:	Max Welz
Second assessor:	Dr. Andrea A. Naghi
Date final version:	29th June 2023

Abstract

This bachelor thesis aimed to investigate whether microcredits are effective as a poverty alleviation tool and whether they exhibit treatment effect heterogeneity. Using the innovative methodology developed by Chernozhukov et al. (2022), the study examined evidence of heterogeneity within and across Randomized Controlled Trials centered on microfinance, exploring four key outcome variables: loan amounts, business revenues, business profits, and consumption. The study analysed three settings - two individual RCTs in Morocco and India, and a pooled analysis across six RCTs conducted in different countries. Turning to results, mix evidence for microfinance effectiveness was found: while microcredit increased loan amounts and business revenues, it did not significantly improve profits and consumption. However, the study found significant heterogeneity in treatment effects across all variables, confirming the theories suggesting that heterogeneous treatment effects may be obscured by the lack of transformative impact of microcredit interventions. Evidence suggested that both site-level and household-level covariates influenced this heterogeneity. Men and households with previous business experience tend to benefit more from enhanced microcredit access. Also, lower interest rates as well as larger loan sizes as a percentage of income resulted in higher impacts of microcredit. Policy implications of these findings include the potential for a more targeted microfinance approach, varying by urban and rural contexts and different households' characteristics.

1 Introduction

Despite the recent progress made in reducing global poverty over the past few decades, it remains a daunting challenge that requires urgent attention. According to the United Nations (UN), as of 2015, about 736 million people still lived on less than US \$1.90 a day, and many lack food, clean drinking water and sanitation. About 10 percent of the world's population lives in extreme poverty[1] and women are more likely to be poor than men because they tend to have less paid work, education, and own less property (United Nations, 2015b). Despite accelerated economic growth in countries such as China and India aiding millions to escape poverty, progress is still inconsistent (United Nations, 2015b). The UN has set a goal to eradicate poverty in all its forms by 2030 as part of its Sustainable Development Goals (SDGs). This task not only consists of lifting people out of extreme poverty, but also providing access to basic needs such as food, clean water and sanitation (United Nations, 2015b). With new threats brought on by climate change, conflict and food insecurity, it is essential to keep developing strategies to combat this issue and bring people out of poverty. Over the years, Randomized controlled trials (RCTs) have been presented as a critical tool to address global poverty. Notably, one primary approach has emerged from these trials. Multiple RCTs such as Augsburg et al. (2015), Tarozzi et al. (2015), Banerjee, Duflo, Glennerster and Kinnan (2015), Angelucci et al. (2015), Attanasio et al. (2015), and Crépon et al. (2015), have proposed microcredit as a tool for poverty alleviation. Microcredit was envisaged to aid poor households by encouraging entrepreneurship and fostering economic self-reliance. Specifically, it has been proposed as a practical and effective alternative to traditional finance, intended to empower the poor with funds to maintain or create economic activities (Yunus, 1999; Khandker, 1998).

However, concerns were raised about the potential downside of creating credit bubbles (Meager, 2019). Despite their varied designs and implementations, these studies consistently demonstrated modest, albeit not transformative, positive impacts on metrics like business earnings and consumption (Banerjee, Karlan & Zinman, 2015). While the approach presents a viable strategy towards achieving the SDGs, the enthusiasm for microcredit as a revolutionary poverty reduction tool has not been met with the anticipated transformative impact (World Bank Group, 2016). Rather, it has spurred a robust backlash (Banerjee, Duflo, Glennerster & Kinnan, 2015). Why has microcredit fallen short of making a significant impact? One plausible answer lies in the heterogeneous treatment effects that may be obscured by the moderate average outcomes observed. Indeed, as indicated by Chernozhukov et al. (2022) and Angelucci et al. (2015), underlying variations in the effects of microcredit could be critical to understanding its real impact and potential for poverty alleviation. Understanding heterogeneity in poverty-targeting programs is not only relevant for evaluating the universal applicability of a program's impact, but also for shedding light on the mechanisms underlying the program.

Therefore, this research will investigate the possibility of heterogeneity in treatment effects (HET) within and across the scope of multiple RCTs targeting poverty using microfinance.

[1] Poverty is defined by the UN as the "lack of income and productive resources to ensure livelihoods" and it manifests itself with hunger and malnutrition, limited access to education and other basic services, social discrimination and exclusion, as well as the lack of participation in decision-making (United Nations, 2015a).

Specifically, it will investigate the HET across four key outcome metrics as described by Meager (2019): the amount of loans, business profits, business revenues, and consumer spending. The main research question that will guide this paper can be articulated as follows:

- *Is there evidence for the presence of treatment effect heterogeneity within and across the Randomized Control Trials targeting poverty using microfinance ? Which covariates are driving this potential heterogeneity?*

To investigate this research question, the research presented in this paper will focus on detecting and understanding heterogeneity in three different settings, each distinctly aimed at reducing poverty by enhancing access to microcredit to poor households. The first setting is the RCT performed in Morocco by Crépon et al. (2015). Note that this study has previously been explored by Chernozhukov et al. (2022) and is therefore a replication of their results as a critical first step to ensure validity and robustness of the investigation of heterogeneity in this context. Subsequently, the analysis will be extended, with the analysis being applied to different contexts. The first extension investigated is the RCT conducted in India by Banerjee, Duflo, Glennerster and Kinnan (2015). The design of this RCT was very similar to the one in Morocco with the exception of the characterisitcs collected and it will enable to broaden the understanding of the heterogeneity impacts of microcredit initiative to a different context. Then, the second extension presented in this analysis is a pooled analysis investigating heterogeneity across the six RCTs conducted in Bosnia and Herzegovina, Ethiopia, India, Mexico, Mongolia, and Morocco simultaneously. This pooled analysis will provide robust and generalize insights about the heterogeneous impacts of microcredit access on poverty reduction. In particular, conducting a pooled analysis and two metastudies allows to conduct a comparative investigation of the similarilities and differences of the considered RCTs. This approach will therefore shed light on the following sub-questions:

1. As highlighted by Meager (2019), HET across studies could be attributed to site-level variables as well as household-level variables. Hence, the first sub-question will probe: *"Is the potential heterogeneity in treatment effects across the settings driven by site-level covariates or rather by household-level covariates?"*.
2. Considering the variety of the contexts investigated in this study, a second sub-question is : *"Is there a particular context where enhancing microcredit access has proven to be particularly beneficial for the population?"*.

The research questions presented in this paper will be answered by applying the methodology developed by Chernozhukov et al. (2022). The authors presented a generic approach to using any available machine learning (ML) regression learner to "predict and make inference on heterogeneous treatment effects" (Chernozhukov et al., 2022). Instead of attempting to obtain uniformly valid inference on the conditional average treatment effect (CATE), the study focuses on generating reliable estimations and interpretations of CATE features. These features include the Best Linear Predictor (BLP), the Sorted Group Average Treatment Effects (GATES), and Classification Analysis (CLAN). Together, they provide information on detectable heterogeneity (via

BLP), treatment effects for different groups (via GATES), and the particular covariates associated with this heterogeneity (via CLAN). The approach developed by the authors is particularly useful because of its lack of strong assumptions. Indeed, unconfoundededness and propensity score being bounded away from zero and one are the only two assumptions required to perform the analysis and answer the research questions (Chernozhukov et al., 2022). Their approach offers numerous benefits when exploring the diversity of the impacts of microfinanace treatments. Indeed, in this kinds of settings, our understanding of the origin of heterogeneity remains largely uncertain (Chernozhukov et al., 2022). It cannot be denied that the factor "previously owned a business" has consistently demonstrated its capacity to predict the differences in treatment effect, as described by Meager (2019). Nonetheless, our knowledge regarding other potential predictors of this heterogeneity is limited. Given the high-dimensional nature of these settings, Chernozhukov et al. (2022)'s approach allows us to incorporate a large number of characteristics in any form. Utilizing their CLAN estimation, we can identify the characteristics of the most and least affected sub-populations. This ability to identify specific characteristics can provide invaluable insights for welfare analysis, or for devising strategies to target households that are most likely to benefit from access to microfinance.

In terms of results, it is found that microcredits do not have a significant impact on most of the outcome variables in the three settings. However, heterogeneity in treatment effects has been found within and across the three settings investigated. In Morocco, there was HET on the amount of loans, business outputs and business profits and this heterogeneity was largely driven by unobservable village-level covariates. The RCT in India was characterized by strong HET on business revenues and profits as well as consumption. The most impacted groups typically had more prior business experiences, lower literacy rates, resided in less populated areas and lived in areas with fewer businesses. Finally, the pooled analysis showed significant heterogeneity in treatment effects across all outcome variables. Households with previous business experience were consistently more impacted by microcredit. In terms of study design, implementations of RCT with larger loans as a percentage of income and lower interest rates proved to achieve larger impact on business outcomes while larger impacts on consumption were found in rural areas. It was also found that different strategies are required to obtain a significant impact on business outputs compared to obtaining significant impacts on standards of livings as represented by consumption.

The main contribution of this paper is that it is the first to exhaustively explore the origins and impacts of heterogeneity in treatment effects within and across multiple RCTs concerning microcredit enhancement using the innovative approach of Chernozhukov et al. (2022). The RCTs investigated in this paper offer a rich and varied background for exploration of heterogeneity in treatment effect of microcredit. The unique context of each RCT strengthens the validity of the findings and provides a comprehensive understanding of microcredit impacts in diverse settings. The results obtained illuminate on the different impacts of these microcredit programs across various subgroups of the poor populations. Ultimately, the insights gained in this research can serve as guides for program implementers, enabling them to fine-tune their interventions to better meet the specific needs and circumstances of diverse targeted subgroups.

For instance, this could by done by targeting households with previous business experience with higher loan sizes as a percentage of income and lower interest rates.

The reminder of this paper is organized as follows. The upcoming Section 2 provides a description of the existing literature on the topic. Then, Section 3 provides a description of the data that is used in the studies, and Section 4 describes which methodology will be applied to identify the potential HET. Section 5 provides the results of the research, and finally concluding remarks are presented in Section 6.

2 Literature Review

Microfinance is typically characterised by the provision of small loans, savings account, insurance, and other financial products to low-income individuals or those who have been traditionally underserved by conventional financial institutions (Latifee, 2003). Microcredit, a subset of microfinance, specifically refers to the provision of small loans to help spur entrepreneurship and self-employment, often targeting the very poor who lack collateral or a credit history (Latifee, 2003). According to Sen (1976), poverty is more than a lack of income; it is a deprivation of capabilities to live a life one values. It includes factors like inadequate participation in societal activities, lack of education, limited access to resources, and poor health (Sen, 1976). The motivation behind enhancing access to microcredit as a tool to alleviate poverty comes from the role of developed financial services in stimulating economic growth. As proposed by King and Levine (1993) and De Gregorio and Guidotti (1995), such financial development could lead to a uniform increase in income across all segments of society, potentially alleviating poverty, as suggested by Soubbotina and Sheram (2000), Besley and Burgess (2003) and Bourguignon (2004). Other authors argue that orientating financial services towards the poor can contribute substantially to reducing income inequalities (Datt & Ravallion, 1992; Ravallion, 2001). More directly, microfinance initiatives, specifically microcredit, have been proposed as a practical and effective alternative to traditional finance, intended to empower the poor with funds to maintain or create economic activities (Yunus, 1999; Khandker, 1998).

The appeal of microcredit goes beyond income enhancement, as it is believed to lead to broader socio-economic benefits. Access to microcredit can result in improved access to essential services such as health and education (Morduch & Haley, 2002). Its potential to address market imperfections in the financial sector, foster more inclusive economic growth and reduce inequality is also recognized (Beck et al., 2007). Moreover, microcredit programs focused on poverty alleviation, such as those cited by Latifee (2003), can provide financial and business services to the very poor, creating opportunities for self-employment. Finally, Yunus' work in Bangladesh also revealed a significant portion of poor families benefitting from microcredit, demonstrating its ability for poverty reduction (Yunus, 2007).

However, as enthusiasm for microcredit grew, so did scrutiny and criticism. The initial optimism around microcredit was based largely on anecdotal evidence and successful replication of models like the Grameen Bank (Meager, 2019). The Grameen Bank was founded by Muhammad

Yunus in Bangladesh and was one of the earliest and most influential models of microcredit. Its innovative approach consisted of providing small loans to groups of poor women who could use these funds to start or expand microenterprises, thus aiming to empower them economically (Yunus, 1999). Nevertheless, systematic evidence of microcredit's impacts on poverty alleviation remained inconclusive (Banerjee, Karlan & Zinman, 2015; World Bank Group, 2016). Critics began to highlight concerns about over-indebtedness and a trend towards commercialisation that might compromise microcredit's initial goal of serving the poor. Shetty (2013) suggested that the industry's shift towards profitability was causing a dilution of the initial goals of poverty alleviation and empowerment, particularly for women.

Six studies grew out of this debating context. The studies were conducted to produce causal proof concerning the effects of microcredit on its target users (Banerjee, Karlan & Zinman, 2015). In particular, the six studies conducted in Bosnia-Herzegovina (Augsburg et al., 2015), Ethiopia (Tarozzi et al., 2015), India (Banerjee, Duflo, Glennerster & Kinnan, 2015), Mexico (Angelucci et al., 2015), Mongolia (Attanasio et al., 2015) and Morocco (Crépon et al., 2015) were RCTs. According to Hariton and Locascio (2018), "RCTs are prospective studies that measure the effectiveness of a new information or treatment". They are widely used in economics to estimate the causal effect of a treatment on an outcome. They involve randomly assigning units to either a treatment or a control group to measure the causal effect of the treatment on an outcome. This random assignment ensures that any differences in outcomes between the two groups can be attributed to the treatment alone. RCTs are very popular in the literature because the average difference between the treatment and the control group is an unbiased estimate of the typical causal effects for the trials. Also, precise probabilistic statements can be made indicating how unusual the observed difference would be under specific hypothesized causal effects (Rubin, 1972). Moreover, RCTs minimize selection biases, often encountered in observational studies, by randomization. In the context of microcredit, such biases could originate from both demand-side and supply-side factors (Banerjee, Karlan & Zinman, 2015).

