Microcredit and Women's Empowerment:

Examining Heterogeneous Treatment Effects and External Validity

A thesis submitted by

Vasudha Ramakrishna

in partial fulfillment of the requirements for the degree of

Master of Science
in
Economics

Tufts University

May 2020

Adviser: Cynthia Kinnan

Abstract

Microcredit has had mixed impacts on measures of women's empowerment in causal evaluations, making it hard to generalize its effects to different contexts. I attempt to model this variation by testing for within study heterogeneous treatment effects to answer questions about generalizability. I use machine-learning methods to model heterogeneous treatment effects using experimental data from India. I then use this to estimate treatment effects in another context, validating it with actual effects using experimental data from Morocco. I find that there is no heterogeneity in the treatment effect in the case of women's outcomes in India, making it less of a concern when generalizing to other contexts.

Acknowledgements

I would like to thank my advisor Professor Cynthia Kinnan, Professor Kyle Emerick and Professor Gilbert Metcalf for their guidance, helpful comments, discussions and support. I am also grateful for the support and encouragement of my close friends, parents and sister.

Contents

List of Tables

List of Figures

1 Introduction

Microcredit, which has emerged as one of the most popular tools for development assistance, prescribes small, typically short-term loans to entrepreneurs who otherwise do not have formal lending options. The evidence for the efficacy of Microcredit through a set of studies in India, Mexico, Mongolia, Morocco, Ethiopia and Bosnia and Herzegovina, finds little support for transformative effects on average, though these modest average affects could mask important heterogeneity (Banerjee, Karlan, & Zinman, 2015). Further analyses of these studies do find that access to microcredit could have large benefits for some households while others see little or no benefit, which may be hard to identify using measures of average impacts (Angelucci, Karlan, & Zinman, 2013; Banerjee, Breza, Duflo, & Kinnan, 2019; Meager, 2019). For instance the evidence from India finds that the increases in profits resulting from exposure to microcredit are concentrated in the upper tail, rendering it ineffectual to many enterprises, which offers important insights to policy makers who are concerned with reducing inequality (Banerjee, Duflo, Glennerster, & Kinnan, 2015).

Given that microcredit has largely targeted women who are often excluded from credit markets, and has been considered by policy makers to empower women by giving them greater access to resources, it is important to examine the gender dimension. There is mixed evidence for the impact of microcredit on measures of women's empowerment (measured by household decision making, number of teenagers in school, mobility, counts of female young children among others). Evidence supporting an increase in financial decision making by women and no increase in the intra-household conflict in Mexico (Angelucci, Karlan, & Zinman, 2015), contrasts the Indian experience which finds no evidence to suggest any changes in decision making or social outcomes

(Banerjee, Duflo, et al., 2015) or that from Bosnia where an increase in the labor supply of 16-19 year-olds in the household's business which could be indicative of a *fall* in women's empowerment (Augsburg, De Haas, Harmgart, & Meghir, 2015).

The mixed evidence on women's empowerment makes it important to test for heterogenous effects. Given the evidence for heterogeneity in treatment effects on other outcomes such as business profits, revenue etc (Angelucci et al., 2013; Banerjee et al., 2019; Chernozhukov, Demirer, Duflo, & Fernández-Val, 2018; Meager, 2019) it is reasonable to expect similar heterogeneity in measure's of women's empowerment. There is hence an opportunity to understand how women who differ in key baseline characteristics may respond differently to microcredit, and whether access to credit disproportionately benefits certain groups of women and not others. For instance, households with a woman owned business at baseline could experience larger effects on empowerment because it grants these women greater access to resources and is effective only when coupled with some existing agency. Or these households might face negative effects on measures of women's empowerment if female business ownership increases hours worked for women disproportionately, deepening intra-household inequalities. On the other hand, it could be that all households experience similar effects because women's bargaining is not dependent on financial independence but is governed by social and cultural norms(Jayachandran, 2019). If we were to expect that changes to business outcomes resulted in changes in women's empowerment, it is possible that given the evidence for prior business ownership having positive effects for business outcomes these households might also see positive effects for women's empowerment measures. I look to investigate this question in the context of experimental data from India and Morocco.

Results of impact evaluations also need to be considered in the context of failures of external validity, which make it hard to extrapolate and or scale up results. These could emerge due to spill over effects, changes to programme design on scaling, and implementation challenges (Banerjee et al., 2016) and or other anomalies, among other reasons. For instance, the 'No lean Season' intervention a travel subsidy to enable migration for work in the agricultural lean season in Bangladesh was found to increase incomes, consumption and welfare (Bryan, Chowdhury, & Mobarak, 2014) but when tested at scale similar results were not found. Based on experimental and programmatic data the investigators found this to be caused largely by changes to program design when it was scaled (Levy & Sri Raman, 2018).[1] Bold, Kimenyi, Mwabu, and Sandefur (2018), in the context of a teacher hiring experiment in Kenyan government schools, contrary to the positive effects found in the experiment observed that when it was implemented as a nationwide policy yielded no impact on student test-scores. This disparity was caused by implementational challenges engendered by political opposition and not by any differences in teachers who were sampled in the study.

In a randomized setting bias due to treatment selection is addressed in the design of the RCT. So, assuming the results of the RCT are internally valid we would then consider issues of generalizability due to sample selection, which may be harder to address because ensuring that the sample is fully representative of the target population is subject to a lot of challenges. Given that there could be many observed covariates and unobserved differences between the experimental sample and the actual target population, generalizability and hence external validity could come into question. In the context of extrapolat-

[1]The investigators find that this was largely caused by "mistargeting", where those who received the subsidy ended up being those who would have migrated without the subsidy anyway.

ing to larger populations, conducting tests at scale has been telling for many programmes that have been successful with smaller pilots or as demonstration studies but have failed to scale up due to spill-over effects (de Chaisemartin & Navarrete, 2019), political backlash when scaled (Bold et al., 2018), and changes to program design and implementation (Bird et al., 2019; Bryan, Mobarak, Naguib, Reimao, & Shenoy, 2018) . This exercise is important given the heavy cost to funding such programmes, the opportunity cost of not implementing alternative interventions and the human costs given that these programmes look to bring people out of poverty (Kondylis & Loeser, 2019). Apart from extrapolating to a larger population, extrapolating over time has also proven to be difficult given the heterogeneity in shocks over time (Rosenzweig & Udry, 2016).

What is critical for this study however is to see if results on the effect of microcredit on empowerment variables for women in a particular context for instance India, can accurately predict the direction and magnitude of the effect on the same measures in another context, Morocco. When comparing between contexts the true effect can vary because of heterogeneity in treatment effects. This could emerge from heterogeneity across individuals for example stemming from differences in returns to capital for borrowers and non-borrowers (Beaman, Karlan, Thuysbaert, & Udry, 2020) or differences in benefits from microcredit based on entrepreneurial ability (Banerjee et al., 2019).[2] I thus look at asking questions of external validity by modeling individual treatment heterogeneity in India and Morocco. While this has been done in the context of microcredit for household business profit, expenditures, and revenues as well as consumption patterns by Meager (2019), these questions have not

[2] It could also emerge due to implementation and study level heterogeneity that could be harder to model, but in the context of microcredit Meager (2019) finds that economic variables are better predictors of variation in effects when compared to study protocols.

been looked at in the context of social outcomes such as measures of women's empowerment. Generalizing effects on women's empowerment however is challenging to aggregate given the highly localized nature of variables measured (Meager, 2016). Meager (2019) uses a Bayesian Hierarchical framework and a meta-analytic approach to aggregate results across studies and test for external validity. Given that I have chosen to look at measures of women's empowerment which are not consistently measured if at at all across multiple studies, her methodology does not lend itself well for this analysis because my analysis is restricted to two cites. I instead use machine learning methods to identify the key features of treatment heterogeneity (Chernozhukov et al., 2018) in one context to see its predictive power in the other context.

