UHERO
THE ECONOMIC RESEARCH ORGANIZATION
AT THE UNIVERSITY OF HAWAI'I

The Miracle of Microfinance Revisited: Evidence from Propensity Score Matching

BY

Inna Cintina
Inessa Love

Working Paper No. 2014-14
March 2014

UNIVERSITY OF HAWAI'I AT MANOA
2424 MAILE WAY, ROOM 540 • HONOLULU, HAWAI'I 96822
WWW.UHERO.HAWAII.EDU

The Miracle of Microfinance Revisited: Evidence from Propensity Score Matching

Inna Cintina and Inessa Love*

Abstract

We provide new evidence on the effectiveness of microfinance intervention for poverty alleviation. We apply the Propensity Score Matching (PSM) method to data collected in a recent randomized control trial (RCT) in India by Banerjee et al. (2014). The PSM method allows us to answer an additional set of questions not answered by the original study. First, we explore the characteristics of MFI borrowers relative to two comparison groups: those without any loans and those with other types of loans, predominantly from family and friends and money lenders. Second, we compare the impact on expenditures of MFI borrowers relative to these two comparison groups. We find that microfinance borrowers have higher expenditures in a number of categories, notably durables, house repairs, health, festivals and temptation goods. The differences are stronger relative to those without any loans. Our results suggest that microfinance can make a larger difference for households previously excluded from other credit sources. However, some of the increased expenditures are unlikely to lead to long-term benefits and there is no significant difference in total expenditures. We also present suggestive evidence of negative spillovers, i.e. non-participants reducing some categories of expenditures, while MFI participants "pick up the tab."

---

* Inna Cintina is a Research Economist with UHERO at the University of Hawaii at Manoa; Inessa Love is Associate Professor at the University of Hawaii at Manoa and Senior Economist (on leave) at the World Bank. We thank Tim Halliday and participants of the World Bank DECRG seminar series for useful comments. All errors are our own. The arguments discussed in the paper do not represent the views of the World Bank or its member countries. Contact author: Inessa Love, ilove@hawaii.edu, 808-956-7653.

1. Introduction

The impact of microfinance on poverty alleviation has become a topic of intense debate in recent academic and policy literature. Originally touted as a means for poor people to escape poverty, more recent reports suggest that the impact is likely to be small and often mixed, with such negative effects as over-indebtedness, leading to illegal organ sales and suicides in extreme cases.[1]

The key challenge of evaluating the impact of a microfinance program is to make sure any observed outcomes are due to the program itself and would not have occurred without the program. Thus, it is not sufficient to compare those with a microloan and those without it because people who obtain a microloan may be fundamentally different from those that don't. Randomized Control Trials (RCTs) have increasingly become the preferred method of evaluation for many development economists (Duflo et al., 2008). However, an important limitation of RCT in evaluating microfinance effectiveness is that researchers cannot randomly assign the recipients to receive a microfinance loan for two main reasons. First, not everyone in a random treatment group would want to obtain a loan, which will result in a selective take up. Second, the financial institution has to ensure the borrower's creditworthiness and thus cannot allocate loans randomly. Both of these problems make it difficult for an RCT to evaluate the impact of microfinance on the individual level.[2]

To avoid these problems, many recent RCTs evaluate the impact of the microfinance introduction to specific geographic areas.[3] In these studies, the microfinance is offered in some areas (villages, slums, towns), but not in others. Regardless of how many people actually take up the microfinance, the outcomes are compared across areas – i.e., a treated area is compared with a non-treated one. However, such studies can only produce the intention to treat estimates (ITT), which is the average impact of making microfinance available in an area (i.e., averaged over those who take it and whose who do not), or the Local Average Treatment Effect (LATE) if the Instrumental Variable (IV) estimator with random assignment as an instrument for take up was

[1] As reported by BBC news, October 2013. http://www.bbc.com/news/world-asia-24128096?SThisFB

[2] Occasionally, these problems can be addressed in the research design. For example, Karlan and Zinman (2010) avoid the second problem by using marginally rejected pool of applicants, but they do not address the first one – i.e. the selection to demand a loan, as they use a pool of loan applicants, who revealed their demand for loan by applying for one. Therefore, their results cannot be generalized to an average person.