These implementations of RCTs on microcredit to better understand its impact found mixed results and revealed key nuances to consider when evaluating the effectiveness of microcredit initiatives. A recurring observation across these studies was the modest take-up rates of credit among potential microentrepreneurs, with estimates ranging from around 17 to 31% (Augsburg et al., 2015; Tarozzi et al., 2015; Banerjee, Duflo, Glennerster & Kinnan, 2015; Angelucci et al., 2015; Attanasio et al., 2015; Crépon et al., 2015). It was noted that predicting microcredit take-up was challenging due to unobserved heterogeneity in borrowing and lending decisions. This could be particularly relevant in Bosnia, India, and Mexico, where borrowing from other Microfinance Institutions (MFIs) increased, maybe as compensation for low take-up rates (Banerjee, Karlan & Zinman, 2015). It is important to highlight that the studies did not provide substantial evidence of transformative effects on the average borrower, despite the investment in business growth (Banerjee, Karlan & Zinman, 2015). The lack of transformative effects does not necessarily indicate a failure of microcredit but it rather points towards more subtle effects. Furthermore, the studies provided little support for the harshest criticisms of microcredit mentioned above (Banerjee, Karlan & Zinman, 2015).

With regards to specific outcomes, the studies measured profits, income, consumption standards as well as living standards and female empowerment. They found potential growth in business sizes and profits. For instance in Morocco, where access to microfinance had a positive effect on assets, resulting in an estimated impact of 1,448 Moroccan Dirhams (MAD), the Moroccan currency (Crépon et al., 2015). However, none of the six studies found a statistically significant increase in total household income (Banerjee, Karlan & Zinman, 2015). In terms of the other crucial outcomes, they revealed varying effects of microcredit, highlighting the complex and multifaceted impact on poverty alleviation (Augsburg et al., 2015; Tarozzi et al., 2015; Banerjee, Duflo, Glennerster & Kinnan, 2015; Angelucci et al., 2015; Attanasio et al., 2015; Crépon et al., 2015). These initial findings suggest that while microcredit may not have a revolutionary effect in elevating individuals or communities out of poverty, it appears to offer greater autonomy in decision-making and opportunities for self-sustainability. These studies collectively underline the average treatment effects of microcredit, which, while indicating positive trends, are not consistently high. As a consequence, Banerjee, Karlan and Zinman (2015) argue that this highlights the need to delve deeper into the analysis and understand the variations, heterogeneity, and effectiveness of individual components within microcredit programs. The authors accentuate the importance of recognizing that microcredit, while generally beneficial, can have differing impacts across the spectrum.

This is not surprising. Indeed, the notion that the effects of microcredit may be heterogeneous is a recurrent theme in RCTs that assess microfinance initiatives. Its study is paramount for both policy design and welfare evaluations. A lack of evidence on average effects might conceal heterogeneity, suggesting the presence of potential winners and losers from microcredit expansions (Chernozhukov et al., 2022). The literature consistently hints at the possibility of such heterogeneity in the impacts of microcredit (Banerjee, Karlan & Zinman, 2015; Meager, 2019; Chernozhukov et al., 2022; Banerjee et al., 2017). For instance, experiences from Morocco and India showed significant positive effects on business profits for certain segments, while negative impacts were seen elsewhere (Crépon et al., 2015; Banerjee, Duflo, Glennerster & Kinnan, 2015). In Mexico, there was evidence of enhanced financial decision-making power among certain demographics (Angelucci et al., 2015). These findings imply heterogeneity in impacts across the distribution and suggest that research is needed to understand the variations and nuances of microcredit effectiveness.

Heterogeneity is not confined solely to individual households but is also evident in site-by-site variations. For instance, the relationship between microfinance and poverty varies across countries and is influenced by the targeting strategy of the MFIs (Bangoura et al., 2016). Furthermore, a central piece in the heterogeneity story is the potential presence of behavioral tendencies that might lead some borrowers to harm themselves (Banerjee, Karlan & Zinman, 2015). It appears that microcredit can play a strengthening role even in the presence of such behavioral deviations. Several studies have considered subpopulations to explore this heterogeneity further. Banerjee et al. (2017) found that the impacts of microcredit were persistently different six years after the microcredit was introduced, with greater impacts on business outcomes for those who already has businesses compared to those who did not. These findings imply the significance of entrepreneurial ability, reinforcing the importance of focusing on borrowers at the intensive margin.

Is it therefore clear that the existing literature confirms the need of investigating the possible heterogeneity of program participants. A deeper understanding of the differential effects of microcredit expansions can aid in customising interventions that better serve the varying needs of different individuals. This understanding can also guide policymakers to identify households likely to reap maximum benefits from microcredit programs and those that might require additional support or services. Recognizing this heterogeneity also helps in identifying potential mechanisms underpinning the success or failure of a program, which is paramount in a world with increasing concerns about debt traps. Furthermore, insights into distributional effects of these programs are critical in designing screening and targeting technologies that aim to maximize benefits while minimizing harm. Also, concerns about the external validity of these studies remain, emphasizing the importance of understanding heterogeneity across different contexts. Building upon the work of Athey and Wager (2021), policy learning is relevant in the study of heterogeneous treatment effects. The authors highlight the need of understanding individual characteristics and their impact on treatment assignment rules. This concept is highly applicable in the study of RCTs in the six countries, where the effects of microcredit interventions can vary substantially across recipients due to unique individual characteristics and contextual constraints. Recognising and investigating these variations can lead to refined policy designs, ultimately improving the effectiveness of microfinance programs. Therefore, investigating heterogeneity to predict treatment effects remains an essential undertaking for future research in microfinance. It will equip policymakers with critical information to inform program design and implementation, ultimately improving the lives of the poorest and most vulnerable households.

The study of heterogeneity in treatment effects has seen significant progress over the years and has gone from using subgroup analysis to the emergence of machine learning methods to handle high dimensional datasets and complex interaction effects. The details about the evaluation of the methodologies of the field go beyond the scope of this paper and are therefore omitted in the main text. Please refer to Appendix A.1 or to the book of Hastie et al. (2009) for more information on ML methods and their use in this context.

3 Data

The data sets used in this study come directly from the studies investigating the impacts of microcredit in combatting poverty. All the studies were published by the *American Economic Association*, and despite each one being independently conceived and executed, all authors agreed to employ analogous outcomes and estimation techniques. Additionally, the data used in each study is available in a convenient format on the *American Economic Journal: Applied Economics*[2]. It is important to remember that the study focuses on heterogeneity in treatment effects on four household outcome variables: the amount of money borrowed, revenues from

[2]The journal in which all the studies are published can be reached at `https://www.aeaweb.org/issues/360`. This site provides access to the individual dataset for each of the six studies.

self-employment activities, profits from self-employment activities and consumption. The choice of these variables is motivated by the availability of the variables in all the different RCTs as well as well as the heterogeneity analysis of the RCT in Morocco by Chernozhukov et al. (2022). In their analysis in Morocco, the authors did not have a variable measuring revenues from self-employment activities. Instead, they analyzed output from self-employment activities. The same strategy is employed here. The control variables for each study will be explained in details in the following subsections. For the individual studies in Morocco and India, indicators for missing observations at baseline are also included as controls. The setting is slightly different for the pooled analysis across the six countries and is explained in section 3.3.

3.1 RCT in Morocco

The first RCT to be considered is the one in Morocco, where Al Amana, the largest microfinance institution in country as of December 2012, was involved. The main product that is offered is a group liability loan, as stated in Crépon et al. (2015). Such groups are composed of three to four members who jointly ensure loan repayment. The loan amounts per member can vary from 1,000 to 15,000 Moroccan Dirhams (MAD), equivalent to US$125 to US$1,855 based on market exchange rates. To qualify, candidates must be within the age range of 18 to 70 years, possess a national ID card, provide a residency certificate, and have been engaged in an economic activity excluding non-livestock agriculture for a minimum of 12 months (Crépon et al., 2015). Contrary to most of the implementing MFIs, Al Amana is a nonprofit organisation and does not restrict its loans exclusively to women. To conduct the experiment, Al Amana identified 162 villages in their intervention region. As detailed by Crépon et al. (2015), these villages were arranged into 81 pairs, taking into account certain characteristics such as the number of households, proximity to community centers, existing infrastructure, and the nature of activities conducted by the households, as well as their agricultural activities. Each pair was randomly divided into a treatment village and a control village. In the treatment villages, representatives began promoting microcredit and providing loans. They visited the villages weekly, engaging in promotional activities that included door-to-door campaigning, meetings with current and potential clients, and liaisons with village associations, cooperatives, and women's centers (Crépon et al., 2015). Control groups did not receive any intervention. The MFI started its intervention in 2007 and conducted endline household survey between 2008 and 2010. The overall sample size of the study was 5,524 households, of which 2,730 were assigned to the treatment group and the remaining 2,794 to the control group.

The covariates investigated in this research include initial household characteristics like the quantity of members, number of adults, age of the household head, engagement in animal husbandry, non-agricultural activities, and instances where another household member responded to the survey. Additionally, 81 village pair fixed effects were incorporated. Given the absence of village-level characteristics in the dataset, these village pair fixed effects can be considered as a comprehensive set of proxy variables for unobserved village-level characteristics. To maintain consistency with the original analysis of Crépon et al. (2015), standard errors are adjusted to account for clustering at the village level. The descriptive statistics for the variables used in the

Table 1: Descriptive Statistics of Households in Morocco

	All	Treated	Control
Outcome Variables			
Total amount of loans	2359.404	2929.569	1802.299
Total output from self-employment activities (past 12 months)	32499.088	35148.117	29910.734
Total profit from self-employment activities (past 12 months)	10102.149	11034.919	9190.744
Total monthly consumption	3011.862	2996.003	3027.374
Baseline Covariates			
Number of household members	3.879	3.872	3.886
Number of members 16 years old or older	2.604	2.601	2.607
Household head age	35.976	35.936	36.014
Declared animal husbandry self-employment activities	0.415	0.426	0.404
Declared non-agricultural self-employment activities	0.146	0.129	0.164
Borrowed from any source	0.210	0.224	0.196
Spouse of head responded to self-employment section	0.067	0.074	0.061
Member responded to self-employment section	0.044	0.048	0.041
Number of observations	5,524	2,730	2,794

Note. The values presented are the means of the variables for the whole sample, the treatment group and the control group. All monetary values are expressed in Moroccan Dirhams, MAD.

Morocco analysis are displayed in Table 1. All monetary variables are expressed in MAD. The table displays comparable characteristics between treatment and control group households. The intervention's impact is reflected in the differences in loans, outputs, profits, and consumption, which are 1,127, 5,237, 1844, and -31 respectively. Particularly for consumption, the mean total monthly consumption dropped from 3,027 MAD in control households to 2,996 MAD in treated households, suggesting that the intervention led to a reduction of 31 MAD in consumption.

3.2 RCT in India

The next RCT evaluated in this research involves Spandana, the most significant and rapidly expanding MFI in India during the period of study, as reported by Banerjee, Duflo, Glennerster and Kinnan (2015). Spandana's offering was also a group-loan product where each group comprised 6 to 10 women who collectively shared responsibility for the group's loans. The first loan amounted to Indian Rupees (Rs.) 10,000, and if all group members successfully repaid their loans, they qualified for second loans ranging from Rs. 10,000 to 20,000 (equivalent to US$200 to US$400 based on market exchange rates). Eligible clients had to be female, between 18 and 59 years old, have lived in the same area for at least a year, possess valid identification and residential proof, and at least 80 percent of the women in a group had to own their own homes. Spandana is a for-profit operator, meaning that they charge interest rates sufficient to make profits (Banerjee, Duflo, Glennerster & Kinnan, 2015). To conduct the experiment, the MFI identified 104 neighborhoods in Hyderabad where it could implement its product. As in Morocco, the areas were paired based on per capita consumption and per-household debt, resulting in 52 pairs of neighborhoods. One neighborhood from each pair was randomly selected to be the treatment group. Spandana opened branches in treatment communities but not in control ones. The MFI began operating in the treatment areas between 2006 and 2007 and conducted a first endline survey 15 to 18 months after the start of the treatment. Two years later, a second endline survey was conducted among the same households. However, by that time, Spandana and other MFIs had begun their operations in both groups. Therefore, the proportions of households borrowing was relatively similar across the two groups. This second endline survey is not considered in this study and the analysis focuses on the first endline survey. The overall sample size of the study was 6,236 households, of which 2,970 are part of the treatment group.

Table 2: Descriptive Statistics of Households in India

	All	Treated	Control
Outcome Variables			
Total amount of loans	59,804.286	62,948.036	56,347.219
Total revenues from self-employment activities (past 30 days)	4,652.641	5,180.870	4,071.767
Total profits from self-employment activities (past 30 days)	968.642	1,172.107	744.898
Total monthly consumption	-0.013	-0.013	-0.014
Baseline Covariates			
Has business for year or more before endline 1	0.261	0.263	0.259
Area population	309.754	304.338	315.710
Total outstanding debt in area	33,786.037	31,612.275	36,176.443
Total number of businesses in area	7.093	6.747	7.473
Area mean monthly per-capita expenditure	994.607	1,005.134	983.032
Area literacy rate (HH heads only)	0.618	0.622	0.613
Area literacy rate	0.681	0.683	0.680
Number of observations	6,236	2,970	3,266

Note. The values presented are the means of the variables for the whole sample, the treatment group and the control group. All monetary values are expressed in Indian Rupees, Rs.