This paper relates to several strands of literature, it first adds to the evidence on the impact of microcredit, and treatment effect heterogeneity, external validity of experimental results and finally evidence relating to credit and its effect on women's empowerment. The rest of the paper is organized as follows. Section 2 briefly reviews the microcredit literature and the strategies used to model heterogeneous treatment effects and investigate external validity. Section 3 provides details on the interventions in the two contexts India and Morocco and the data sources. Section 4 lays out the empirical strategy. Section 5 presents the results. Section 6 concludes.

2 Literature Review

In this section I first discuss the current evidence on the impact of microcredit in general, and associated literature on heterogeneous treatment effects, to motivate the choice of variables for the analyses. I then do a more specific methodological review relevant to modeling heterogeneous treatment effects and investigating external validity.

2.1 Microcredit Evidence

The key evidentiary base for microcredit comes from evaluations of Microcredit programmes that overcome the classic endogeneity emerging from lender targeting and self-selection by randomizing access to microcredit based on individual credit-scores (Augsburg et al., 2015; Karlan & Zinman, 2011) or to entire neighborhoods (Angelucci et al., 2015; Attanasio, Augsburg, De Haas, Fitzsimons, & Harmgart, 2015; Banerjee, Duflo, et al., 2015; Crépon, Devoto, Duflo, & Parienté, 2015; Tarozzi, Desai, & Johnson, 2015). Banerjee, Karlan, and Zinman (2015) review results across these studies and find the key features to be -modest take-up of microcredit, substitution among other forms of credit, modest effects on business profits, income and consumption, and mixed implications for social outcomes such as schooling and empowerment of women.[3] While social outcomes are less comparable across the six studies all of them study the impact of microcredit on child schooling finding null effects, except in the case of Bosnia where schooling for 16-19 year olds sees a significant decline and is accompanied by increased labour supply to the household business (Augsburg et al., 2015)[4]. Only four of the studies investigate effects on measures of women's empowerment and three of them find no effects (Banerjee, Karlan, & Zinman, 2015). In contrast Angelucci et al. (2015) in the context of Mexico, find improvements in women's financial decision making in the extensive margin, particularly for low power women; and an increase in the number of household issues a woman has a say in. While the study in Bosnia does not explicitly measure female decision making, and or empowerment, the

[3] The take-up rates were the least in India (18 percent), Mexico (19 percent) and Morocco (17 percent), another concern for statistical power are the low differentials in take-up rates between treatment and control groups.

[4] This effect however disappears when the authors correct for multiple hypothesis testing.

decline in schooling may be indicative of a lack of any impact on women's household bargaining, given that increases in women's influence on household decisions has been found to have positive effects on children's participation in school (Duflo, 2012). Again in the Mongolian case absent an explicit measure, the observed increase in female entrepreneurship (Attanasio et al., 2015) could be indicative of some positive change for women's empowerment. It is important to consider here that this difference in results could either be because the effects on empowerment are more driven by the specifics and modalities of the program (eg, does it explicitly promote empowerment) than by increased access to microcredit alone, or it could be due to heterogeneous effects between different groups of women and I will focus on the latter.

2.1.1 Heterogeneous impacts for Microcredit

It is reasonable to expect that microcredit could have different results for different households, because it can affect many outcomes (such as business, consumption, health, education)(Kinnan, 2020). Angelucci et al. (2013) note that the modest results on average might be concealing important heterogeneity. In order to uncover heterogeneous treatment effects and distributional effects of Microcredit in Mexico, they : (1) compare standard deviations for treatment and control groups and find evidence for a treatment effect on the standard deviations, which are however not readily explained by observable characteristics; (2) look at quantile treatment effects, finding significant treatment effects for business expenses, profits and revenues in the right tail of the distribution; (3) compare sub-groups based on several baseline characteristics such as prior business ownership, education, urban location among others, and find that prior business ownership, location, and informal savings experience drive the differential effects relative to the other characteristics.[5] Banerjee et al. (2015) find that for those with a pre-existing business before the expansion of lending by Spandana (lending agency) there is more than a doubling of profits, and this effect is driven by the upper-tail of the distribution. They also test the sustainability of these effects in a follow-up looking specifically at heterogeneity by entrepreneurial ability and find that 'gung-ho'(those with an existing business) entrepreneurs experience most of the gains from increased access to credit (Banerjee et al., 2019). Chernozhukov et al. (2018) in the context of Morocco find heterogeneity for amount of money borrowed, business output and profits, and that those who are most likely to see gains own a non-farm business at baseline. Microcredit takes many forms and is extended in many contexts, and this heterogeneity Pritchett and Sandefur (2015) argue, makes experimental evidence from one context poor predictors in other contexts. In order to obtain more generalizable conclusions on microcredit Meager (2019) uses a Bayesian hierarchical analysis to aggregate across studies and finds that most of the cross-study variation is due to sampling variation. To isolate the source of variation she investigates the role of baseline covariates and finds that having a prior business yields larger effects on average across studies for business profits, revenues and expenditure as well as household consumption and expenditure on consumer durables. She also looks at the relative role of study level characteristics and household characteristics in explaining the variation, finding the latter to better predict variation thus suggesting veritable differences in results across studies.

[5] Other variables include: income level, prior formal credit experience, prior formal bank account experience, and prior informal savings group experience.

2.2 Methodology Review

In this section I discuss some of the dominant strategies to investigate external validity, and then more specifically testing generalizability by modeling heterogeneous treatment effects. While meta-analysis is an important strategy for such questions I do not discuss it in detail given that it does not lend itself well when the studies under consideration are so few.

2.2.1 Strategies to Test External Validity

Estimating the treatment effects for a target population from empirical evidence has been visited extensively in the literature. It can simply be done using post-stratification estimation techniques, which involve reweighting treatment effects, based on the distribution of the target population. For instance, if the study sample was more educated than the target population, observations from less educated households could be up-weighted to reflect their weights in the target population.This approach may not be the ideal method when there are multiple variables to account for, unless sample sizes are large enough and there is enough variability such that there are sufficient sample units for all sub-groups intersecting between different variables (Stuart, Cole, Bradshaw, & Leaf, 2011).