[3] See Attanasio et al. (2013), Augsburg et al. (2013), Banerjee et al. (2014), Beaman et al. (2014), Angelucci, Karlan and Zinman (2013), Crepon et al. (2014), Desai, Johnson and Tarozzi (2013).

used. Neither method can produce the estimate of the impact of microfinance on the individuals or households that actually take out the loans (i.e. the Average Treatment Effect on the Treated or TOT), which is of critical policy importance.[4] This gap can be filled with the Propensity Score Matching (PSM) method and it pursued in this paper.

The PSM method creates a statistical comparison group of individuals without microfinance loans (i.e., non-participants) that has similar observable characteristics to the individuals with microfinance loans (i.e., participants). Thus, the comparison group is created to be observationally equivalent to the group of participants. While controlling on observables will reduce many of the significant differences between participants and non-participants, it cannot address the differences in unobservable characteristics, such as the entrepreneurial talent of the borrower, their time and risk preferences or their social support network. It is likely that such latent factors affect the selection of people to obtain an MFI loan and the outcomes of interest, such as poverty status. Thus, the PSM approach is not a perfect solution to identify the exact impact of microfinance on individual level. Nevertheless, multiple studies that compared performance of PSM estimators relative to experimental results have argued that PSM can produce accurate estimates under certain conditions. (see Heckman et al. 1997, 1998a, 1998b, Diaz and Handa, 2006). As we discuss below, our data satisfies all of these conditions. Moreover, the same authors argued that the bias due to unobservables is small relative to the bias due to observables. In addition, the PSM method has been used successfully to evaluate impact of different programs in a wide variety of settings (see Ravallion, 2008 for a survey). Thus, PSM appears as an appropriate method to apply in an effort to evaluate microfinance effectiveness and has an important advantage of allowing a direct comparison of borrowers to non-borrowers.

We apply the PSM method on the data collected by a recent randomized experiment by Banerjee et al. (2014). The PSM methodology we use allows us to answer an additional set of questions unanswered by the RCT. Specifically, our study addresses ***four main questions***. First,

[4] In the case of MFI loans, ITT usually measures the impact of microfinance availability (i.e., in a village) on average individual or household outcomes, such as consumption or expenditure. However, because of low take up, to increase power researchers often choose non-random sampling; they are referred to as *convenience samples*. This further clouds the interpretation of the results. For example, according to Banerjee et al. (2014) they estimate “the impact of microfinance becoming available in an area on likely clients”, which is “neither the effect on those who borrow nor the average effect on the neighborhood.” In contrast, LATE estimates the impact of credit on marginal households which take-up the credit when offered (i.e., those whose behavior is changed by the instrument, in this case increased the availability of credit). In the case of heterogeneous impacts, the LATE estimator is not equal to the parameter of interest, i.e. TOT (see Ravallion, 2008).

what are the main characteristics of microfinance borrowers? The knowledge of the borrowers' characteristics is important for microfinance program targeting, especially in light of low microfinance take-up identified in many of the recent studies (see footnote 3). Second, how do microfinance borrowers differ from those using other types of informal loans? The rich dataset allows us to compare whether characteristics of those who borrow from microfinance organizations are different from those who have no other loans vs. those who borrow from other sources (predominantly from family and friends or from money lenders).