In this RCT, the covariates are calculated as area-level baseline values such as area population, total businesses, average per capita expenditure, fraction of households heads who are literate, and fraction of all adults who are literate (Banerjee, Duflo, Glennerster & Kinnan, 2015). There is also a variable indicating whether the household had previous business experiment. Baseline household characteristics have not been collected. Standard errors are adjusted for clustering at the area level to stay consistent with the original analysis of Banerjee, Duflo, Glennerster and Kinnan (2015). Table 2 shows descriptive statistics for the variables used in the India analysis. All monetary variables are expressed in Indian Rupee, Rs. In terms of baseline covariates, it seems that treated and control households have similar characteristics. The difference between the treatment and the control group after the intervention on loans, outputs, profits, and consumption are respectively 6,601, 1,109, 427, and 0.001. For instance, households in the treatment groups had an average increase in total revenues from self-employment activities of 1,109 Rs. compared to control groups.

3.3 Joint analysis across the six countries

The joint analysis consists of investigating heterogeneity across the RCTs conducted in Bosnia and Herzegovina, Ethiopia, India, Mexico, Mongolia and Morocco. The treatment interventions, involving the local MFI opening a microcredit branch in treated villages, areas or neighborhoods, were consistent across the countries. However, some also included promotions and additional services, accounted for in the analysis via a dummy variable indicating such additional service usage. In addition to the different baseline characteristics that have been collected in each site, the studies differ across several variables. As outlined in Table 3, variations include loan sizes and interest rates, randomization methods, targeting strategies (e.g., focusing on urban areas or women), the use of promotions, and the microcredit market saturation, the latter marked by the "Current Market" variable ranging from 0-3. This variable stems from Meager (2019)'s analysis, highlighting the importance of credit market saturation. Additionally, the time gap varied significantly across different sites. These differences underscore the importance of considering specific contextual factors when evaluating the impact of microcredit interventions.

Table 3: Study Characteristics

	Bosnia	Ethiopia	India	Mexico	Mongolia	Morocco
In urban area	Mix	No	Yes	Mix	No	No
Target Women	No	No	Yes	Yes	Yes	No
Has Promotion as well	No	No	No	Yes	No	Yes
Group Loan?	No	Yes	Yes	Yes	Mix	Yes
Loan Size	1,012$	150$	200$	538$	435$	1,188$
APR	22%	12%	24%	100%	120%	13.5%
Time Gap	14	36	14	27	19	24
Current market	2	1	3	2	1	0
Loan as % of income	9%	118%	22%	6%	36%	21%
Loan term (in months)	14	12	12	14	Mix	16
Village Randomisation	No	Yes	Yes	Yes	Yes	Yes
Year	2010	2006	2007	2012	2009	2009
USD PPP Conversion	$\frac{1}{0.88}$	$\frac{1}{2.29}$	$\frac{1}{11.09}$	$\frac{1}{9.18}$	$\frac{1}{513.24}$	$\frac{1}{4.31}$
USD to 2009 USD	$\frac{100}{101.653}$	$\frac{100}{94.729}$	$\frac{100}{97.101}$	$\frac{100}{106.121}$	1	1

Note. Year indicates the year of collection of the first endline surveys. *USD PPP Conversion* indicates the PPP conversion factor from local currency units to US $ at the corresponding year (World Bank, 2023). *USD to 2009 USD* indicates the conversion factor to 2009 US $ to account for inflation according to the Consumer Price Index (CPI) (US Department of Labor Bureau of Labor Statistic, 2023).

In this multi-country study, covariates collected varied across the six studies under review. Only baseline characteristics available in at least four of the six countries were included in the analysis. For the remaining countries, missing values were replaced with the median. To facilitate a pooled analysis, comparability of variables across different RCTs, each conducted in a unique country with its specific currency and timeline, was ensured by standardizing all variables to US Dollars at Purchasing Power Parity (PPP) over a fortnight. This method follows the approach taken by Meager (2019) in their analysis. All monetary variables were first converted to US Dollars using the World Bank PPP conversion factor[3], which serves as a price deflator and currency converter to account for price level disparities between countries World Bank (2023). Subsequently, to address inflation, the monetary variables were converted to 2009 US Dollar values using the CPI converter provided by US Department of Labor Bureau of Labor Statistic (2023). For profits, revenues, and consumption, additional transformations were performed to render them biweekly and comparable across studies. After data cleaning and processing, the pooled analysis comprised a total of 20,478 observations, with 10,684 assigned to the treatment group.

Table 4: Descriptive Statistics of the pooled analysis

Variable	All	Treated	Control
Outcome Variables			
Total Amount of Loans	2,518.370	2,693.432	2,327.400
Biweekly revenues from self-employment activities	186.326	199.018	172.480
Biweekly profit from self-employment activities	52.380	54.953	49.574
Biweekly consumption	226.911	222.717	231.487
Baseline Covariates			
Number of household members	4.075	4.121	4.022
Number of members 16 years old or older	2.742	2.768	2.711
Age	37.208	37.292	37.116
Household has a previous business	0.359	0.361	0.357
Gender is female	0.982	0.978	0.986
At least primary school	0.677	0.689	0.663
In urban area	0.606	0.604	0.608
Number of observations	20,478	10,684	9,794

Note. The values presented are the means of the variables for the whole sample, the treatment group and the control group. All monetary values are expressed in 2009 USD at PPP.

Along with the unique characteristics of each study, baseline covariates considered in this analysis include the number of household members, the number of adults, age, and indicators for previous business experience, gender, base level education completion, and urban residency. To control for potential spillovers, country-level and area-level fixed effects are incorporated. Standard errors are clustered at the country level, as justified by Meager (2019). The descriptive statistics for these variables, presented in Table 4, show that the treatment and control groups exhibit similar characteristics. The average treatment effects on loans, outputs, profits, and consumption are respectively 366, 27, 5, and -9 USD. For example, households in treatment groups borrowed an average of 366 USD more than those in control groups.

[3]The conversion rate used is the PPP conversion factor from local currency units to US $ at the time the endline survey was collected. Cf. Table 3 for the exact rates.

4 Methodology

4.1 General setting

This paper investigates the setting of a randomized control trial, trying to understand if an event X causes an event Y. Here, Y represents the observed outcome variable for a given observation. The causal effect of a treatment on an individual unit is measured as the difference between the outcome variable with and without the treatment. However, the fundamental problem in causal inference is that we cannot simultaneously observe the same unit in treated and untreated states, leading to the concept of potential outcomes (Holland, 1986). These denote the hypothetical outcomes under treatment and control states, represented as Y_{1i} and Y_{0i}, respectively for observation i. Let Z represent specific characteristics affecting the observations in the study. The average treatment effect can be estimated as:

$$ATE = E[Y_1 - Y_0].$$

By randomly assigning individuals to treatment or control groups, RCTs eliminate bias, making potential outcomes independent of the treatment.

Of interest here are the "baseline conditional average" (BCA) and the "conditional average treatment effect" (CATE). The former, denoted as $b_0(Z)$ can be considered as the average performance for the observation that did not undergo the treatment, given Z. Mathematically, it can be expressed as:

$$b_0(Z) := E[Y_0|Z].$$

It provides a baseline against which the effect of the program can be assessed by comparing it to the actual outcomes observed for the treatment group. The latter, denoted by $s_0(Z)$, is the expected difference in performance if a subject is in the treatment group, versus if they are not, again considering their specific characteristics. The CATE is represented mathematically as:

$$s_0(Z) := E[Y_1 - Y_0|Z] = E[Y_1|Z] - E[Y_0|Z].$$

The main goal is to determine the impact of the treatment by estimating the BCA and the CATE. For this purpose ML techniques help predict expected outcomes given all subjects' characteristics and their treatment status. The methodology developed by Chernozhukov et al. (2022) is particularly useful in this context. Instead of directly estimating the BCA or CATE, they suggest strategies to estimate and obtain inference on key features of $s_0(Z)$ rather than $s_0(Z)$ itself, overcoming the need for strong assumptions about ML estimators properties. This approach, referred to as Generic ML, follows three stages. First, it plits the data $(Y_i, D_i, Z_i)_{i=1}^N$ into a main sample, denoted by $Data_M$, and an auxiliary sample, $Data_A$, where D_i is a binary variable indicating treatment group membership. Second, it uses the auxiliary sample $Data_A$ and some ML technique to obtain estimates of baseline and treatment effects of $b_0(Z)$ and $s_0(Z)$, known as proxy predictors:

$$z \mapsto B(z) = B(z; Data_A) \text{ and } z \mapsto S(z) = S(z; Data_A).$$

This step does not require consistency, as the estimates will likely not be consistent across different learners. Finally, it uses these proxies to estimate and infer the key features of the CATE using the main sample $Data_M$. These features include:

1. The Best Linear Predictor (BLP)
2. The Sorted Group Average Treatment Effects (GATES)
3. The Classification Analysis (CLAN)

There are two types of uncertainties involved in this process. First, the "estimation uncertainty" which is conditional on the auxiliary sample in step two, and second, the "splitting uncertainty" which arises from data partitioning in step one. These challenges are addressed by using many data splits to produce robust estimators. Specifically, after setting a significance level α, the key features are calculated across the data splits, and the median of each key feature parameter is then taken across the splits. This procedure, known as "Variational Estimation and Inference" (VEIN), provides inference on each key feature parameter, ensuring a size control level of 2α. This process can be replicated with multiple machine learners. In this specific research, 100 splits into main and auxiliary samples are used, a method that was previously applied by Chernozhukov et al. (2022) on the Moroccan dataset, yielding consistent and robust outcomes.

4.2 Best Linear Predictor

The BLP aims to answer the question: *"Is there evidence of treatment effect heterogeneity?"*. It is defined by

$$BLP[s_0(Z)|S(Z)] := \beta_1 + \beta_2(S(Z) - E[S(Z)]),$$

and can be estimated using Ordinary Least Squares (OLS) or Weighted Least Squares (WLS). In this context, $S(Z)$ is an unobserved proxy estimated via ML using the auxiliary sample data $Data_A$. Once the predicted scores are estimated, coefficients β_1 and β_2 are estimated using OLS or WLS. The inputs to the regressions are therefore the ML proxy $S(Z)$ and the observed treatment effects $s_0(Z)$. In these calculations, $\beta_1 = E[s_0(Z)]$ is the ATE, and $\beta_2 = \frac{Cov[s_0(Z),S(Z)]}{Var[S(Z)]}$ represents the level of heterogeneity in treatment effect. The later captures the relationship between the CATE $s_0(Z)$ and the proxy predictor $S(Z)$. It can be interpreted as the slope of the relationship between the treatment effect and the proxy. If β_2 is significantly different from zero, it indicated evidence of treatment effect heterogeneity. Essentially, the BLP is a filtered predictor of the CATE compared to the ML proxy. The estimates of interest are asymptotically normal and valid confidence bounds can be constructed. If $S(Z)$ is a perfect proxy, then $\beta_2 = 1$. In general, the coefficient will be different than 1 because of noise in $S(Z)$. If the proxy is complete noise, and if there is no heterogeneity, $\beta_2 = 0$. In general $\beta_2 \neq 0$ if there is heterogeneity in $s_0(Z)$ and $S(Z)$ predicts it well. Therefore, evidence of treatment effect heterogeneity is obtained by testing if β_2 is statistically different from zero.

4.3 Sorted Group Average Treatment Effects

The GATES aims to answer the question: *"What are the treatment effects for the different groups of households?"*. Thus, the goal is to find out how treatment effect varies across households. To achieve this, groups are created to explain as much variation in $s_0(Z)$ as possible. Mathematically, K groups are defined as $G_k := S(Z) \in I_k$, for $k = 1, ..., K$, and where $I_k = [l_{k-1}, l_k)$ divide support of $S(Z)$ into regions. After the groups have been defined, the primary parameters of interest are γ_k which are the average treatment effects per subgroups and defined as:

$$\gamma_k := E[s_0(Z)|G_k].$$

The most common strategy to estimate them is via Weighted Residuals. In particular, the outcome variable Y is expressed as a function of the covariates X_1 and a weighted sum of treatment interactions $[D - p(Z)]$ with group indicators $I(G_k)$. The full weighted linear projection is given by:

$$Y = \alpha_0' X_1 + \sum_{k=1}^{K} \alpha_k \cdot [D - p(Z)] \cdot I(G_k) + v, \quad E[w(Z)vW] = 0,$$

where $W := (X_1', W_2')$, X_1 contains a vector of functions of Z, and $W_2 := (\{[D - p(Z)] \cdot I(G_k)\}_{k=1}^{K})'$. This way, the model identifies the GATES by exploiting variation within each group defined by $S(Z)$, and control variables in X_1 are used to reduce estimation noise. In practice, an empirical analog of this model is fitted to the data. This involves replacing the theoretical quantities with their observed counterparts in the sample. For each observation i in the sample, Y_i is regressed on X_{1i} and a set of group-specific treatment interaction as:

$$Y_i = \hat{\alpha}_0' X_{1i} + \hat{\alpha}' W_{2i} + \hat{v}_i, \quad i \in M.$$

The estimated coefficients $\hat{\alpha} = (\hat{\alpha}_1, ..., \hat{\alpha}_K)$ provide the estimated GATES.