In some cases, studies have used RCT results to estimate the treatment effects on the target population, by using matching and weighting techniques and existing observational data of the target population. Stuart et al. (2011) use propensity score based metrics to quantify the similarities between the participants of the RCT and the target population. The propensity scores are used to predict programme participation, and then to measure differences between the trial participants and the target population and then finally to match and re-weight the outcomes to the population. Hartman, Grieve, Ramsahai, and Sekhon (2015) in a similar fashion match participants based on observable characteristics in a RCT creating 1-1 matched pairs, they then use maximum entropy weighting to adjust the distribution in the experimental sample to that of the target population in an observational dataset, and finally compare the mean outcomes of the re-weighted sample and the target population. Matching on the likelihood of being in the study sample is useful if the experimental sample is from the same population, and the common support assumption for being in the experimental sample holds, in my case given that the two samples are from different populations, India and Morocco this assumption will not hold.

Another strategy has been to use information about the importance of observed and unobserved characteristics which may affect participation, to estimate bounds for the target population moments (Isaiah & Oster, 2018). Gechter (2015) derives bounds on the average treatment effect for the context of interest and verifies the applicability of these bounds using experimental data. The bounds act as a measure of uncertainty (emerging from unobserved differences between the two contexts) for the treatment effects in the target population. He checks the predictions against measured average treatment effects of a remedial education program implemented in two cities of India, Bombay and Vadodara assuming that the treated and untreated outcomes are dependent.

2.2.2 Modeling Heterogeneous Treatment Effects

A measure of generalizability is the variation in the effect size across different studies for a particular intervention. Hence while extrapolating across multiple contexts a key approach could be to model heterogeneity in treatment effects to explain this variation. Heterogeneity in treatment effects become especially significant if they emerge from study level characteristics (for example in the context of microcredit these would be the interest rate or the lenders targeting strategy), and could be explained away if specific study contexts are accounted for (Vivalt, 2020), these can be difficult to model however because they may be endogenous to the setting. The key strategies used in the Meta-analysis literature to model differences between studies depends on assumptions relating to the source of the cross-study variation. Fixed effect models are used if we can safely assume that the cross-study differences emerge from sampling errors, however if we suspect that contextual factors affect the effects between the two studies random effect models are used. In the Microcredit context the latter is the case (Vivalt, 2020). Meager (2019) leverages Bayesian hierarchical models to estimate the variation across studies in the context of Microcredit. Her approach essentially helps separate the true cross-study variation from the within study sampling variation. This exercise showed that the true underlying variation was amplified by the within study variation, suggesting that the findings relating to business outcomes and consumption were reasonably externally valid.

While modeling heterogeneity within a study we could use a parametric approach (appropriate for a smaller set of covariates relative to the sample size), which estimates a full set of interactions where the covariates are all indicators and partition the data (Athey & Imbens, 2017). Ideally we would want to study sub-group effects along pre-specified strata because splitting the sample across multiple dimensions could risk specification searching, machine learning tools can in such situations be leveraged (Chernozhukov et al., 2018). Athey and Imbens (2016) use regression tree methods, which split the sample along sub groups that have the most differences in treatment effects and estimate the Conditional Average Treatment Effect for each household. This strategy has been used by Beaman et al. (2020) in the context of a two-stage grants and lending programme in Mali to test whether there is heterogeneity in those who select into the loan program. Chernozhukov et al. (2018) note that in order to obtain consistent inference these methods require certain assumptions which may not always be met, and hence suggest an ML strategy that investigates features of the Conditional Average Treatment Effects instead. They test for heterogeneity in the context of microcredit using experimental data from Morocco by dividing the training data by treatment status, and applying machine-learning methods to each group separately and then project treatment effects onto control attributes. I use this approach because a metanalysis is not feasible with a small number of studies, and it makes for a tractable approach given that it employs a fewer set of less restrictive assumptions.

3 Context and Data

In this section I briefly describe the study contexts in India and Morocco, and the data which is sourced from the randomized studies done by Banerjee et al. (2015) and Crepon et al. (2015).

3.1 India

The intervention in India was rolled-out in Hyderabad, a city in the state of Andhra-Pradesh where a large lending organization- Spandana started lending in 2006.[6] The intervention involved loans for a period of 12 months with a weekly payment frequency and the loan was sized at 22 percent of income. The borrowers were women who were aged 18-59 and had to have resided in the

[6] In this period other Micro Finance institutions had also started operation in both treatment and control neighborhoods.

same area for at least one year with valid identification and residential proof. The program was randomized at the neighborhood level where 52 of the 104 neighborhoods in the sample were assigned to treatment. Two rounds of data were collected one exactly a year after the intervention which started between 2007-08 after which the control households also received access, and the second in mid 2010. The data from the first endline survey sampled 6,862 households, and in the second endline they resampled 6,142 households.[7] The take up rate was very low at 18 percent in endline 1 and 17 percent in endline 2. Table 2 reports the balance test between treatment and control groups for time-invariant baseline characteristics, and suggests that the sample was mostly composed of male household heads, who were moderately likely to have wage employment and likely to have had no education. The average household size is between 5 to 6 members, with an average of 1 child of school going ages. The households are less likely to own land and moderately likely to have a business at baseline, and any outstanding loan at baseline.[8]

3.2 Morocco

The evaluation in Morocco was implemented (in 2006) in a rural region where Al Amana (MFI) was planning to start lending in 162 villages, divided into 81 pairs of similar villages. One village in each pair was assigned to treatment. The borrowers were men and women aged 18-70 who were required to hold a national ID card, have a residency certificate, and have had an economic activity other than non-livestock agriculture for at least 12 months. The loans were for 16 months on average with weekly, bi-monthly and monthly repayment structures, and it was sized at 21 percent of income. Before the randomiza-

[7] In the original study the authors do not find differential attrition to be a concern.

[8] Treatment groups are less likely to have an outstanding loan, the difference is significant at the 5 percent level.

tion a baseline was collected with 4,465 households and two years later an endline was conducted where 4,118 households were found.[9] An additional 1,433 households were interviewed at endline to increase power making the sample size a total of 5,551 households. There was a similar low take-up rate of 17 percent.Table 3 reports the balance test between treatment and control groups for similar baseline characteristics, and suggests that the sample was mostly composed of male household heads, who were less likely to have wage employment, highly likely to have had no education. The average household size is between 5 to 6 members, with an average of 2 children of school going ages. The households are highly likely to own land and moderately likely to have a business at baseline, and any outstanding loan at baseline.[10]

4 Empirical Methods

This section describes the strategies I use to model treatment heterogeneity. For the purpose of my analysis I consider the data from India to be my trial sample, and the Morocco data to be my target or holdout sample. I hence look to model the treatment heterogeneity in the India data, and use those estimates to validate the treatment effects in the Morocco data.

4.1 Stratifying and Reweighting

As a first approach to model treatment heterogeneity I estimate the average Intent to Treat effect for different subgroups in the India data. I create these sub groups based on base-line covariates, which have been found to be con-

[9]The baseline was collected in four waves, which enabled the authors to sample those with the highest probability to become clients in the subsequent rounds. So the sample consisted of both a random sample, and high probability borrowers.