Third, what is the impact of microfinance on consumption and expenditures of average borrowers relative to non-borrowers, and is it the same or different relative to those borrowing from other informal sources? In other words, we can identify whether the impact of microfinance depends on the comparison group we use – i.e., those without any loans vs. those with other types of loans. From the policy perspective it is important to know whether microfinance provides same or different benefits relative to other types of credit. To our knowledge such comparison has not been done in the previous literature. Importantly, by comparing the magnitudes of estimates obtained using two comparison groups, we can shed some light on one source of biases commonly associated with microfinance borrowing (e.g., entrepreneurial spirit or talent). Such biases would be more pronounced in comparison of MFI borrowers to those without loans and, as we argue below, would be reversed in comparison to those already borrowing from other sources. Thus, our estimates on two comparison groups can be thought of as straddling the biased upward estimates (relative to those without loans) and biased downward estimates (relative to those with other loans).

Fourth, are there any spillovers from MFI participants on non-borrowers? Spillovers, i.e. positive or negative impacts on non-participants, are important both for academics, since they inform the choice of evaluation methods that can be used, and for policy makers, which need to evaluate the aggregate impacts and provide targeted services. Combining the PSM method, which allows for impact evaluation on individual level, with the RCT set up, we can generate "spillover – free" estimates by comparing MFI borrowers from the treated areas with comparison groups constructed from the control areas. This allows us to evaluate whether our baseline estimates, which compare MFI participants and non-participants coming from the same treated areas contain any spillovers.

We have four main results.[5] First, we present a profile of MFI borrowers, who are more likely to be middle aged, have low education and be relatively poor (i.e., they have overcrowded living conditions, no landline, and receive government assistance). They are more likely to have prior MFI experience. They are also more likely to have had an adverse shock – like a health accident or loss of property due to fire, theft, flood or accident. The characteristics of MFI borrowers are mostly similar when compared to households without any loans vs. those with other types of loans. However, we also document a few interesting differences.[6]

Second, we find significant increase in many of the expenditure categories. In general, the comparison of impact on MFI borrowers relative to those without loans and relative to those borrowing from other sources yields many similar results: increased durable purchases, home repairs, festivals, and temptation goods. However, the magnitudes of the impact are often larger in comparison of MFI borrowers to those without any loans. While the magnitudes appear rather large, in some cases indicating an increase of up to 90%, these results should be taken with some perspective: the categories of expenditure that we find significantly increased represent a relatively minor share of total expenditures. Thus, food and non-durable expenditures, which are the largest shares of the total expenditure, show no significant changes. This explains why we don't find a significant increase in the total expenditure either.

Third, we find some suggestive evidence of negative spillovers. In other words, non-participants in the treatment areas reduce their expenditures on durables, festivals, temptation goods, and home repairs. We hypothesize that the MFI borrowers are more often "pick up the tab" for these types of expenses, which benefit other members of the community, which are able to spend less.

Fourth, our results do not support the assumption that entrepreneurial spirit is likely to bias upward the comparison of MFI borrowers and non-borrowers. Instead, the results point to the increased 'non-productive' expenditures (such as festivals, temptation goods, and home repairs). These expenditures while improving utility in the short term are unlikely to lead to any significant long term changes.

[5] Our results are robust to including neighborhood fixed effects (and seem mostly unaffected by these effects) and are largely robust to four different matching methods. We also find that of the four matching methods we use, the nearest neighbor with replacement offers the largest standard errors and is least consistent with other methods.

[6] For example, we find that those with a health-related accident are more likely to have an MFI loan relative to those without any other loans, but are not more likely to get MFI loan among those that already have other types of credit.

Finally, we compare the results from the PSM method to RCT results obtained on the same dataset by Banerjee et al. (2014). Two of our main results – the increase in durables and the lack of overall increase in total expenditures – are the same using both methods. This provides some validation for using the PSM method in evaluating microfinance effectiveness. However, PSM can also provide more nuanced evidence, such as comparing borrowers to different comparison groups of interest. Thus, whenever rich data exists, PSM can be a useful evaluation method with significantly lower costs relative to an RCT.