4.4 Classification Analysis

Finally, CLAN aims to answer the question: *"What characteristics are associated with treatment effect heterogeneity?"*, when there is evidence of heterogeneity. In particular, the average characteristics of the most and least affected groups G_1 and G_K are compared using two-sample t-tests. The average characteristics of the groups are defined as:

$$\delta_k := E[g(Y, D, Z)|G_k].$$

In the empirical analysis, CLAN parameters are estimated by taking the empirical expecation over the main dataset $Data_M$ of size N denoted as $E_{N,M}$. The average across all observations in the main sample is expressed as:

$$\hat{\delta}_k = \frac{E_{N,M}[g(Y_i, D_i, Z_i) G_{k,i}]}{E_{N,M} G_{k,i}},$$

where $G_{k,i} = I\{S(Z_i) \in I_k\}$, and $I_k = [l_{k-1}, l_k)$ divide the proxy into quantiles. Specifically, l_k values are thresholds based on the CATEs used to divide the distribution of the proxy $S(Z)$ into quantiles to define the groups G_k. Each group G_k contains those individuals for whom the score function $S(Z)$ falls within the above defined interval I_k. The outputs of the CLAN are the characteristics that are associated with the heterogeneity in the CATE.

4.5 Machine learning methods

As mentioned, this procedure can be performed with any ML method. In this research, the ML methods employed align with those used in the heterogeneity evaluation of the RCT in Morocco, performed by Chernozhukov et al. (2022). In particular, four methods are considered to estimate the proxy predictors: Elastic Net, Extreme Gradient Boosting, Support Vector Machines, and Random Forest. The reason behind the selection of these methods is their demonstrated efficiency in handling complex, high-dimensional datasets, as well as their compatibility with the methodology of Chernozhukov et al. (2022). In terms of hyperparameters, the defaults ones are used for all of the methods except for Elastic Net where the parameter to control the mixing of Ridge ($\alpha = 0$) and Lasso ($\alpha = 1$) is set to $\alpha = 0.5$ to obtain an equal balance of Lasso and Ridge penalties and combining the benefits of both methods.

4.6 Evaluation measures

To guide the selection of ML proxies, Chernozhukov et al. (2022) suggested goodness-of-fit measures. In particular, two performance measures are computed for each learner, one based on the BLP of CATE and one based on the GATES. For the CATE, the performance measure is:

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

and for the GATES:

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

For the first performance measure, maximizing $\hat{\Lambda}$ is the same as maximizing the correlation between $S(Z)$ and $s_0(Z)$, or equivalent to maximizing the R^2 in the regression of $s_0(Z)$ on $S(Z)$. For the latter, the procedure is the same with the only difference that the data is divided into subgroups. Therefore the best machine learner is the one that attains a higher $\hat{\Lambda}$ for the BLP of CATE, and the one that attains a higher $\hat{\bar{\Lambda}}$ for the GATES.

4.7 Additional adjustments

The procedure for detecting heterogeneity, as detailed in Algorithm 1 in Appendix A.2, is adjusted for potential correlation among observations. For the meta analyses, fixed effects are created at the pair or cluster level, and standard errors are adjusted at the village level. For the pooled analysis, country fixed effects are added, and standard errors are adjusted at the country level, following the recommendations of Abadie et al. (2017). The entire analysis will be performed using the **GenericML** package by Welz et al. (2022), and the code will be executed in parallel using six cores on a 1.4 GHz Quad-Core Intel Core i5 processor.

5 Results

5.1 Heterogeneity analysis in Morocco

In Morocco, the propensity score is defined as constant and is therefore the mean of the variable D, which is $p(Z_i) = 0.49$ for all the observations. The comparison of the four ML methods is first presented before diving deeper into the potential heterogeneity, as shown in Table 5. Given its large performance metrics, Random Forest performs best for the amount of loans and profits, while Elastic Net excels for total output and consumption. Accordingly, the subsequent analysis will focus on the top-performing method for each variable.

Table 5: Comparison of ML Methods: Microfinance Availability in Morocco

	Elastic Net	Boosting	SVM	Random Forest
Amount of Loans				
Best BLP(Λ)	545,580	560,062	452,612	1,835,148
Best GATES($\bar{\Lambda}$)	2,203,307	2,013,648	2,290,265	2,680,322
Output				
Best BLP(Λ)	76,733,172	12,255,390	32,584,559	19,031,281
Best GATES($\bar{\Lambda}$)	150,984,159	72,186,949	100,044,258	116,359,104
Profit				
Best BLP(Λ)	6,617,312	2,640,473	7,202,861	14,762,207
Best GATES($\bar{\Lambda}$)	19,638,743	18,606,405	23,799,971	33,224,878
Consumption				
Best BLP(Λ)	9,924	8,563	*3,685*	8,526
Best GATES($\bar{\Lambda}$)	37,942	32,788	*23,084*	32,132

Note. Medians over 100 splits in half. *_* For the analysis of *Consumption*, SVM did not provide accurate estimation, therefore this column is replaced by the Neural Network estimation, as was done in the original analysis of heterogeneity on this dataset by Chernozhukov et al. (2022).

Table 6: BLP of Microfinance Availability in Morocco

	ATE (β_1)	HET (β_2)
Amount of Loans	1,138.369	0.317
	(273; 1,897)	(-0.016; 0.661)
	[0.01]***	[0.059]*
Output	5,228	0.288
	(-2,045; 12,813)	(0.001; 0.558)
	[0.159]	[0.047]**
Profit	1,554	0.186
	(-2,539; 5,551)	(-0.021;0.396)
	[0.472]	[0.074]*
Consumption	-64.55	0.144
	(-215.185;86.55)	(-0.192;0.467)
	[0.404]	[0.352]

Note. Medians over 100 splits in half. 90% confidence intervals are in parentheses and p-values in square brackets. ***, **, and * indicate significance at the 1%, 5% and 10% level, respectively.

The results of the BLP of CATE, presented in Table 6, include the ATE and HET loading parameters β_1 and β_2. The amount of loans is the only variable with a statistically significant ATE at the 10% level, consistent with Crépon et al. (2015)'s findings. The scale of the ATEs aligns with the findings from the authors and with the unconditional ATE (1,127, 5,237, 1,844, & -31, cf. section 3.1), as expected by definition of the randomization. In terms of heterogeneity, the hypothesis that HET is zero is rejected at the 10% level for the amount of loans, business outputs, and business profits, suggesting heterogeneity in microfinance's effects on these variables. Yet, no significant heterogeneity is found for consumption. This implies that while microfinance has heterogeneous impacts on business-related outcomes, it does not appear to alter living standards (represented by consumption) in a detectable way (Chernozhukov et al., 2022). One possible explanation brought forward by Chernozhukov et al. (2022) is that households most likely to borrow and profit from microfinance compensate by reducing labor supply, aligning with Crépon et al. (2015)'s findings.

Table 7: GATES of 20% Most and Least Affected Groups in Morocco

	20% Most (γ_5)	20% Least (γ_1)	Difference ($\gamma_5 - \gamma_1$)
Amount of Loans	2,792.9	-142.0	2,889.4
	(546.5; 4,811)	(-2,168.3; 2,033)	(-103.5; 5,865)
	[0.013]**	[0.904]	[0.053]*
Output	23,902.7	-1,362.4	25,221.5
	(2,850.2; 44,028)	(-13,985.7; 9,569)	(1,199.4; 48,046)
	[0.025]**	[0.831]	[0.040]**
Profit	10,478.4	-1,316.4	11,471.9
	(-141.4; 21,703)	(-9,273.4; 6,926)	(-2,283.5; 25,667)
	[0.050]**	[0.752]	[0.102]
Consumption	24.40	-327.10	341.00
	(-256.58; 332.1)	(-816.70; 126.9)	(-259.24; 901.6)
	[0.850]	[0.150]	[0.277]

Note. Medians over 100 splits in half. 90% confidence intervals are in parentheses and p-values in square brackets. ***, **, and * indicate significance at the 1%, 5% and 10% level, respectively.

Next, we examine the GATES. Households are split into five groups based on quantiles, with the ATE estimated per group. Coefficients $\gamma_1 - \gamma_5$ are presented in Figures 1-4 in Appendix A.2, revealing significant variation in the effects on loan amount, outputs and profits. The GATES are further investigated by comparison of the most and least affected groups in Table 7. The difference in GATES of the two groups is statistically different from zero at least at the 10% significance level for the amount of loans and business outputs and near-significant for profits. For consumption, the hypothesis that the difference is zero cannot be rejected. Notably, no evidence suggests any significant negative impact on profit and output for the least affected groups, mitigating concerns of adverse effects.

The final heterogeneity analysis is investigating what drives the heterogeneity in the data using CLAN. In this RCT, two sets of covaraites were used to predict heterogeneity: baseline household characteristics and village pair fixed effects. Consistent with the discoveries of Chernozhukov et al. (2022), the predictive capacity for treatment effect heterogeneity is significantly attributed to the village pair fixed effects, as most heterogeneity drivers identified by the CLAN were the same village pair fixed effects. This indicates that village-level variables account for a substantial portion of the variability in treatment effects. Nevertheless, these variables are not easily interpretable, implying challenges in predicting individually who would embrace or benefit from microfinance. On a more positive note, it highlights the benefits of studying heterogeneity in different contexts, such as the RCT in India and the pooled analysis across six countries.

Finally, the results obtained are consistent with Chernozhukov et al. (2022)'s analysis, using the same ML methods. Minor variations in GATES results can be attributed to different seeds, splits, and system configurations affecting the random number stream. Additionally, different hyperparameters have been used in the original analysis. Re-running the analysis using the authors' original code produced almost identical results, confirming consistency between the methods. Nevertheless, the motivation for presenting the results obtained with the package is that is ensures replicability and comparability across the data sets.

5.2 Heterogeneity analysis in India

Next, the results for the RCT conducted in India by Banerjee, Duflo, Glennerster and Kinnan (2015) are presented. The propensity score of the analysis is again the mean of the variable D and is $p(Z_i) = 0.52$ for all the observations. First, the comparison of the ML methods is presented before diving into the potential heterogeneity in treatment effect. The performance metrics for the different methods are presented in Table 8. It is found that Elastic Net performs best for the amount of loans, SVM for the profits, and Boosting outperfoms the other methods for business revenues and consumption because the values of Λ and $\bar{\Lambda}$ are larger for these methods. For the remainder of the analysis, the results presented for each outcome variable will be the ones obtained from the best method for each outcome.

Table 8: Comparison of ML Methods: Microfinance Availability in India

	Elastic Net	Boosting	SVM	Random Forest
Amount of Loans				
Best BLP(Λ)	153,070,479	103,643,807	136,028,303	28,553,073
Best GATES($\bar{\Lambda}$)	380,308,843	327,794,473	482,218,527	316,717,190
Revenue				
Best BLP(Λ)	11,559,108	49,757,953	41,473,842	9,762,983
Best GATES($\bar{\Lambda}$)	21,902,941	26,350,719	34,598,950	16,504,139
Profit				
Best BLP(Λ)	1,241,105	666,514	1,973,616	327,217
Best GATES($\bar{\Lambda}$)	2,657,700	2,520,371	3,250,901	2,705,730
Consumption				
Best BLP(Λ)	0.001	0.009	0.009	0.001
Best GATES($\bar{\Lambda}$)	0.006	0.017	0.015	0.006

Note. Medians over 100 splits in half.

Table 9: BLP of Microfinance Availability in India

	ATE (β_1)	HET (β_2)
Amount of Loans	7,379.863	0.498
	(-6,032.043; 20,988.294)	(-0.188; 1.338)
	[0.271]	[0.122]
Revenue	1,128.877	1.646
	(-1,213.384; 3,605.590)	(0.362; 2.980)
	[0.344]	[0.010]***
Profit	346.905	1.095
	(-378.989; 1,116.920)	(0.210; 2.041)
	[0.320]	[0.011]**
Consumption	-0.016	0.548
	(-0.074; 0.038)	(0.209; 0.913)
	[0.585]	[0.001]***

Note. Medians over 100 splits in half. 90% confidence intervals are in parentheses and p-values in square brackets. ***, **, and * indicate significance at the 1%, 5% and 10% level, respectively.

The coefficients β_1 and β_2 in Table 9 represent the ATE and HET loading parameters in the BLP of CATE using ML proxies $S(Z)$. The estimates of the ATEs of microfinance availability in India are consistent with the findings of Banerjee, Duflo, Glennerster and Kinnan (2015) and are similar to the unconditional ATE (6,601, 1,109, 427, and 0.001, cf. section 3.2). None of the ATE are statistically significant at least at the 10% level for the best ML proxies. Therefore, microfinance availability did not have an impact on any of the outcome variables at the first endline. Banerjee, Karlan and Zinman (2015) suggest that the lack of evidence of transformative effects on the average borrower may hide the possibility of transformative effects - good for some, bad for others - on certain subgroups of microlenders' target populations. Indeed, the hypothesis that HET is zero is rejected at the 5% level for profits, and at the 1% level for revenues and consumption, suggesting that microfinance has varied effects on different business-related outcomes. Interestingly, these effects lead to a significant immediate improvement in the quality of life, as depicted by consumption, for the households most positively influenced. This differs from the scenario in Morocco, where households most likely to borrow from microfinance often offset the increased profits by reducing their labor supply (Crépon et al., 2015).