[10]Both having a baseline business and any outstanding loans are different between treatment and control groups at the 5 percent level.

sequential in past empirical work, or deemed important by economic theory (similar to Angelucci et al 2013).[11] I am unable to explore a wide variety of covariates because of considerations of ensuring comparability between variable definitions in the trial and target populations; and have to restrict my analysis to time-invariant covariates given that baseline households were not resurveyed for the follow-up rounds in the India study (final list of covariates are listed in Table 1).[12] After checking for balance for these covariates, and constructing them to be comparable between the two data sets I check that they have enough variation to ensure large enough sample sizes for the subgroups so that I have enough power to detect an effect if it exists.[13] I use a combination of fully saturated and forest type models to partition[14] the trial data, and estimate the Average ITT for each of the sub groups. I do this using the following equation:

$$Y_{in} = \alpha + \sum_{k=1}^{k} \beta_k 1(s_k = 1)T_n + \sum_{k=1}^{k} \gamma_k 1(s_k = 1) + X_n + \epsilon_{in} \tag{1}$$

Here Y_{in} is the outcome, which is an indicator of high empowerment i.e if the respondent takes an above median value on an index of women's empowerment[15] measured for each individual i in neighborhood n, T is an indicator for

[11]Consistent with Banerjee et al. (2015), and Crepon et al. (2015) I focus on the reduced form estimates here given that microcredit access is randomized at the area level and investigate comparisons between treatment and control neighborhoods. The ITT is also preferred because of concerns of potential spillovers biasing results.

[12]The baseline data collected was not representative of the neighborhoods because the authors are concerned that the sampling strategy was not followed rigorously and the baseline sample turned out to be too small, given low takeup. Hence the households interviewed at baseline are not the same as those from the follow-up surveys, which means that there is no household level baseline data available. The authors use the baseline data only to control for neighborhood level covariates.

[13]For example marital status of the household head has limited variation, with the sample predominantly consisting of married individuals.

[14]This is done by dichotomizing the variables, if they are continuous variables I use indicators of whether they fall above the median values.

[15]In India this index is measured as the weighted sum of z scores of 16 social outcomes: indicators for women making decisions on each of food, clothing, health, home purchase and repair, education, durable goods, gold and silver, investment; levels of spending on school

living in a treated neighborhood, X is a vector for strata fixed effects, and s_k is a vector of indicators for belonging to different sub-groups created by the interaction of several baseline characteristics. For example, Figure 1 describes one such iteration where the sub-groups are defined by having a baseline business, high or low education, female headed business, and land ownership splitting the data into 8 subgroups. I am interested in seeing if any of these treatment effects are significant (β_k's) or significantly different from each other, and or if the magnitudes for the subgroups are different or opposite in sign and hence might aggregate to zero on average.

After estimating the treatment effect for each of the sub-groups defined by the s_k I estimate the average treatment effect for the target population by reweighting the sub-group treatment effects based on the relative weights in and the appropriate sample weights I compare the estimate to the actual Average ITT for Morocco to see if they are comparable.

A major limitation of this approach however is that selecting the sample splits could be viewed as specification searching, and I could be detecting significant effects purely because of chance. I would also be restricting expectations of heterogeneity to linear combinations of baseline characteristics, and may miss non-linear relationships (Beaman et al., 2020). Hence, I complement this analysis with machine learning methods.

tuition, fees, and other education expenses; medical expenditure; teenage girls' and teenage boys' school enrollment; and counts of female children under one year and one to two years old. In Morocco this is measured as the sum of 14 standardized measures: at least one woman in the household has currently an own activity, decides by herself on activity assets, buys activity assets herself, decides by herself on activity inputs, buys inputs herself, decides what to produce, commercializes production, decides by herself on commercialization, makes sales herself, had an own activity in the past five years, is allowed to go to the market by herself, is allowed to take public transportation by herself, is allowed to visit family by herself, is allowed to visit friends by herself. Given that these are measured differently I use an indicator for having an above median score in the respective index for comparability.

4.2 Machine Learning Methods

I first use a simple Machine learning approach of modeling heterogeneity across 100 sample splits. After standardizing the outcome variable and baseline covariates (Variables Listed in Table 1[16]) and randomly splitting the sample of the India data set, I use an ML algorithm to predict women's empowerment in the training sample.The idea is to let the algorithm determine the most important characteristics for predicting women's empowerment. Then in the left out sample I estimate the following equation:

$$Y_{in} = \alpha + \beta_1 T_n + \beta_2 T_n * \hat{s}_i + \beta_3 \hat{s}_i + \epsilon_{in} \tag{2}$$

Here $\hat{s}_i$ is the predicted value of the outcome women's empowerment in the sample not used to train the data. I am interested in seeing if $\hat{s}_i$ (which is a function of baseline covariates) informs us of treatment heterogeneity, by looking at the coefficient on the interaction term β_2 This method allows me to obtain a tractable interpretation of people most and least affected by looking at the coefficient on the interaction term. This method is however less flexible as it assigns a restriction on the heterogeneity index to not vary for different households.

I then follow Chernozhukov et al. (2018) who suggest Machine Learning methods for heterogeneity analysis to get around the problem of overfitting engendered from sample splitting when sub-samples have not been preregistered. Based on their approach I focus on estimating features of the CATE -Conditional Average Treatment Effect (the difference in the expected potential outcomes between treated and control groups conditional on covariates) which are: Best Linear Predictor (BLP) of the CATE; second, the Sorted

[16] I also include indicators of whether the baseline values are missing.

Group Average Treatment Effects (GATES) or average treatment effect by heterogeneity groups; and third, the Classification Analysis (CLAN) or the average characteristics of the most and least affected units.

I first rescale the outcome and baseline covariates to be between 0 and 1 before training the data. And then randomly split the data into the main and auxiliary sample, which are equal in size, and in the auxiliary sample, obtain the ML estimators of the outcome for the control group and the treatment group separately. Then in the main sample I estimate the predicted CATE as the difference between the predicted values of the outcome variable for the treatment and control groups. This yields a predicted treatment effect $\hat{s_i}$ as a function of the baseline characteristics z_i. I then use equation 3 to regress the outcome on treatment interacted with heterogeneity index in the main sample (the data not used to estimate the heterogeneity score) to test if there is any treatment effect heterogeneity.

$$Y_{in} = \alpha_1 + \alpha_2 \hat{b_0} + \beta_1 \hat{s_{0i}} + \beta_2 T_n * \hat{s_{0i}} + \gamma T_n + X_n + \epsilon_{in} \text{[17]} \tag{3}$$

Here the coefficient on the interaction term (β_2) is key, where if it is found to be significant there is heterogeneity and the linear combination of zi's is a relevant predictor. The coefficient γ can be interpreted as the average treatment effect. The GATES estimate is found by dividing the households into K = 5 groups based on the quintiles of the ML proxy predictor s(z) and by estimating the average effect for each group. And the CLAN analysis involves determining which covariates explain the heterogeneity the most, using bivariate and multivariate regressions, if any heterogeneity exists. As in the estimation in Chernozhukov et al. (2018) I look at the median values for each of these features across a 100 sample splits.

[17] Here the $\hat{s_{0i}}$ is actually s(z)- E(s(z))

5 Results

5.1 Partitioning and Reweighting

The results from the various iterations of sample splitting are reported in Tables 5-7, for women's empowerment measured at both endline 1 (column 1) and 2 (column 2). In each of the estimations none of the sub-groups have significant treatment effects. These are also not significantly different from each other based on results from individual t-tests. The sub-group specific treatment effects are hence imprecisely estimated and not significantly different from zero or from each other. This could be because I am unable to use a rich set of covariates for the sample splitting exercise due to data availability concerns, engendered by the lack of household level baseline covariates for India, and considerations of sufficient variation and comparability with the Moroccan covariates. Unsurprisingly the re-weighted sub-group treatment effects do not match the average effects in Morocco. To use the data more efficiently, I next explore machine learning methods.