Some earlier papers used other non-experimental methods to estimate the impact of microfinance on actual borrowers, most notably Pitt and Khandker (1998) and Khandker (2006). However, these studies are still surrounded by controversy; see for example the re-evaluation by Roodman and Morduch (2014) and a response to re-evaluation by Pitt (2014). In light of this mixed evidence, our paper serves as an important addition to a scant non-experimental microfinance evaluation literature.

To the best of our knowledge only Floro and Swain (2012) have previously used the PSM method to study the impact of microcredit on individual level.[7] Their paper is also set in India, but they study the impact of bank-connected Self Help Groups (SHG) rather than loans from a specialized microfinance institution. In addition to this, our study has several other important differences. First, we compare MFI borrowers to two distinct groups of controls: those without any loans and those with loans from other sources, such as family and friends and money lenders. Second, we have a larger and richer dataset, which allows us to use a larger number of control variables in PSM estimation. This increases the quality of matches and minimizes bias.[8] Third, Floro and Swain (2012) focus primarily on an indicator of vulnerability, which they measure as the variance of consumption, and average food expenditures. We have a much wider set of outcomes, including purchases of durables, education, health expenditures, home repairs, and other expenditure categories, which allows for broader focus. Fourth, we use the data that has been used in an RCT evaluation, which allows us to compare the performance of these two

[7] Crepon et al. (2014) also estimate a model of the propensity to borrow. They use this model to increase the power of their randomized design by sampling households with a high propensity to borrow and to evaluate the existence of spillovers. They do not use PSM method to match participants to non-participants as we do here.

[8] Floro and Swain (2012)'s control sample includes only 51 observations for non-participants, relative to nearly 700 participants. This implies that the same control observation has to be matched to nearly 14 treatment observations, on average. In our sample the treatment and control groups are much more balanced, which suggests likely smaller bias and variance.

methods and evaluate the presence of spillovers. Thus, our paper is a significant extension of Floro and Swain (2012).

The rest of the paper is organized as follows. Section 2 discusses PSM methodology; Section 3 describes our data; Section 4 presents our results; Section 5 contains a discussion and caveats; and Section 6 concludes.

2. Methodology

PSM constructs a statistical comparison group that is based on a model of the probability of participating in the program conditional on a set of observed characteristics, X. Ravallion (2003) refers to PSM as the "observational analog" to an experiment: "just like an experiment, PSM equalizes the probability of participation across the population — the difference is that with PSM it is the conditional probability, conditional on the X variables."

Suppose T is a binary variable indicating whether an individual has participated in the program (e.g., obtained an MFI loan). The propensity score is given by the probability of participating in the program given observed characteristics: $P(X) = Pr(T = 1|X)$. It is assumed that X are unaffected by the individual's participation in the program.[9]

An important assumption for validity of PSM is *conditional independence*, which states that given a set of observable covariates X that are not affected by treatment, potential outcomes Y are independent of treatment assignment T.[10] This condition, which is also referred to as *conditional exogeneity of placement*, allows the unobserved counterfactual (i.e., the outcome of treated if they had not participated in the treatment) to be replaced by the observed outcome of the control group (Ravalion, 2008). In other words, this condition implies that the uptake of the program is based entirely on observable characteristics, and hence the differences in outcomes between treated and controls can be attributed to the treatment. While this assumption is inherently untestable, it can be more credibly invoked if there are rich observable data on control

---

[9] Ideally, the X variables are observed pre-program. Unfortunately, we do not have pre-program data. Therefore, we are careful in selecting control variables that are unlikely to be affected by the program, as we discuss below.

[10] The second identifying assumption is the presence of the common support, which can be tested and conditioned on. In essence, this means that treatment units have to be similar to control units in terms of observed characteristics. Formally, $P\ (T_i = 1|X_i) < 1$ is the weakest required assumption for estimation of the treatment effect on the treated. Rosenbaum and Rubin (1983) show that, under the two main assumptions – (1) conditional independence and (2) presence of a common support – matching on $P(X)$ is as good as matching on X.

variables (i.e., the X vector) that would allow one to control for as many of the relevant characteristics that can affect program participation, and the institutional setting in which the program takes place is well understood (Caliendo and Kopeinig, 2008).