The estimates of the GATES are presented next. As in Morocco, the households are divided into $K = 5$ groups based on the quantiles of the ML proxy predictor and the average effect are estimated for each group. The coefficients $\gamma_1 - \gamma_5$ along with joint confidence bonds are displayed in Figures 5 - 8 in Appendix A.2, alongside the ATE and its interval. These figures reveal quite clearly that there are groups of winners, the most affected groups, for which the GATES on revenues, profits and consumption are significantly different from zero. There also appears to be some group of losers, the least affected groups, for which the GATES on consumption is negative and significantly different from zero, cf. Figure 8. These groups are likely to drive the heterogeneity in the treatment effect found in the BLP analysis. Comparing these groups in Table 10, the analysis shows that the difference of GATES of the two groups is significantly different from zero for the three outcome variables at least at the 5% level, and particularly

Table 10: GATES of 20% Most and Least Affected Groups in India

	20% Most (γ_5)	20% Least (γ_1)	Difference ($\gamma_5 - \gamma_1$)
Amount of Loans	26,663	-8,175	32,245
	(-15,074; 60,411)	(-36,981; 17,923)	(-20,331; 84,178)
	[0.174]	[0.598]	[0.192]
Revenues	10,548	-3,028	14,254
	(195; 19,955)	(-9,435; 2,996)	(1,358; 26,112)
	[0.043]**	[0.273]	[0.028]**
Profit	2,673.44	-1,566.94	4,282.53
	(-128.50; 5,476.60)	(-3,591.39; 588.60)	(612.53; 8,232.90)
	[0.064]*	[0.158]	[0.022]**
Consumption	0.163	-0.136	0.295
	(0.029; 0.291)	(-0.276; 0.003)	(0.096; 0.509)
	[0.016]**	[0.058]*	[0.004]***

Note. Medians over 100 splits in half. 90% confidence intervals are in parentheses and p-values in square brackets. ***, **, and * indicate significance at the 1%, 5% and 10% level, respectively.

pronounced in consumption at the 1% level. For the "losers", consumption shows a significant negative effect, potentially due to households decreasing consumption to enhance investment as suggested by Chernozhukov et al. (2022). The remainder of the outcomes show no significant negative impact, which is reassuring. The "winners", on the other hand, reflect significant positive impacts on consumption, profits, and revenues, possibly due to increasing spending on durables, as Banerjee, Duflo, Glennerster and Kinnan (2015) noted. For these households, the increase in durable spending outweighs the reduced spending on non-essentials or "temptation goods".

After presenting evidence on the heterogeneity of treatment effects for the four outcomes, we examine what drives this heterogeneity in the data using CLAN. Unlike in Morocco, the India RCT incorporated area-level baseline values such as area population, average per capita expenditure, and fraction of all adults who are literate. Therefore, if these covariates drive the heterogeneity, it could potentially be possible to identify village-level success factors in microfinance, and guide future policy. Table 11 shows the average baseline characteristics of the most and least affected groups for the four more significant characteristics identified by CLAN for the heterogeneity-affected outcomes. Area population, total number of businesses and the previous business experience consistently appear for the three outcomes. Regarding profits, the most impacted groups have more business experience. This result is expected. Indeed, Meager (2019) found that the differential effect in profits is robust and generalizeable across studies. Research works like Banerjee et al. (2017) and Banerjee, Karlan and Zinman (2015) discovered that credit substantially influenced the business results of individuals who established businesses before the introduction of microfinance, more than those who didn't have businesses beforehand. A similar result was found in the analysis in Morocco and this confirms the benefits of targeting households with previous business experience. The results are more contradicting regarding revenues and consumption. On the one hand, regarding revenues, the most impacted groups live in more populated areas with more businesses, but the households in these areas have less businesses experience. On the other hand, for consumption, the most affected live in less populated areas with fewer businesses, but the households in these areas have more business experience. This counterintuitive result suggests that the impact of microcredit on consumption is not merely a function of business density and population, but could also be influenced by other factors. A possible explanation for such results is differences in market competition and population density. Indeed, in areas with more businesses, increased competition might lead to higher revenues due to a larger customer base. On the other hand, in areas with fewer businesses, households with more business experience might be able to better optimize their consumption, leading to a larger impact (Jovanovic, 1982). It is also possible that the population factor plays a role, as it can be expected that more populated areas could provide larger markets, boosting revenues, while less populated areas may see reduced competition and, therefore, lower consumption (Desmet & Henderson, 2015; Porter, 2008).

Table 11: CLAN of Microfinance Availability in India

	20% Most (δ_5)	20% Least (δ_1)	Difference ($\delta_5 - \delta_1$)
Revenue			
Total number of businesses in area	7.917 ***	6.273 ***	1.611 ***
Has business for year or more before endline 1	0.398 ***	0.546 ***	-0.139 ***
Total outstanding debt in area	39,596 ***	33,324 ***	6,359 ***
Area Population	327.850 ***	304.953 ***	21.147 ***
Profit			
Total number of businesses in area	8.721 ***	4.731 ***	3.927 ***
Area Population	286.1 ***	403.5 ***	-115.6 ***
Area literacy rate	0.667 ***	0.695 ***	-0.027 ***
Has business for year or more before endline 1	0.429 ***	0.362 ***	0.066 **
Consumption			
Area Population	285.700 ***	341.170 ***	-55.860 ***
Has business for year or more before endline 1	0.421 ***	0.252 ***	0.158 ***
Area literacy rate	0.669 ***	0.695 ***	-0.028 ***
Total number of businesses in area	6.592 ***	8.187 ***	-1.533 ***

Note. Medians over 100 splits in half. ***, **, and * indicate significance at the 1%, 5% and 10% level, respectively.

Interestingly, the most affected groups for profits and consumption live in areas with lower literacy rates and smaller populations. Contrary to expectations and research from Budiono et al. (2021), areas with higher literacy rates did not leverage the borrowing opportunity. Nonetheless, several studies back up the claim that microcredit can have significant impact in areas of lower literacy rates. For instance, Coleman (2006) showed that microfinance often provides the poor and illiterate with the financial resources necessary to improve their economic situation. Also, Cull et al. (2007) show that microcredit can be particularly impactful in less literate rural areas.

In general, the most impacted groups generally possess more prior business experience, lower literacy rates, live in less populated areas, and areas with fewer businesses. These conditions suggest that microfinance can significantly benefit a particular population subset in India.

5.3 Heterogeneity analysis across six countries

Finally, the results of the pooled analysis across the six countries are presented. By design, the propensity score is the mean of the variable D, which is $p(Z_i) = 0.52$ for all the observations. In terms of the best performing ML methods, it is found that that Random Forest performs best for the amount of loans, revenues and profits, while Elastic Net performs best for consumption, as shown in Table 12. Consequently, the rest of the analysis will be focused on these methods.

Table 12: Comparison of ML Methods: Microfinance Availability Across Countries

	Elastic Net	Boosting	SVM	Random Forest
Amount of Loans				
Best BLP(Λ)	66,740	163,473	18,916	559,551
Best GATES($\bar{\Lambda}$)	279,820	303,642	337,320	504,041
Revenue				
Best BLP(Λ)	19,817.500	28,042.500	76.900	31,230.900
Best GATES($\bar{\Lambda}$)	5,302	3,794	2,722	11,603
Profit				
Best BLP(Λ)	60.320	85.620	37.750	368.140
Best GATES($\bar{\Lambda}$)	484.000	404.500	593.400	1,095.400
Consumption				
Best BLP(Λ)	332.106	35.012	1.559	135.998
Best GATES($\bar{\Lambda}$)	355.270	65.730	79.420	274.630

Note. Medians over 100 splits in half.

Table 13: BLP of Microfinance Availability Across Countries

	ATE (β_1)	HET (β_2)
Amount of Loans	406.021	0.475
	(66.165; 633.732)	(0.334; 0.541)
	[0.011]*	[0.000]***
Revenue	33.214	0.790
	(-1.514; 66.093)	(0.158; 1.424)
	[0.064]*	[0.012]**
Profit	7.837	0.193
	(-11.160; 22.431)	(0.044; 0.329)
	[0.437]	[0.009]***
Consumption	-1.584	0.335
	(-7.752; 4.649)	(0.108; 0.593)
	[0.601]	[0.002]***

Note. Medians over 100 splits in half. 90% confidence intervals are in parentheses and p-values in square brackets. ***, **, and * indicate significance at the 1%, 5% and 10% level, respectively.

The results of the BLP of CATE for the four outcome variables using the ML proxies are presented in Table 13. The coefficients β_1 and β_2 corresponding to the ATE and the HET parameters in the BLP are reported. The estimated ATEs of microfinance availability are mostly consistent with the findings of the full pooling analysis of Meager (2019) who finds ATE of 22.5, 7.3,and 4.6 for revenues, profits, and consumption respectively. The author did not estimate the ATE for the amount of loans. These ATE are also similar to the unconditional ATE (366, 27, 5, and -9, cf. section 3.3). The ATE for the amount of loans and revenues are statistically significant at least at the 10% level. Microfinance availability does not seem to have a significant impact on profit and consumption. Note that this exactly what has been found for the analysis in Morocco.

Regarding heterogeneity, the hypothesis that HET is zero is rejected at the 5% level for revenues and at the 1% level for the amount of loans, profits and consumption. These findings indicate that the presence of microfinance had heterogeneous impacts on business-related outcomes, which appear to manifest in a noticeable immediate influence on the quality of life, as illustrated by consumption. Note that these results differ from the analysis of Meager (2019) that found that most of the heterogeneity was sampling variation. However, in her study, the author was aggregating evidence from the various studies and did not perform a deep heterogeneity analysis with a large number of covariates, as has been done here.

Table 14: GATES of 20% Most and Least Affected Groups Across Countries

	20% Most (γ_5)	20% Least (γ_1)	Difference ($\gamma_5 - \gamma_1$)
Amount of Loans	1,423.800	-114.980	1,484.770
	(438.290; 2,371.900)	(-395.890; 263.600)	(356.050; 2,704.800)
	[0.005]***	[0.462]	[0.010]***
Revenues	230.833	-42.864	277.689
	(14.848; 429.180)	(-119.402; 37.260)	(-18.551; 543.210)
	[0.033]**	[0.280]	[0.065]*
Profit	62.174	-6.667	70.705
	(16.107; 102.668)	(-35.532; 19.392)	(17.268; 125.165)
	[0.005]***	[0.628]	[0.006]***
Consumption	18.222	-32.315	50.606
	(4.090; 33.074)	(-50.033; -16.649)	(31.861; 67.120)
	[0.014]**	[0.000]***	[0.000]***

Note. Medians over 100 splits in half. 90% confidence intervals are in parentheses and p-values in square brackets. ***, **, and * indicate significance at the 1%, 5% and 10% level, respectively.

Next, the GATES is estimated. Consistently with previous analyses, the households are divided into $K = 5$ groups based on the quintiles of the ML proxy predictor and the average effect is estimated for each group. The estimated GATES coefficient γ_1 - γ_5 along with their joint confidence bands are presented in Figures 9 - 12 in Appendix A.2. The figures reveal that there are groups of winners, the most affected groups, for which the GATES on all outcome variables are significantly different from zero. It is likely that these groups are driving the heterogeneity in treatment effect found in the BLP analysis. The most and least affected groups are compared in Table 14. The difference of GATES in these two groups is significantly different from zero at least at the 10% level for all variables. Looking at the most affected groups, there is significant evidence of positive impacts on all outcome variables, suggesting that there are indeed certain subgroups that can strongly benefit from accessing microcredit. Regarding the least affected groups, there is no evidence of significant negative impacts on the amount of loans, revenues and profit, attenuating the concerns that there are adversely affected households. However, there is negative and significant effect on consumption for this group. The exact same dynamics have been found in the analysis of the RCT in Morocco and this suggests that investment might be lumpy and that some households might cut back on consumption to increase investment as suggested by Chernozhukov et al. (2022).

After presenting evidence on the heterogeneity of treatment effects for the four outcome variables, it is interesting to examine what drives this heterogeneity in the data using CLAN. In Morocco, the results suggested that village pair fixed effects explained a significant part of the heterogeneity in treatment effects. In India, it was found that prior business experience and area population were characteristics consistently linked with heterogeneity. For the pooled analysis, village and country pair fixed effects were used as proxy for village-level and country-level characteristics. In addition, specific study variables are included as controls as well as baseline household characteristics. Therefore, if the heterogeneity appears to be driven by study specific variables or baseline household characteristics, specific study designs could be implemented in the future to target the populations most likely to benefit from microcredit.

Table 15: CLAN of Microfinance Availability Across Countries

	20% Most (δ_5)	20% Least (δ_1)	Difference ($\delta_5 - \delta_1$)
Amount of Loans			
Loan as a % of income	21.318***	14.430***	6.867***
APR	35.410***	54.170***	-20.310***
Gender is female	0.801***	0.845***	-0.045***
Household has a previous business	0.436***	0.296***	0.135***
Revenue			
Household has a previous business	0.616***	0.432***	0.189***
APR	28.640***	41.760***	-12.180***
At least primary school	0.560***	0.611***	-0.060***
Gender is female	0.784***	0.827***	-0.041***
Profit			
Loan as a % of income	18.769***	14.063***	4.663***
APR	30.060***	55.670***	-25.400***
Gender is female	0.788***	0.870***	-0.079***
In urban area	0.528***	0.429***	0.105***
Consumption			
Number of members 16 years old or older	2.088***	2.686***	-0.624***
Loan as a % of income	11.290***	14.490***	-3.000***
Number of household members	4.421***	4.921***	-0.506***
In urban area	0.395***	0.493***	-0.077***

Note. Medians over 100 splits in half. ***, **, and * indicate significance at the 1%, 5% and 10% level, respectively.