5.2 Machine Learning Methods

Table 8 reports results from the simple machine learning exercise, the average ITT's with a point estimate of 0.024 standard deviations and -0.04 standard deviations for endline 1 and endline 2 respectively are comparable to the ones from the original paper (Table 4, 0.013 and -0.009 for endline 1 and endline 2 standard deviations respectively), but the heterogeneity index does not vary significantly when interacted with treatment. It does have a negative coefficient which means that households having those characteristics see a small reduction in women's empowerment, but this effect is not statistically significant at conventional levels.

Table 9 presents results of the Best Linear Predictor (BLP) of CATE using the ML proxies s(z) for the women's empowerment measured at both the end lines. The table reports estimates of the Average Treatment Effect and the Heterogeneity which are measured by the coefficients β_1 and γ from equation 3 respectively; the 90 percent confidence intervals and the median p-values from the 100 sample splits. The estimated average intent to treat estimates are consistent with the results in Table 4. This means that the ML estimators are accurate predictors of the treatment effects. These estimates are similar to those found in the original paper, where I fail to detect a significant effect at conventional significance levels, and the null results are precisely estimated. The results are consistent across both the ML estimators the ridge and lasso. I fail to reject the null hypothesis that the heterogeneity is zero, for measures of women's empowerment at both endline 1 and endline 2, again the results are consistent across both the ML estimators. These results suggest that microfinance availability does not have heterogenous impacts on measures of women's empowerment.

The next analyses looks at the Group Average Treatment Effects (GATES) estimates, where I divide the households into five groups based on quintiles of the $\hat{s_{0i}}$ and estimate the average ITT for each group. Figures 2 and 3 show that there is no difference between the groups and they are all equally unaffected at endline 1; Figures 4 and 5 suggest similar results for endline 2, as evident from the overlapping confidence bands. Table 9 reports the treatment effects for the most and least affected groups, and these are not statistically different from each other.[18]

These results may appear unsurprising because the original studies found that on average, exposure to microcredit had null effects on measures of

[18]Since I do not detect any heterogeneity I do not do the classification analysis, which would tell us what was driving the heterogeneity if there was any.

women's empowerment. However given the evidence on heterogeneity for other outcomes, the same pattern could have extended to social outcomes. The results suggest that this is not the case, and it is evident that households (those with a prior business) who experienced significantly higher gains in business outcomes (Banerjee et al., 2019) did not experience any change in women's bargaining power within the household in either the short term or long term. While coefficients sometimes take negative values for different groups these are not statistically significant hence allaying concerns of negative effects on some groups (Angelucci et al., 2013). Given that the take-up rates are very low it may explain why I don't find much heterogeneity.

6 Robustness checks

I use different methods of capturing heterogeneity first using strata indicators, and then by predicted women's empowerment as a function of covariates and find null results consistently. In order to establish that these results reflect the fact that there is truly no heterogeneity to detect I need to ensure these results are robust to other specifications of the outcome variable and splits of the data. I redo the partitioning exercise by assigning households to quartiles based on women's empowerment and still find null results. In future work I could also look at two individual components of the women's empowerment index which I can obtain for Morocco: teenage girls' and teenage boys' school enrollment; and counts of female children under one year and one to two years old. Second I could estimate the machine learning approach for business outcomes and see if those are able to predict any heterogeneity in women's empowerment. Finally given that Morocco had only one MFI(Alamana) the complier population is slightly different because they would not all be always takers so I might have more power to detect heterogeneity, it would be useful to reverse the analyses by using Morocco as the trial population and India as the testing population.

7 Conclusion

Overall, the partitioning exercise showed a lack of evidence for heterogeneous treatment effects for different sub-groups of women, this could have been because of data availability concerns. The more data driven machine learning approach however confirms these findings. The algorithm accurately predicts the average intent to treat effects, but fails to find heterogeneity. Given that proponents of microcredit have suggested often strongly that microcredit empowers women, these results are important. Additionally, there is no strong evidence to suggest that microcredit hurts certain women disproportionately, this allays fears that access to credit for women could negatively affect them for example they might experience more domestic violence in retaliation (Angelucci, 2008).

These results might potentially point towards important lessons for targeting and designing microcredit interventions. If the objective is to improve women's empowerment the interventions might need to be tied with complementary workshops or campaigns specifically targeting women's outcomes and a change in social norms and or attitudes. These could include discussions about gender equality like in the study by Dhar, Jain, and Jayachandran (2018) where the authors use a school based intervention to target gender attitudes and find that the discussions actually made gender attitudes more progressive; or interventions that specifically target women's self efficacy for instance McKelway (2019) finds that a psychosocial intervention targeting women's self efficacy leads to increases in women's employment.

The absence of heterogeneity also means that microcredit interventions do not need to be designed differently to cater to different sub-groups, and a one size fits all approach might be adequate. It is important to note here that there are limitations with measuring women's empowerment due to it being highly context specific and prone to biases (Peterman & Seymour, 2018). Moreover it might take longer to experience changes in bargaining power despite experiencing positive changes in business related outcomes so lengthening the time span for analysis such as in Banerjee et al. (2019) might uncover more insights. Finally, in the absence of standardized measures for women's empowerment it is difficult to extrapolate results to different contexts.

References

Angelucci, M. (2008). Love on the rocks: Domestic violence and alcohol abuse in rural mexico. *The B.E. Journal of Economic Analysis Policy*, *8*(1).

Angelucci, M., Karlan, D., & Zinman, J. (2013, June). *Win Some Lose Some? Evidence from a Randomized Microcredit Program Placement Experiment by Compartamos Banco* (NBER Working Papers No. 19119). National Bureau of Economic Research, Inc. Retrieved from https://ideas.repec.org/p/nbr/nberwo/19119.html

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

Athey, S., & Imbens, G. (2016, July). *Recursive partitioning for heterogeneous causal effects.* https://www.pnas.org/content/113/27/7353.abstract.

Athey, S., & Imbens, G. (2017). The econometrics of randomized experiments. *Handbook of Economic and Field Experiments*, *1*.

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., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukherji, S., ... Walton, M. (2016, December). *From proof of concept to scalable policies: Challenges and solutions, with an application* (Working Paper No. 22931). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w22931 doi: 10.3386/w22931

Banerjee, A., Breza, E., Duflo, E., & Kinnan, C. (2019, October). *Can microfinance unlock a poverty trap for some entrepreneurs?* (Working Paper No. 26346). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w26346 doi: 10.3386/w26346

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., 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

Beaman, L., Karlan, D., Thuysbaert, B., & Udry, C. (2020, March). *Selection into credit markets: Evidence from agriculture in mali.*

Bird, K. A., Castleman, B. L., Denning, J. T., Goodman, J., Lamberton, C., & Rosinger, K. O. (2019, August). *Nudging at scale: Experimental evidence from fafsa completion campaigns* (NBER Working Papers No. 26158). National Bureau of Economic Research, Inc. Retrieved from https://ideas.repec.org/p/nbr/nberwo/26158.html

Bold, T., Kimenyi, M., Mwabu, G., & Sandefur, J. (2018, 12). Experimental evidence on scaling up education reforms in kenya. *Journal of Public Economics*, *168*, 1-20. doi: 10.1016/j.jpubeco.2018.08.007

Bryan, G., Chowdhury, S., & Mobarak, A. M. (2014, September). Underinvestment in a profitable technology: The case of seasonal migration in bangladesh. *Econometrica*, *82*.