The important question is how well PSM method performs relative to experimental methods. Fortunately, a number of studies have established that PSM can provide fairly accurate estimates under certain conditions. Heckman et al. (1997, 1998a, 1998b) analyze performance of various matching schemes relative to experimental estimators.[11] They find that propensity score matching performs well if three conditions are met: 1) using a rich set of control variables, 2) using the same survey instrument for treated and controls, and 3) comparing participants and non-participants from the same local market.

Our data satisfies all three of these conditions. First, we have a very rich set of control variables. As we describe below, the data come from a detailed household surveys and provide ample individual and household characteristics which we use to control for observable factors affecting participation in microfinance. Specifically, we use characteristics of eligible female (i.e., aged 18-59) including her age, education, and nature of employment. In addition, we have details for the male head of household, details on the household composition, and details on the dwelling. Second, the same survey instrument was used for participants and control group. Third, participants and control group come from the same local markets. To further satisfy this requirement, we only use slums in which microfinance was introduced and compare users to non-users. Thus, we believe that our rich data and setting provide solid justification for using PSM method.

While PSM cannot control for unobservable characteristics affecting program participation, Heckman et al. (1997, 1998a, 1998b) argue that the bias coming from unobservable characteristics is small, relative to the bias coming from the incorrect use of observable characteristics (i.e., comparing units outside of the common support). More recently a meta-study by Glazerman et al. (2003) also finds that bias of non-experimental estimates was lower when the comparison group was drawn from within the evaluation itself rather than from a national dataset and locally matched to the treatment population. Another relevant study by Diaz and Handa (2006) evaluate performance of PSM relative to randomized experiment using

[11] Dehejia and Wahba (2002) and Smith and Todd (2005) also compare experimental and non-experimental evaluations of impact of training program on labor market outcomes in the US.

Mexican conditional transfer program PROGRESSA. They find that in cases when the outcomes are measured using comparable surveys, the bias arising from PSM is negligible. To summarize, even though PSM cannot eliminate the bias arising from unobservable characteristics, previous research indicates that this bias is likely to be small.

The PSM method has successfully been used in many different settings. For example, Jalan and Ravallion (2003a) use PSM to study the gains in child health from access to piped water in rural India, Jalan and Ravallion (2003b) study impact of the workfare program in Argentina, Godtland et al. (2004) study the impact of agricultural extension program in Peru, Chen, Mu, and Ravallion (2009) study the effects of the World Bank–financed Southwest China Poverty Reduction Project.

One of the advantages of the PSM is its semi-parametric nature, which imposing fewer constraints on the functional form of the treatment model (i.e., it does not have to be linear), as well as fewer assumptions about the distribution of the error term relative to the regression based models. The main estimator capturing the Average Treatment effect on the Treated (TOT) can be simply written as the difference in the average outcomes for the treated and weighted average outcomes for the comparison group:

$$TOT = \frac{1}{N}\left(\sum_{i \in T} Y_i^T - \sum_{j \in C} w_{ij} Y_i^T\right)$$

where, Y is an outcome of interest (such as consumption or expenditure), N is the number of participants i and w_{ij} is the weight used to aggregate outcomes for the matched nonparticipants j. The weight depends on the matching method used.