The average baseline characteristics of the most and least affected groups for the four most significant characteristics that pop up from the CLAN for all outcome variables are presented in Table 15. In terms of baseline household characteristics, gender and previous business experience often come up as characteristics of the most and least affected groups. For the amount of loans, revenues and profits, it is found that men tend to be more affected than women. While it is mostly women that are being targeted by microcredit, this result it not surprising. Indeed, Fletschner (2009) and Agier and Szafarz (2013) show that women tend to face greater constraints in growing their businesses compared to men even though they are targeted by microcredit. The authors argue that this is often due to patriarchal norms, women's limited mobility, their lack of control over resources and lower financial literacy (Fletschner, 2009; Agier & Szafarz, 2013). For the amount of loans and revenues, it is found that households with previous business experiences are more affected than other households. This results fits nicely with the previous results obtained and the literature suggesting that households or individuals with previous business experience are more likely to effectively use loans to increase revenues and profits (Banerjee et al., 2017; Meager, 2019; Chernozhukov et al., 2022; Banerjee, Karlan & Zinman, 2015).

Turning to study specific characteristics, it is found that the loan size as a percentage of income, the annual percentage rate (APR) and the implementation of the study in urban areas are strongly correlated with the most and least affected groups. In terms of the APR, it is found that for the amount of loans, revenues and profits, the most affected groups tend to be targeted by much lower interest rates. This aligns with the fundamental economics of borrowing. Indeed, Karlan and Zinman (2010) indicate that lower interest rates reduce the cost of borrowing, making it more affordable for low-income households to take loans and invest in business activities. Furthermore, lower rates can potentially increase loan repayment rates and decrease the likelihood of over-indebtedness, leading to better business performance and potentially higher consumption. The most impacted groups in terms of the amount of loans tend to have larger loan sizes as a percentage of income. Banerjee, Duflo, Goldberg et al. (2015) found that larger loans as a proportion of household income have a more significant impact on business growth and consequently poverty reduction. Larger loans can enable households to make more substantial and potentially more profitable investments. However, it is then surprising to find that the most impacted groups in terms of consumption tend to have lower loan sizes. Finally, there is a contrasting impact of microcredit across urban and rural areas. For profits, microcredit has a larger impact on urban areas. This is explained in the literature. In particular, Giné and Townsend (2004) show that urban areas typically have more vibrant and diverse economies, which can create greater opportunites for business profitability. In contrast, the impact of microcredit on consumption is higher in rural areas. Banerjee, Duflo, Goldberg et al. (2015) found that microcredit often helps smooth consumption, particularly in rural areas. Rural areas often have higher levels of vulnerability and in such context, households may use microcredit to stabilize consumption rather than investing their funds in businesses creations.

A large part of the heterogeneity is also driven by particular village-level and country-level characteristics which have not been collected across most of the studies. In order to get a precise idea of the heterogeneous impact of microcredit, combining the existing covariates with site-level covariates would be very beneficial. However, study characteristics and baseline household level covariates account for part of the heterogeneity. From this pooled analysis, it can be concluded that there is strong heterogeneity in the impacts of microcredit on the four outcome variables across the six countries. Households with previous business experience appear to be consistently more impacted by microcredit and should be targeted for future interventions. In terms of study design, when possible, loan should be larger as a percentage of income and interest rates should be lower to achieve larger impacts on business outcomes. Finally, it also seems that different targeting strategies should be used for different outcome variables. Indeed, larger impacts on consumption are found in rural areas with loans that are smaller as a percentage of income, whereas the contrary is found for business revenues and profits.

6 Conclusion

The main goal of this paper was to investigate whether microcredits are effective as a tool to alleviate poverty and whether they exhibit evidence of treatment effect heterogeneity. Specifically, using the innovative methodology developed by Chernozhukov et al. (2022) to obtain reliable estimations and interpretations of CATE features, the main research question investigated in this paper is: *Is there evidence for the presence of treatment effect heterogeneity within and across the RCT targeting poverty using microfinance? Which covariates are driving this potential heterogeneity?*. In particular, HET has been explored across four key outcome variables: the amount of loans, business profits, business revenues, and consumer spending. In addition to detecting evidence of treatment effect heterogeneity, this research has also aimed to discover whether the potential heterogeneity was driven by site-level covariates or rather by household-level covariates. Finally, it has tried to identify specific contexts that could be associated with successful microcredit intervention. To investigate this research question, the research presented in this paper focused on three different settings, each distinctly aimed at reducing poverty by enhancing access to microcredit to poor househods. The first two settings were individual RCTs in Morocco and India, respectively. Then, a pooled analysis investigating heterogeneity across six RCTs conducted in Bosnia and Herzegovina, Ethiopia, India, Mexico, Mongolia, and Morocco simulatenously was conducted to provide robust and generalizable insights about the heterogeneous impacts of microcredit access on poverty reduction.

The results obtained in this research provide mixed evidence regarding the effectiveness of enhanced access to microfinance. On the one hand, the finding that access to microfinance resulted in significant increases in the amount of loans and business revenues across the observed countries confirm previous researches that show that microfinance has the potential to increase the amount of loans and business revenues (Armendáriz & Morduch, 2010; Banerjee, Karlan & Zinman, 2015). On the other hand, the fact that microcredit did not translate into significant improvements in profit and consumption diverge from the anticipated transformative impacts of microfinance as proposed by Morduch and Haley (2002) and Yunus (2007). However, this suggest that while microfinance may boost business outcomes, it may not necessarily improve the overall financial well-being of the targeted population, as suggested by Banerjee, Karlan and Zinman (2015).

In terms of heterogeneity, the results obtained in this study counter the arguments that the benefits of microfinance are uniform across different demographics and regions as suggested by Meager (2019). The experiments in Morocco and India offer compelling case studies in this regard. While microcredit did not significantly impact most outcome variables in both countries, substantial heterogeneity in treatment effects was observed. In particular, in Morocco, heterogeneity was largely driven by unobservable village-level covariates, affecting the amount of loans, business outputs, and profits. In contrast, in India, factors like previous business experience, lower literacy rates, less populated areas, and fewer businesses largely drove the heterogeneity, with effects on business revenues, profits, and consumption. Consistent with previous studies (Angelucci et al., 2015; Augsburg et al., 2015; Chernozhukov et al., 2022), the results of the pooled analysis suggest that the impact of microfinance varies significantly across different groups. For instance, men and households with previous business experience are found to benefit more from microfinance, suggesting a need for further research into these groups' unique needs and circumstances. These findings align with the idea that these households might possess necessary entrepreneurial skills that are crucial for leveraging the benefits of microfinance (Fafchamps et al., 2014; Banerjee et al., 2017; Meager, 2019). Furthermore, this research emphasized the relevance of loan size, APR, and urban versus rural context in determining the effectiveness of microfinance. The finding that lower interest rates and larger loan sizes as a percentage of income result in higher impacts of microcredit aligns with previous studies, suggesting that the cost and amount of microcredit significantly affect their utility (Karlan & Zinman, 2010). Interestingly, the findings highlighted the differential impact of microfinance in urban and rural contexts, which supports existing studies suggesting that the effectiveness of microcredit is contingent on the broader economic environment (Giné & Townsend, 2004; Banerjee, Duflo, Goldberg et al., 2015).

Addressing the sub-questions raised in this research, the heterogeneity in treatment effects across the settings appears to be influenced both by site-level covariates and household-level covariates. For example, the study in Morocco underscored the importance of unobservable village-level covariates, confirming that site-level variables significantly contribute to heterogeneity. At the same time, household-level factors played a crucial role in India, where previous business experience, literacy rates, and local business density significantly influenced treatment outcomes. As for the second sub-question, the study was unable to single out a specific context where enhancing microcredit access has proven to be particularly beneficial for the population across all key outcome variables. Nevertheless, certain contexts were linked to more pronounced effects on certain outcome variables, such as business revenues and consumption in India.

In terms of policy implications, this research underlines the importance of a more targeted approach to microfinance. Based on the findings obtained here, it could be beneficial to target households with previous business experience, provide larger loans as a proportion of income, and set lower interest rates to enhance microcredit's impact. Additionally, distinct strategies might be more effective for urban versus rural areas. In rural areas, smaller loans could be useful to help households smooth consumption, while in urban areas, larger loans might promote business growth and profitability.

This research makes a significant contribution to the literature by revealing the complex nature of microfinance impacts, highlighting the importance of taking into account individual, household, and contextual factors when assessing the effectiveness of microcredit programs. This more nuanced understanding can help inform the design and implementation of future microcredit interventions, making them more tailored and, hence, potentially more effective. However, certain limitations of this research must be noted. Not all variables were uniformly collected across the various studies, which may affect the comparability of findings. Also, data collection was not evenly distributed across all countries, leading to potential biased. Finally, some RCTs did not collect either household level or area level covariates, potentially limiting the depth of the analysis. These issues underline the need for improved methodological rigour in future research.

To conclude, there are several promising avenues for further research. Further research into the most affected groups' needs and circumstances is of paramount importance to target them accordingly. More extensive data on village-level and country-level characteristics would allow for a more nuanced understanding of the context-dependent nature of microcredit effectiveness. In addition, collecting additional household-level characteristics that are the same across multiple countries would allow for a deeper understanding on the characteristics of the most and least affected households. Moreover, examining the impacts of microfinance within difference socio-economic and cultural contexts can shed more light on the factors that may moderate the effect of microcredit. Finally, future research could also focus on combining microfinance with other interventions such as financial literacy training and investigating the potential heterogeneity in more complex settings.

References

Abadie, A., Athey, S., Imbens, G. W. & Wooldridge, J. (2017). *When should you adjust standard errors for clustering?* (Tech. Rep.). National Bureau of Economic Research.

Agier, I. & Szafarz, A. (2013). Microfinance and gender: Is there a glass ceiling on loan size? *World development*, *42*, 165–181.

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

Armendáriz, B. & Morduch, J. (2010). *The economics of microfinance.* MIT press.

Athey, S. (2018, January). The impact of machine learning on economics. In *The economics of artificial intelligence: An agenda* (p. 507-547). University of Chicago Press. Retrieved from http://www.nber.org/chapters/c14009

Athey, S. & Imbens, G. (2016, July). Recursive partitioning for heterogeneous causal effects: Table 1. *Proceedings of the National Academy of Sciences*, *113*, 7353-7360. doi: 10.1073/pnas.1510489113

Athey, S. & Wager, S. (2019, January). Estimating treatment effects with causal forests: An application. *Observational Studies*, *5*, 37-51. doi: 10.1353/obs.2019.0001

Athey, S. & Wager, S. (2021). Policy learning with observational data. *Econometrica*, *89*(1), 133–161.

Attanasio, O., Augsburg, B., De Haas, R., Fitzsimons, E. & Harmgart, H. (2015, January). The impacts of microfinance: Evidence from joint-liability lending in mongolia. *American Economic Journal: Applied Economics*, *7*(1), 90-122. Retrieved from https://www.aeaweb.org/articles?id=10.1257/app.20130489 doi: 10.1257/app.20130489

Augsburg, B., De Haas, R., Harmgart, H. & Meghir, C. (2015, January). The impacts of microcredit: Evidence from bosnia and herzegovina. *American Economic Journal: Applied Economics*, *7*(1), 183-203. Retrieved from https://www.aeaweb.org/articles?id=10.1257/app.20130272 doi: 10.1257/app.20130272

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

Banerjee, A., Duflo, E., Glennerster, R. & Kinnan, C. (2015, January). The miracle of microfinance? evidence from a randomized evaluation. *American Economic Journal: Applied Economics*, *7*(1), 22-53. Retrieved from https://www.aeaweb.org/articles?id=10.1257/app.20130533 doi: 10.1257/app.20130533

Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parienté, W., ... Udry, C. (2015). A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science*, *348*(6236), 1260799. Retrieved from https://www.science.org/doi/abs/10.1126/science.1260799 doi: 10.1126/science.1260799

Banerjee, A., Karlan, D. & Zinman, J. (2015, January). Six randomized evaluations of microcredit: Introduction and further steps. *American Economic Journal: Applied Economics*, *7*(1), 1-21. Retrieved from https://www.aeaweb.org/articles?id=10.1257/app.20140287 doi: 10.1257/app.20140287

Bangoura, L., Khary Mbow, M., Lessoua, A. & Diaw, D. (2016). Impact of microfinance on poverty and inequality a heterogeneous panel causality analysis 1. *Revue d'économie politique*(5), 789–818.

Beck, T., Demirgüç-Kunt, A. & Levine, R. (2007). Finance, inequality and the poor. *Journal of economic growth*, *12*, 27–49.

Besley, T. & Burgess, R. (2003). Halving global poverty. *Journal of economic perspectives*, *17*(3), 3–22.

Bourguignon, F. (2004). *The poverty-growth-inequality triangle* (Tech. Rep.). working paper.

Budiono, S., Purba, J. T. & Purba, A. (2021). Fighting poverty gap through literacy and electricity: A case from indonesia.

Chernozhukov, V., Demirer, M., Duflo, E. & Fernández-Val, I. (2022). *Generic machine learning inference on heterogenous treatment effects in randomized experiments.*

Coleman, B. E. (2006). Microfinance in northeast thailand: Who benefits and how much? *World development*, *34*(9), 1612–1638.

Crépon, B., Devoto, F., Duflo, E. & Parienté, W. (2015, January). 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. Retrieved from `https://www.aeaweb.org/articles?id=10.1257/app.20130535` doi: 10.1257/app.20130535

Cull, R., Demirgu¨ç-Kunt, A. & Morduch, J. (2007). Financial performance and outreach: a global analysis of leading microbanks*. *The Economic Journal*, *117*(517), F107-F133. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0297.2007.02017.x` doi: https://doi.org/10.1111/j.1468-0297.2007.02017.x

Datt, G. & Ravallion, M. (1992). Growth and redistribution components of changes in poverty measures: A decomposition with applications to brazil and india in the 1980s. *Journal of development economics*, *38*(2), 275–295.

De Gregorio, J. & Guidotti, P. E. (1995). Financial development and economic growth. *World development*, *23*(3), 433–448.