Bryan, G., Mobarak, M., Naguib, K., Reimao, M. E., & Shenoy, A. (2018, May). *No lean season 2017 evaluation.* AEA RCT Registry.

Chernozhukov, V., Demirer, M., Duflo, E., & Fernández-Val, I. (2018, June). *Generic machine learning inference on heterogenous treatment effects in randomized experiments* (Working Paper No. 24678). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w24678 doi: 10.3386/w24678

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

de Chaisemartin, C., & Navarrete, N. (2019, January). *The direct and spillover effects of a nationwide socio-emotional learning program for disruptive students.*

Dhar, D., Jain, T., & Jayachandran, S. (2018, December). *Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in india* (Tech. Rep.). National Bureau of Economic Research.

Duflo, E. (2012, December). Women empowerment and economic development. *Journal of Economic Literature*, *50*(4), 1051-79. Retrieved from https://www.aeaweb.org/articles?id=10.1257/jel.50.4.1051 doi: 10.1257/jel.50.4.1051

Gechter, M. (2015). *Generalizing the results from social experiments : Theory and evidence from mexico and india .*

Hartman, E., Grieve, R., Ramsahai, R., & Sekhon, J. S. (2015, June). From sample average treatment effect to population average treatment effect on the treated: combining experimental with observational studies to estimate population treatment effects. *Journal of the Royal Statistical Society. Series A (Statistics in Society).*

Isaiah, A., & Oster, E. (2018). *Weighting fro external validity* (Tech. Rep.).

National Bureau of Economic Research.

Jayachandran, S. (2019, September). *Social norms as a barrier to women's employment in developing countries.*

Karlan, D., & Zinman, J. (2011, June). Microcredit in theory and practice: Using randomized credit scoring for impact evaluation. *Science*, *332*.

Kinnan, C. (2020). *External validity and multiple inference: The case of microredit.*

Kondylis, F., & Loeser, J. (2019, June). *External validity musings.* https://blogs.worldbank.org/impactevaluations/external-validity-musings.

Levy, K., & Sri Raman, V. (2018, November). *Why (and when) we test at scale: No lean season and the quest for impact.* https://www.evidenceaction.org/why-test-at-scale-no-lean-season/. Retrieved from https://www.evidenceaction.org/why-test-at-scale-no-lean-season/

McKelway, M. (2019, December). *Vicious and virtuous cycles:self-efficacy and employment of women in india.*

Meager, R. (2016, October). *Understanding the impact of microcredit expansions:abayesian hierarchical analysis of 7 randomized experiments.*

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

Peterman, A., & Seymour, G. (2018, November). *Opinion: Where is the standardized measure of women's empowerment?*

Pritchett, L., & Sandefur, J. (2015, May). Learning from experiments when context matters. *American Economic Review*, *105*(5), 471-75. Retrieved from https://www.aeaweb.org/articles?id=10.1257/aer.p20151016 doi: 10.1257/aer.p20151016

Rosenzweig, M., & Udry, C. (2016, July). *External validity in a stochastic world* (Working Paper No. 22449). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w22449 doi: 10.3386/w22449

Stuart, E. A., Cole, S. R., Bradshaw, C. P., & Leaf, P. J. (2011, April). The use of propensity scores to assess the generalizability of results from randomized trials. *Journal of the Royal Statistical Society. Series A: Statistics in Society"*.

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

Vivalt, E. (2020). How much can we generalize from impact evaluations. *Journal of the European Economic Association.*

Tables and Figures

Table 1: List of Variables

Variable Name	Description
Age Head	Age of the household head
Age Spouse	Age of the spouse
Age Married Head	Age at marriage of the household head
Age Married Spouse	Age at marriage Spouse
Education Head	Highest Educational attainment for Head
Education Spouse	Highest Educational attainment for Spouse
Wage Employment Head	Indicator for head having wage employment
Wage Employment Spouse	Indicator for spouse having wage employment
Head Male	Indicator for having a male household head
Married head	Indicator for household head being married
Nwomen1845	Number of women between the ages of 18-45 in the household
Nchild417 head	Number of children of the household head aged 4-17
First Child Male	Indicator for the first child of the head being male
land ownership	Indicator for owning land
Female Business	Indicator for having any female run business
Old Business	Indicator for having a previous business
Baseline Loan	Indicator for having an outstanding loan at baseline
Formal	Indicator for having a formal loan at baseline

Table 2: Baseline Characteristics and Balance by Treatment in Andhra Pradesh

Variable	(1) Control Group	(2) Treatment Group	(3) Diff	(4) No of Observations
Head and Spouse Characteristics				
Head age at marriage	21.921	21.943	0.006	5,595
	(4.523)	(4.667)	(0.184)	
Spouse age at marriage	16.530	16.494	-0.095	5,246
	(3.421)	(3.358)	(0.142)	
Head no education	0.690	0.690	0.003	6,773
	(0.463)	(0.463)	(0.017)	
Spouse no education	0.548	0.552	0.006	6,168
	(0.498)	(0.497)	(0.022)	
Head wage employment	0.531	0.546	0.013	5,998
	(0.499)	(0.498)	(0.020)	
Spouse wage employment	0.566	0.523	-0.059*	2,522
	(0.496)	(0.500)	(0.031)	
Male household head	0.896	0.906	0.011	6,785
	(0.306)	(0.292)	(0.009)	
First child is male	13.201	13.238	0.072	6,460
	(10.022)	(9.908)	(0.270)	
Household Characteristics				
Household Size	5.624	5.595	0.025	6,796
	(2.110)	(2.107)	(0.055)	
Number of school going children (4-17)	1.372	1.369	0.023	6,796
	(1.255)	(1.233)	(0.032)	
Women aged 18-45	1.277	1.254	-0.012	6,796
	(0.716)	(0.702)	(0.017)	
Land Ownership	0.242	0.245	-0.004	6,750
	(0.428)	(0.430)	(0.025)	
Baseline business	0.300	0.307	0.008	6,796
	(0.458)	(0.461)	(0.016)	
Female owned business	0.132	0.129	-0.002	6,796
	(0.338)	(0.335)	(0.007)	
Baseline loan	0.399	0.340	-0.060**	4,493
	(0.490)	(0.474)	(0.029)	
Baseline formal loan	0.079	0.078	-0.008	4,211
	(0.270)	(0.268)	(0.010)	
Observations	3,233	3,563	6,796	

The table shows mean values of baseline characteristics for households assigned to the treatment and control households. Columns 1-2 show means and standard deviations. These include controls for each of the strata and the standard errors are clustered at the neighborhood level. Asterisks denote a statistically significant difference with the control group at the 1% ***, 5% **, or 10% * levels.