We report results with four matching methods: (1) nearest neighbor matching with replacement; (2) nearest neighbor matching without replacement; (3) Kernel matching; (4) Stratification matching. The nearest neighbor matching is the most popular matching method and involves finding a closest match in terms of propensity score for each treated observation. We used one to one matching with or without replacement.[12] Stratification matching partitions the common support into different strata and calculates the program's impact within each interval. Kernel matching uses a weighted average of all comparison units, with weight determined by how far each unit is in terms of the propensity score. Using more information to create a match

[12] Allowing for replacement involves a trade-off between bias and variance. Replacement allows each treated unit to find a closer match, thus making treated more comparable to untreated units and reducing the bias. However, using the same comparison unit multiple times increases the variance.

reduces the variance of the estimator. The four matching methods we report use progressively more information to create a match (i.e., from one to one match in nearest neighbor, to strata match that uses all units within each strata, to kernel match that uses all available comparison units). Asymptotically, different matching methods should yield similar results, while in small samples the choice of matching could be important (Heckman et al. 1997). We present four different matching methods to test the robustness of our results. The standard errors for all methods are calculated using bootstrap simulation with 100 repetitions, which takes into account the fact that propensity scores are estimated.

3. Data

Our data come from the randomized experiment of Banerjee et al. (2014) and are described in more detail in their paper.[13] Here we provide only a brief description. In 2005, 52 of 104 poor neighborhoods in Hyderabad were randomly selected for the opening of a microfinance institution Spandana, which used the canonical group lending model and targeted women who may not necessarily be entrepreneurs. Spandana also targeted the "poor, but not the poorest of the poor" (Banerjee et al. 2014). Our data comes from the first wave of the household surveys, conducted about 15-18 month after Spandana openings. For our main analysis we use data from the 52 neighborhoods in which Spandana was opened to make sure our participants and non-participants come from the same local markets (which has been noted to improve PSM performance, as discussed in the previous section). For robustness checks and analysis of spillovers, we also generate an additional control group using data from locations without Spandana (i.e. the control group used in the RCT design). Since the microfinance program was targeted to females in the range of 18-59 years old, the data was collected only on households that have at least one eligible woman in the household.

We have a total of 3,318 households in the main sample. We construct our outcome variables following Banerjee et al. (2014) as monthly adult-equivalent expenditures, adjusted for inflation. The details on variable construction are given in Appendix A1. Because the distributions of expenditures in rupees have significantly long right tails (Figure 1), we use log

---

[13] We thank Esther Duflo for making the data generously available on her website even prior to their paper publication.

transformation on all expenditure variables. To ensure observations with zero expenditures are not dropped, we add one to all zero values before taking logs. Table 1, Panel A reports summary statistics for the outcome variables. Specifically, we have data on total consumer expenditures, total non-durable expenditures, total durable expenditures, "temptation goods" (defined as meals outside of home, alcohol and gambling), health and education (total education expenditures and education fees), expenditures on festivals, home repairs (the questionnaire only asked to report home repairs above 500 Rs).[14]

Table 1, Panel B reports shares of expenditure categories. We first calculate shares of each category, as % of total, for each household, and then report mean, median, etc. across all households. The average durable expenditures are only about 6%, while non-durable are 94% (they should add to 100%, as they do). Food is the largest category of non-durable expenditures, at an average of about 39% of total. Health expenditures and temptation goods are, in contrast, fairly small categories (5-8% on average, with even lower medians).

We have constructed a number of control variables to use in propensity score estimation. Our selection of controls is guided by the condition that they are unaffected by the MFI participation. Since Spandana was targeting women in the 18-59 years old range, we select the oldest eligible woman in the household and include her characteristics, such as age, education, and several employment characteristics. The woman's age, education, and whether she is a head of household are clearly not affected by the MFI borrowing. While it is possible that MFI participation will affect some employment choices of female borrowers or their spouses (as also argued by Banerjee et al. 2014), we only include variables that are unlikely to be affected. Specifically, we include an indicator whether the female has a stable employment (defined as the same place of employment for at least 4 years). This variable cannot be affected by any recent MFI borrowing because by definition it occurred prior to Spandana entering the area. In contrast, self-employment can plausibly be affected by MFI. Similarly, we cannot include indicators for female or male's work, since these can be affected by MFI borrowings if one or both of them start their own business. We also include male education using either the head of the household (if male) or the oldest male permanently residing in the household.