Desmet, K. & Henderson, J. V. (2015). The geography of development within countries. In *Handbook of regional and urban economics* (Vol. 5, pp. 1457–1517). Elsevier.

Fafchamps, M., McKenzie, D., Quinn, S. & Woodruff, C. (2014). Microenterprise growth and the flypaper effect: Evidence from a randomized experiment in ghana. *Journal of development Economics*, *106*, 211–226.

Fletschner, D. (2009). Rural women's access to credit: market imperfections and intrahousehold dynamics. *World development*, *37*(3), 618–631.

Giné, X. & Townsend, R. M. (2004). Evaluation of financial liberalization: a general equilibrium model with constrained occupation choice. *Journal of development Economics*, *74*(2), 269–307.

Hariton, E. & Locascio, J. J. (2018). Randomised controlled trials – the gold standard for effectiveness research. *BJOG: An International Journal of Obstetrics & Gynaecology*, *125*(13), 1716-1716. Retrieved from `https://obgyn.onlinelibrary.wiley.com/doi/abs/10.1111/1471-0528.15199` doi: https://doi.org/10.1111/1471-0528.15199

Hastie, T., Tibshirani, R., Friedman, J. H. & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). Springer.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, *81*(396), 945–960.

Jacob, D. (2021). Cate meets ml: Conditional average treatment effect and machine learning. *Digital Finance*, *3*(2), 99–148.

Jovanovic, B. (1982). Selection and the evolution of industry. *Econometrica: Journal of the econometric society*, 649–670.

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

Khandker, S. R. (1998). Micro-credit programme evaluation: A critical review 1. *IDS bulletin*, *29*(4), 11–20.

King, R. G. & Levine, R. (1993). Finance and growth: Schumpeter might be right. *The quarterly journal of economics*, *108*(3), 717–737.

Latifee, H. I. (2003). Microcredit and poverty reduction. In *International conference on poverty reduction through microcredit. taksim-istambul.*

Meager, R. (2019, January). Understanding the average impact of microcredit expansions: A bayesian hierarchical analysis of seven randomized experiments. *American Economic Journal: Applied Economics*, *11*(1), 57-91. Retrieved from https://www.aeaweb.org/articles?id=10.1257/app.20170299 doi: 10.1257/app.20170299

Morduch, J. & Haley, B. (2002). *Analysis of the effects of microfinance on poverty reduction* (Tech. Rep.). NYU Wagner working paper New York.

Porter, M. E. (2008). The five competitive forces that shape strategy. *Harvard business review*, *86*(1), 78.

Ravallion, M. (2001). Growth, inequality and poverty: looking beyond averages. *World development*, *29*(11), 1803–1815.

Rubin, D. (1972). Estimating causal effects of treatments in experimental and observational studies. *ETS Research Bulletin Series*, *1972*(2), i-31. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/j.2333-8504.1972.tb00631.x doi: https://doi.org/10.1002/j.2333-8504.1972.tb00631.x

Sen, A. (1976). Poverty: an ordinal approach to measurement. *Econometrica: Journal of the Econometric Society*, 219–231.

Shetty, S. (2013). Microcredit, poverty, and empowerment: Exploring the connections. In *The credibility of microcredit* (pp. 163–198). Brill.

Soubbotina, T. P. & Sheram, K. (2000). *Beyond economic growth: Meeting the challenges of global development.* World Bank Publications.

Su, X., Tsai, C.-L., Wang, H., Nickerson, D. M. & Li, B. (2009). Subgroup analysis via recursive partitioning. *Journal of Machine Learning Research*, *10*(5), 141–158. Retrieved from http://jmlr.org/papers/v10/su09a.html

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

United Nations. (2015a). *Ending poverty.* United Nations. Retrieved from https://www.un.org/en/global-issues/ending-poverty

United Nations. (2015b). *Sustainable development goals: United nations development programme.* https://www.undp.org/sustainable-development-goals/no-poverty.

US Department of Labor Bureau of Labor Statistic. (2023). *Consumer price index data from 1913 to 2023.* Retrieved from

https://www.usinflationcalculator.com/inflation/current-inflation-rates/
Varadhan, R. & Seeger, J. D. (2013). Estimation and reporting of heterogeneity of treatment effects..
Welz, M., Alfons, A., Demirer, M. & Chernozhukov, V. (2022). GenericML: Generic machine learning inference [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=GenericML (R package version 0.2.3)
World Bank. (2023). *Databank, world development indicators: Ppp conversion factor, private consumption (lcu per international $).* Retrieved from https://databank.worldbank.org/source/world-development-indicators/Series/PA.NUS.PRVT.P
World Bank Group. (2016, Dec). *Does microfinance still hold promise for reaching the poor?* World Bank Group. Retrieved from https://www.worldbank.org/en/news/feature/2015/03/30/does-microfinance-still-hold-promi
Yunus, M. (1999). The grameen bank. *Scientific American*, *281*(5), 114–119.
Yunus, M. (2007). The nobel peace prize 2006-nobel lecture. *Law & Bus. Rev. Am.*, *13*, 267.

A Appendix

A.1 Additional Literature on Machine Learning and Heterogeneous Treatment Effects

One of the most common approaches in the literature to examine heterogeneous treatment effects is subgroup analysis. It involves evaluating the treatment effect for various subgroups based on a baseline or pre-treatment variable. The treatment effect is separately estimated for each level of the categorical variable used to define mutually exclusive subgroups, and a test for interaction is conducted to determine if there is a statistically significant interaction with the treatment indicator (Varadhan & Seeger, 2013). Since 2006, 40% of RCTs published in top economic journals reported at least one sugroup analysis, where the treatment effects in subgroups formed by baseline covariates was reported (Chernozhukov et al., 2022). However, it is worth mentioning that the interaction test generally has low power to detect differences in subgroup effects (Varadhan & Seeger, 2013).

Beyond subgroup analysis, the literature has seen the emergence of methods that use machine learning to handle high dimensional datasets and complex interaction effects. Machine learning, described by Athey (2018) as "a field that develops algorithms designed to be applied to data sets, with the main areas of focus being prediction (regression), classification, and clustering or grouping tasks", has been instrumental in advancing the understanding of heterogeneity in treatment effects. Combined with economic theory, it can be very helpful in the context of heterogeneity of treatment effects. For instance, Su et al. (2009) introduce an interaction tree procedure to conduct subgroup analysis which consists of using random forests of the interaction trees to extract factors that contribute to the heterogeneity of the treatment effect. Similarly, Athey and Imbens (2016) proposed a recursive partitioning approach combined with regression trees to obtain heterogeneous causal effects. They pioneered a data-driven method for identifying subpopulations with distinct treatment effects to facilitate policy application. This approach demonstrates the potential of machine learning methods in estimating treatment effects and testing hypotheses about treatment variations across different subgroups. In the same line, Athey and Wager (2019) further used this machine learning approach by introducing causal forest methods for treatment effect estimation. Their method yielded not only heterogeneity of treatment effects, but also valid asymptotic confidence intervals for these underlying effects. By overcoming the traditional limitations of random forest methods, they offered a powerful tool for estimating and testing heterogeneous treatment effects even in high dimensional settings (Athey & Wager, 2019). Finally, in this context, Jacob (2021) shows how to use machine learning methods to estimate not only the average but also a personalised treatment effect, the CATE. He presents a toolbox of methods that are specifically designed to estimate the CATE, like the causal BART and the generalised random forest. ML tools have also proven to be useful to deal with accidental imbalances in the sample.

However, in high dimensional settings, absent strong assumptions, generic ML tools may not even produce consistent estimators of the CATE, the difference in the expected potential outcomes between treated and control states conditional on covariates (Chernozhukov et al., 2022). This is where Chernozhukov et al. (2022)'s generic ML methods enter the picture. The authors showed that ML methods can be beneficial in selecting effective treatments from complex RCT designs that have many treatment combinations. Their method offer the potential to improve statistical power by pooling ineffective treatments with the control group. The authors emphasise that ML methods are useful for exploring heterogeneity of treatment effects when researchers have a large number of baseline variables and limited guidance on which variables may be relevant for forming subgroups (Chernozhukov et al., 2022).

A.2 Additional Figures and Tables

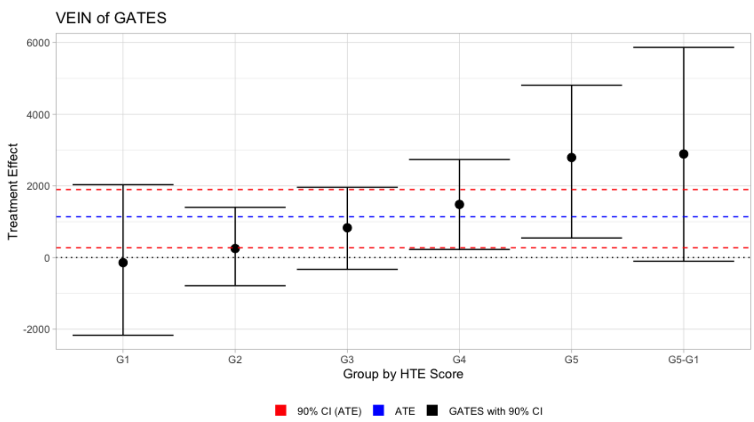

Figure 1: GATES of the amount of loans in Morocco. Point estimates and 90% adjusted confidence intervals uniform across groups based on 100 random splits in half

Algorithm 1 Inference Algorithm

Require: Data $(Y_i, D_i, Z_i, p(Z_i))$ for units $i \in [N] = \{1, \ldots, N\}$
Require: Number of splits N_S (e.g. $N_S = 100$) and significance level α (e.g. $\alpha = 0.05$)
Require: Set of ML or Causal ML methods

1. Generate N_s random splits of $[N]$ into the main sample, M, and the auxiliary sample, A. Over each split apply the following steps:
   - Using A, train each ML method and output predictions B and S for M.
   - Optionally, choose the best or aggregate ML methods.
   - Estimate the BLP parameters.
   - Estimate the GATES parameters.
   - Estimate the CLAN parameters by taking averages in M.
   - Compute the goodness of fit measures in M.
2. If the winning ML methods were not chosen in previous step, median-aggregate the goodness-of-fit measures and chooses the best ML methods.
3. Compute and report the quantile-aggregated point estimates, p-values, and confidence intervals. If previous step is used, compute and report the union of these statistics for all winners.

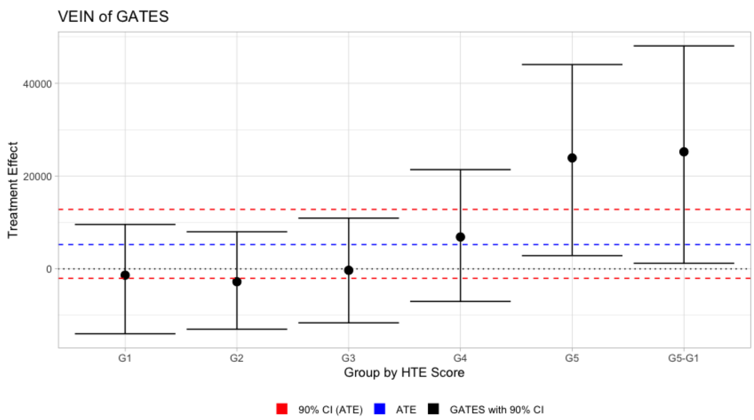

Figure 2: GATES of outputs in Morocco. Point estimates and 90% adjusted confidence intervals uniform across groups based on 100 random splits in half

Table 16: Comparison of ML Methods: Microfinance Availability in India

	Elastic Net	Boosting	SVM	Random Forest
Amount of Loans				
Best BLP(Λ)	34,826,953	133,620,484	60,222,130	230,434,533
Best GATES($\bar{\Lambda}$)	253,487,646	284,487,684	276,472,768	416,155,333
Revenue				
Best BLP(Λ)	32,235,531	59,602,131	6,972,101	76,763,850
Best GATES($\bar{\Lambda}$)	16,465,766	15,465,286	16,852,450	36,763,384
Profit				
Best BLP(Λ)	139,909	252,248	795,901	749,172
Best GATES($\bar{\Lambda}$)	1,352,166	1,042,757	1,827,493	2,579,462
Consumption				
Best BLP(Λ)	0.022	0.005	0.023	0.020
Best GATES($\bar{\Lambda}$)	0.026	0.007	0.027	0.025

Note. Results obtained when using strata-level proxy for the area baseline characteristics. Medians over 100 splits in half.

Table 17: BLP of Microfinance Availability in India

	ATE (β_1)	HET (β_2)
Amount of Loans	9,097	0.241
	(-5,663; 24,123)	(-0.008; 0.482)
	[0.241]	[0.057]*
Revenue	1,646.029	0.632
	(-864.291; 4,133)	(0.238; 1)
	[0.207]	[0.001]***
Profit	496.431	0.398
	(-343.150; 1,352.752)	(-0.542; 1.421)
	[0.268]	[0.345]
Consumption	0.002	0.760
	(-0.050; 0.054)	(0.512; 0.994)
	[0.915]	[0.000]***

Note. Results obtained when using strata-level proxy for the area baseline characteristics. Medians over 100 splits in half. Confidence intervals are in parentheses and p-values in square brackets. ***, **, and * indicate significance at the 1%, 5% and 10% level, respectively.