Table 3: Baseline Characteristics and Balance by Treatment in Morocco

Variable	(1) Control Group	(2) Treatment Group	(3) Diff	(4) No of Observations
Head and Spouse Characteristics				
Male Household Head	0.935	0.938	0.001	4,465
	(0.246)	(0.241)	(0.006)	
Head age	48.068	49.003	1.061**	4,437
	(15.523)	(15.658)	(0.416)	
Head no education	0.677	0.660	-0.016	4,082
	(0.468)	(0.474)	(0.016)	
Head wage employment	0.043	0.040	-0.003	4,435
	(0.202)	(0.195)	(0.005)	
Spouse age	40.007	40.334	0.512	3,943
	(13.001)	(13.434)	(0.392)	
Spouse no education	0.918	0.917	0.001	3,644
	(0.275)	(0.277)	(0.008)	
Spouse wage employment	0.043	0.040	-0.003	4,435
	(0.202)	(0.195)	(0.005)	
Household Characteristics				
Household Size	5.633	5.729	0.131	4,465
	(2.678)	(2.689)	(0.087)	
Number of school going children (4-17)	2.429	2.392	-0.039	4,465
	(2.248)	(2.259)	(0.059)	
Women aged 18-45	3.427	3.470	0.033	4,465
	(2.022)	(2.065)	(0.056)	
Land Ownership	0.612	0.619	0.011	4,446
	(0.487)	(0.486)	(0.017)	
Baseline business	0.216	0.176	-0.034**	4,465
	(0.411)	(0.381)	(0.014)	
Female owned business	0.074	0.065	-0.009	4,465
	(0.262)	(0.246)	(0.009)	
Baseline loan	0.275	0.308	0.046**	4,458
	(0.446)	(0.462)	(0.019)	
Baseline formal loan	0.252	0.321	0.053	1,265
	(0.435)	(0.467)	(0.046)	
Observations	2,266	2,199	4,465	

The table shows mean values of baseline characteristics for households assigned to the treatment and control households. Columns 1-2 show means and standard deviations. These include controls for each of the strata and the standard errors are clustered at the neighborhood level. Asterisks denote a statistically significant difference with the control group at the 1% ***, 5% **, or 10% * levels.

Table 4: Average Intent to Treat Estimates for Women's Empowerment in India

	(1) Women's Empowerment 1	(2) Women's Empowerment 2
treat	0.013	-0.009
	(0.014)	(0.017)
Strata Fixed Effects	Yes	Yes
Mean in Control	-0.00	-0.01
Number of Observations	6796	6079
R squared	0.041	0.017

Column 1 reports the coefficients for Women's Empowerment measured at endline 1, column 2 reports the coefficients for Women's Empowerment measured at endline 2. Treat is defined by assignment to the treatment neighborhood and women's empowerment is a standardized index- a weighted average of z-scores of 16 social outcomes: indicators for women making decisions on each of food, clothing, health, home purchase and repair, education, durable goods, gold and silver, investment; levels of spending on school tuition, fees, and other education expenses; medical expenditure; teenage girls' and teenage boys' school enrollment; and counts of female children under one year and one to two years old. Standard errors are clustered at the neighborhood level. Asterisks denote a statistically significant difference at the 1% ***, 5% **, or 10% * levels.

Table 5: Cell-Specific Regressions in Andhra Pradesh on High women's Empowerment EL1 and EL2

	(1) Women's Empowerment 1	(2) Women's Empowerment 2
Baselinebiz,High-ed,Femalebiz	0.603***	0.422***
(bin1)	(0.055)	(0.076)
Baselinebiz,Low-ed,Femalebiz	0.710***	0.468***
(bin2)	(0.048)	(0.072)
Baselinebiz,High-ed,Nofemalebiz	0.438***	0.288***
(bin3)	(0.047)	(0.072)
Baselinebiz,Low-ed,Nofemalebiz	0.552***	0.427***
(bin4)	(0.062)	(0.082)
NoBaselinebiz,Low-ed,OwnsLand	0.512***	0.399***
(bin5)	(0.057)	(0.078)
NoBaselinebiz,High-ed,OwnsLand	0.530***	0.275***
(bin6)	(0.046)	(0.075)
NoBaselinebiz,High-ed,NoLand	0.545***	0.367***
(bin7)	(0.031)	(0.067)
NoBaselinebiz,Low-ed,NoLand	0.576***	0.367***
(bin8)	(0.037)	(0.066)
treat * bin 1	0.027	0.020
	(0.065)	(0.066)
treat * bin 2	-0.030	-0.050
	(0.068)	(0.071)
treat * bin 3	0.093	0.006
	(0.061)	(0.054)
treat * bin 4	-0.070	-0.063
	(0.069)	(0.066)
treat * bin 5	0.083	-0.083
	(0.071)	(0.062)
treat * bin 6	0.085	0.088
	(0.059)	(0.054)
treat * bin 7	0.019	0.005
	(0.034)	(0.029)
treat * bin 8	0.037	0.034
	(0.045)	(0.036)
Strata Fixed Effects	Yes	Yes
Number of Observations	3327	3327
R squared	0.489	0.527

Column 1 reports the coefficients for Women's Empowerment measured at endline 1, column 2 reports the coefficients for Women's Empowerment measured at endline 2. Treat is defined by assignment to the treatment neighborhood and women's empowerment is an indicator for having an above median value in an index of women's empowerment. Standard errors are clustered at the neighborhood level. Asterisks denote a statistically significant difference at the 1% ***, 5% **, or 10% * levels.

Table 6: Cell-Specific Regressions in Andhra Pradesh on High women's Empowerment EL1 and EL2

	(1) Women's Empowerment 1	(2) Women's Empowerment 2
Baselinebiz,Malehead,Informalloan (bin9)	0.566***	0.311***
	(0.042)	(0.036)
Baselinebiz,Male head,Noinformalloan (bin10)	0.518***	0.309***
	(0.029)	(0.027)
Baselinebiz,Female head,Informalloan (bin11)	1.044***	0.744***
	(0.056)	(0.050)
Baselinebiz,Female head,Noinformalloan (bin12)	0.999***	0.751***
	(0.056)	(0.039)
NoBaselinebiz,Male head,Noinformalloan (bin13)	0.545***	0.333***
	(0.022)	(0.024)
NoBaselinebiz,Male head,Informalloan (bin14)	0.512***	0.282***
	(0.031)	(0.028)
NoBaselinebiz,Female head,Informalloan (bin15)	1.061***	0.715***
	(0.036)	(0.049)
NoBaselinebiz,Female head,Noinformalloan (bin16)	1.019***	0.713***
	(0.039)	(0.045)
treat * bin 9	0.021	0.065
	(0.049)	(0.051)
treat * bin 10	0.058*	0.028
	(0.033)	(0.033)
treat * bin 11	0.005	-0.065
	(0.075)	(0.098)
treat * bin 12	0.018	-0.060
	(0.076)	(0.080)
treat * bin 13	0.040*	0.002
	(0.022)	(0.021)
treat * bin 14	0.063	0.011
	(0.041)	(0.032)
treat * bin 15	-0.009	0.031
	(0.057)	(0.063)
treat * bin 16	-0.032	-0.025
	(0.052)	(0.054)
Strata Fixed Effects	Yes	Yes
Number of Observations	4446	4446
R squared	0.571	0.588