Table 18: GATES of 20% Most and Least Affected Groups in India

	20% Most (γ_5)	20% Least (γ_1)	Difference ($\gamma_5 - \gamma_1$)
Amount of Loans	38,192	-8,162	46,792
	(-8,238; 91,543)	(-26,396; 8,987)	(-6,183; 101,779)
	[0.120]	[0.353]	[0.091]*
Revenues	12,013.2	-4,423.5	15,871
	(2,597.1; 21,622)	(-9,673.3; 1,078)	(5,055.8; 26,969)
	[0.013]**	[0.121]	[0.003]***
Profit	3,121.32	-276.91	3,089.74
	(46.98; 6,142.7)	(-2,424.42; 1,743.2)	(-358.46; 6,831.9)
	[0.046]**	[0.801]	[0.084]*
Consumption	0.248	-0.158	0.411
	(0.135; 0.362)	(-0.249; -0.061)	(0.260; 0.559)
	[0.000]***	[0.001]***	[0.000]***

Note. Results obtained when using strata-level proxy for the area baseline characteristics. Medians over 100 splits in half. Confidence intervals are in parentheses and p-values in square brackets. ***, **, and * indicate significance at the 1%, 5% and 10% level, respectively.

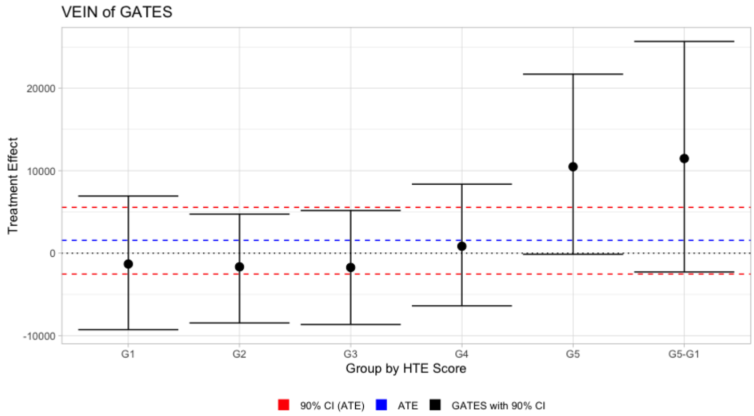

Figure 3: GATES of profits in Morocco. Point estimates and 90% adjusted confidence intervals uniform across groups based on 100 random splits in half

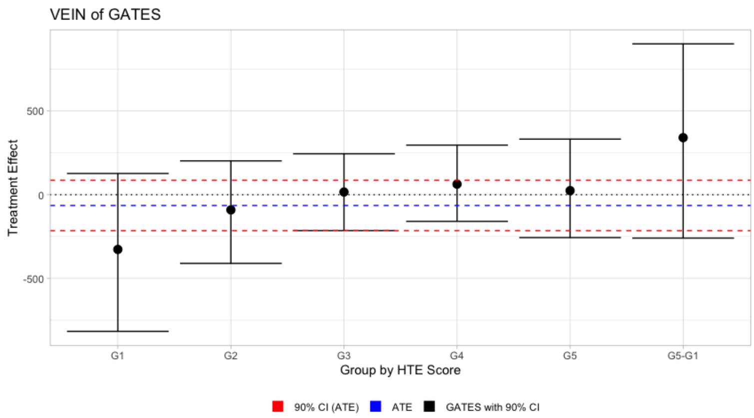

Figure 4: GATES of consumption in Morocco. Point estimates and 90% adjusted confidence intervals uniform across groups based on 100 random splits in half

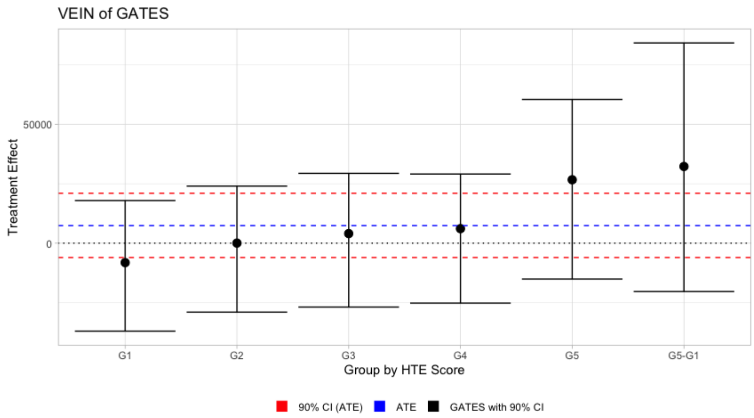

Figure 5: GATES of the amount of loans in India. Point estimates and 90% adjusted confidence intervals uniform across groups based on 100 random splits in half

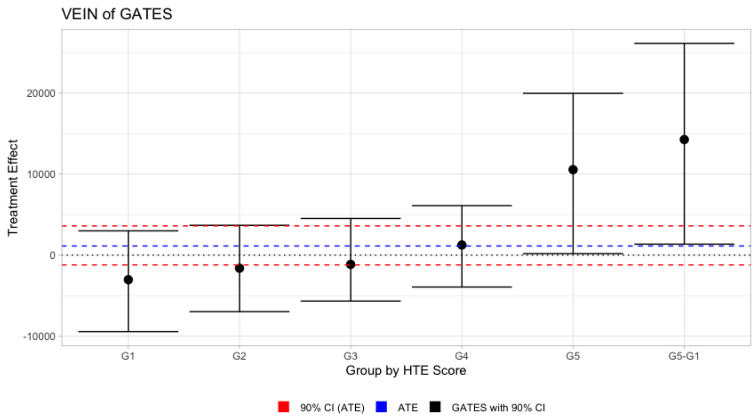

Figure 6: GATES of revenues in India. Point estimates and 90% adjusted confidence intervals uniform across groups based on 100 random splits in half

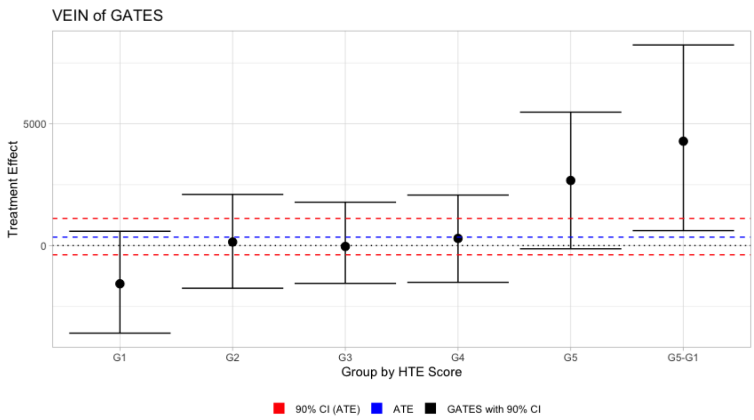

Figure 7: GATES of profits in India. Point estimates and 90% adjusted confidence intervals uniform across groups based on 100 random splits in half

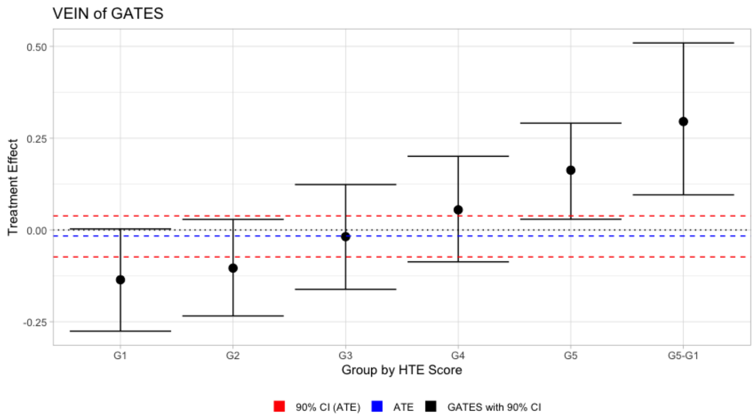

Figure 8: GATES of consumption in Iindia. Point estimates and 90% adjusted confidence intervals uniform across groups based on 100 random splits in half

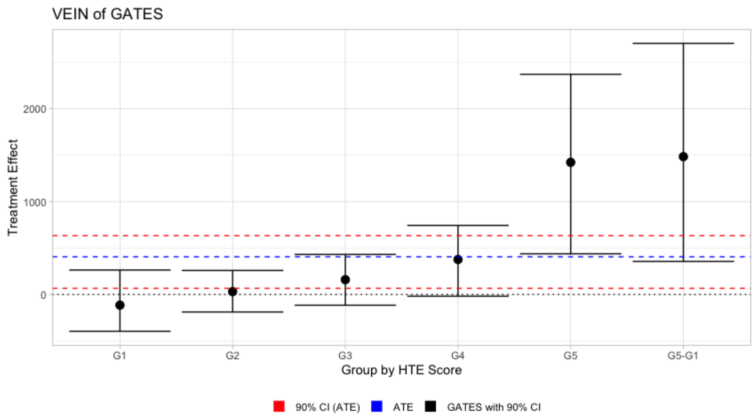

Figure 9: GATES of the amount of loans in pooled analysis. Point estimates and 90% adjusted confidence intervals uniform across groups based on 100 random splits in half

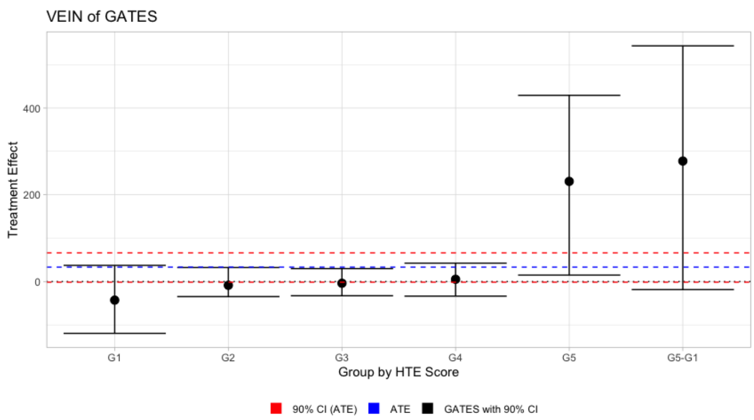

Figure 10: GATES of revenues in pooled analysis. Point estimates and 90% adjusted confidence intervals uniform across groups based on 100 random splits in half

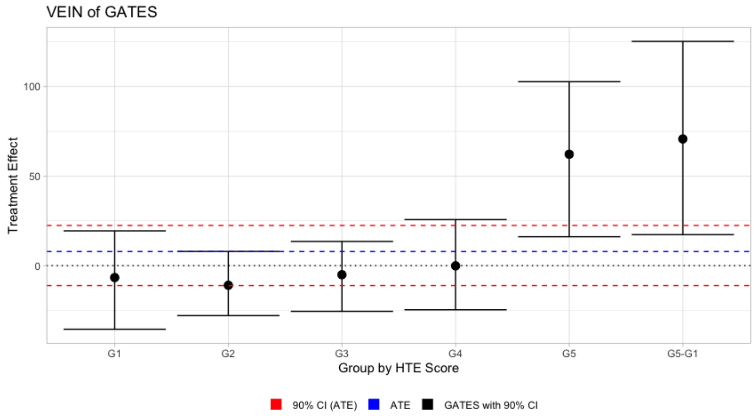

Figure 11: GATES of profits in pooled analysis. Point estimates and 90% adjusted confidence intervals uniform across groups based on 100 random splits in half

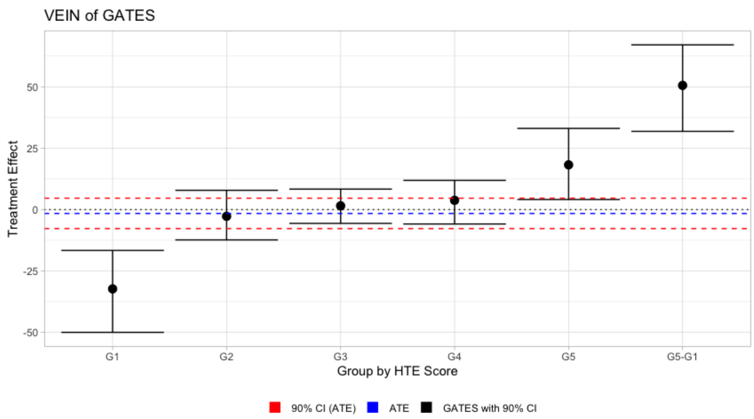

Figure 12: GATES of consumption in pooled analysis. Point estimates and 90% adjusted confidence intervals uniform across groups based on 100 random splits in half

B Programming code

The main analysis is performed using the computer software *R*. Data processing for the pooled analysis is done using *Python*. To replicate the results of the meta analyses and the pooled analysis presented in this thesis, follow these precise steps:

1. **Preparation:** Begin by downloading the ZIP file named *ThesisCode.zip*. Verify that it includes the following three folders: *code*, *data*, and *outputs*.

2. **Morocco Analysis:** To reproduce the results of the analysis in Morocco, start by executing the R script named *preprocessing_morocco.R* located in the *code/preprocessing/* directory. Then, run the script *heterogeneity_morocco.R* found in the *code/heterogeneity/* folder.

3. **India Analysis:** For the results of the analysis in India, execute the R script *preprocessing_india.R* from the *code/preprocessing/* directory, followed by running the script *heterogeneity_india.R* in the *code/heterogeneity/* folder.

4. **Pooled Analysis:** If you want to replicate the pooled analysis results, first, run the Jupyter Notebook file *pooled_preprocessing.ipynb* in the *code/python/* directory. Then, execute the R script *preprocessing_pooled.R* in the *code/preprocessing/* directory and the script *heterogeneity_pooled.R* located in the *code/heterogeneity/* folder.

5. **Outputs:** The output files for each outcome variable necessary for the tables can be found in the *outputs* folder, organized according to the three distinct settings. Figures presenting the GATES are also available in these folders.

By adhering to these instructions, you should be able to successfully reproduce the results presented in this thesis.