Column 1 reports the coefficients for Women's Empowerment measured at endline 1, column 2 reports the coefficients for Women's Empowerment measured at endline 2. Treat is defined by assignment to the treatment neighborhood and women's empowerment is an indicator for having an above median value in an index of women's empowerment.Standard errors are clustered at the neighborhood level. Asterisks denote a statistically significant difference at the 1% ***, 5% **, or 10% * levels

Table 7: Cell-Specific Regressions in Andhra Pradesh on High women's Empowerment EL1 and EL2

	(1) Women's Empowerment 1	(2) Women's Empowerment 2
Baselinebiz,Wage emp,Children417-high (bin17)	0.629***	0.430***
	(0.058)	(0.053)
Baselinebiz, Wage emp,Children417-low (bin18)	0.658***	0.502***
	(0.065)	(0.076)
Baselinebiz,No Wage emp,Children417-high (bin19)	0.543***	0.396***
	(0.038)	(0.038)
Baselinebiz,No Wage emp,Children417-low (bin20)	0.501***	0.453***
	(0.049)	(0.049)
NoBaselinebiz,Wage emp,Children417-low (bin21)	0.563***	0.476***
	(0.037)	(0.044)
NoBaselinebiz,Wage emp,Children417-high (bin22)	0.582***	0.366***
	(0.033)	(0.041)
NoBaselinebiz,No Wage emp,Children417-high (bin23)	0.538***	0.396***
	(0.038)	(0.041)
NoBaselinebiz,No Wage emp,Children417-low (bin24)	0.549***	0.545***
	(0.048)	(0.047)
treat * bin 17	0.040	-0.001
	(0.055)	(0.059)
treat * bin 18	0.013	-0.020
	(0.086)	(0.085)
treat * bin 19	0.036	-0.014
	(0.030)	(0.032)
treat * bin 20	0.044	-0.037
	(0.053)	(0.051)
treat * bin 21	-0.019	0.010
	(0.031)	(0.035)
treat * bin 22	0.040*	0.021
	(0.023)	(0.027)
treat * bin 23	0.032	-0.016
	(0.037)	(0.031)
treat * bin 24	0.039	-0.028
	(0.047)	(0.046)
Strata Fixed Effects	Yes	Yes
Number of Observations	5998	5998
R squared	0.508	0.546

Column 1 reports the coefficients for Women's Empowerment measured at endline 1, column 2 reports the coefficients for Women's Empowerment measured at endline 2. Treat is defined by assignment to the treatment neighborhood and women's empowerment is an indicator for having an above median value in an index of women's empowerment. Asterisks denote a statistically significant difference at the 1% ***, 5% **, or 10% * levels

Table 8: Lasso Prediction of ATE and Heterogeneity

	Lasso	
	ATE	HET
Women's Empowerment (EL1)	0.028	-0.015
	(0.103)	(0.78)
Women's Empowerment (EL2)	-0.004	-.042
	(0.507)	(0.550)

Column 1 reports the Average Treatment Effect coefficient β_1 in equation 2 and Column 2 reports the coefficient β_2 in equation 2. The coefficients are Medians over a 100 sample splits. P-values for the hypothesis that the parameter is equal to zero is reported in paranthesis.

Table 9: BLP of Microfinance Availability

	Lasso		Ridge	
	Ate	Het	Ate	Het
Women's Empowerment (EL1)	0.023	0.24	0.024	0.194
	(0.010-0.033)	(0.027-0.444)	(0.010-0.034)	(-0.002-0.34)
	[0.14]	[0.196]	[0.13]	[0.242]
Women's Empowerment (EL2)	-0.004	0.047	-0.004	0.042
	(-0.022-0.013)	(-0.016-0.228)	(-0.022-0.0126)	(-0.161-0.251)
	[0.553]	[0.545]	[0.549]	[0.547]

Column 1 and 3 report the Average Treatment Effect coefficient γ in equation 3 and Column 2 and 4 report the coefficient β_2 in equation 3. The coefficients are Medians over a 100 sample splits. 90 percent confidence intervals are reported in parenthesis. P-values for the hypothesis that the parameter is equal to zero in brackets.

Table 10: 20 Percent Most and Least Affected Groups

	Lasso		Ridge	
	20% Least	20 % Most	20%Least	20%Most
Women's Empowerment (EL1)	-0.002	0.022	-0.007	0.137
	(-0.057-0.054)	(-0.028-0.058)	(-0.05-0.048)	(-0.028-0.052)
	0.424	0.345	0.415	0.444
Women's Empowerment (EL2)	-0.021	-0.024	-.0154	-.03
	(-0.082-0.048)	(-0.079-0.03)	(-0.072-0.044)	(-0.082-0.026)
	0.461	0.397	0.533	0.376

Column 1 and 2 report the heterogeneity coefficient for the least and most affected groups . The coefficients are Medians over a 100 sample splits. 90 percent confidence intervals are reported in parenthesis. P-values for the hypothesis that the parameter is equal to zero in brackets.

Figure 1: Sample Splitting

	Baseline Business=1	
	High Education=1	High Education=0
Female Business=1	Subgroup 1	Subgroup 2
Female Business=0	Subgroup 3	Subgroup 4

	Baseline Business=0	
	High Education=1	High Education=0
Land Ownership=1	Subgroup 5	Subgroup 6
Land Ownership=0	Subgroup 7	Subgroup 8

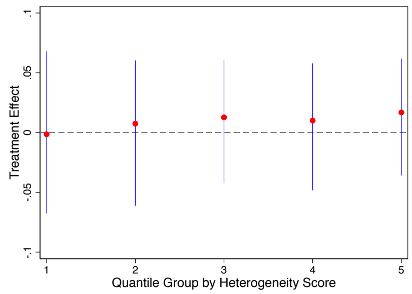

Figure 2: Women's Empowerment EL1-Lasso

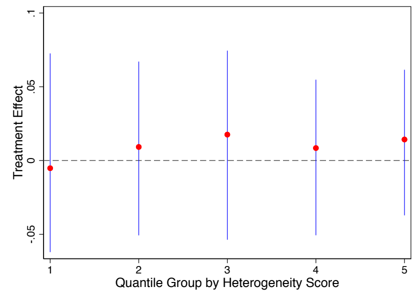

Figure 3: Women's Empowerment EL1-Ridge

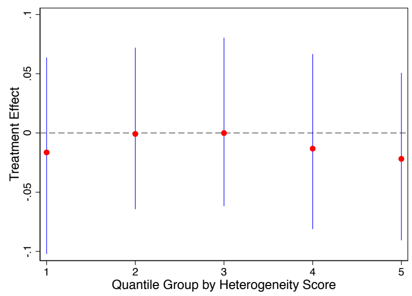

Figure 4: Women's Empowerment EL2-Lasso

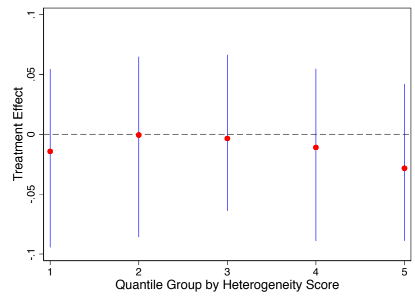

Figure 5: Women's Empowerment EL2-Ridge