[14] In our estimation we use variables defined closely to the ones used by Banerjee et al. (2014), but we have also do robustness tests for some alternatively defined variables. Construction of variables is discussed in Appendix A1.

Household characteristics include the presence of dependents (defined as children under 13), the presence of young children (i.e., aged 0-2), and the number of eligible women in the households. The households with more eligible women are more likely to be Spandana borrowers. We include a dummy if there is only one eligible woman and a dummy for two eligible women, while the omitted category is three or more. While we do not observe the race or caste, the questionnaire includes question on the language of instruction of the children who are in school and we include a dummy for whether any of the children are being educated in English. Previous evidence suggests that children from higher caste families are more likely to be educated in English (Munshi and Rosenzweigh, 2006).

The dwelling characteristics include indicators for whether 50% or more rooms in the house are classified as "pucca" (i.e., solid and permanent, as opposed to housing build of temporary materials), whether there is a landline, whether latrine is not shared with other households, and the source of water (tap and located within premises).[15] In addition, we have an indicator of whether the owner has a title to their house, which is an important source of collateral that can be used for formal borrowing, and whether they own a large plot of land (i.e. 1 acre or more) in the village or in the city. These variables are unlikely to be affected by small MFI loans. In contrast, the type of roof, for example, might be affected if the MFI loan has been used for small house improvement. Importantly, the survey contains question on whether a previous MFI loan has been repaid and the year when the household first borrow from an MFI. We create an indicator of whether the household borrowed and repaid an MFI loan prior to 2006, i.e., before Spandana operations. This captures prior familiarity with microfinance products, and cannot be influenced by current MFI borrowing, since by definition, the loan has been repaid prior to 2006. We also include indicators for whether the household receives government assistance, which could proxy for the poverty status or a disability. Importantly, we include several indicators of adverse shocks faced by the households in the prior year: whether there was a health-related accident, whether any of the household members have lost a job due to layoffs or company closure, and whether there was a lost property due to fire, theft or an accident. These random adverse events proxy for the demand for funds and are good determinants of borrowing, but they clearly cannot be affected by MFI borrowing.

[15] We also have data on electricity, but close to 95% of households in our sample have electricity, so we do not include this variable. In our trial runs it was not statistically significant.

Finally, to save on degrees of freedom and not drop households with missing data (i.e., those that answered “I don’t know” or refused to answer), we replace these observations with zero values and add dummies to capture the average impact of those with the missing characteristics. Specifically, we include dummies for missing education (for males and females), missing years of employment, missing language of instruction and missing distance to water source. We do not report these dummies in regressions since their interpretation is unclear. Table 1, Panel C presents summary statistics on our control variables.

Table 2, Panel A provides information on borrowing patterns. Surprisingly, only about 12% of all households in our sample report that they have no loans. Multiple loans are much more prevalent than single loans: only 19% report having only one loan, while nearly 70% of households have more than 2 loans, and out of these more than 20% of households have 5 or more different loans outstanding at the time of the survey.

Table 2, Panel B provides the distribution of the sources of these loans. Note that in this table the household will be counted multiple times if they have more than one loan. Thus, out of total households about 20% have a loan from Spandana, and 13% have a loan from another MFI. In total, we have 687 Spandana borrowers and 435 other MFI borrowers (178 of them are in both groups). Thus, total there are 944 borrowers from either Spandana or another MFI. This is a large group of people, nearly 30% of the sample. The largest category of other types of creditors are money lenders (37% households have money lenders loans), followed by family and friends (at 33.6%). Shopkeepers and chit funds are about 17% each. Interestingly, 18% of these households have a commercial bank or other financial company loan. Thus, from this table it appears that even before Spandana entered these areas, these households hardly suffered from lack of credit availability. Of course, the cost and terms of credit is another story.

Table 2, Panel C reports loan combinations that are of specific interest to our study. Those with Spandana loans or other MFI loans also have loans from all other types of borrowers. However, there is some weak evidence of substitution (as also documented by Banerjee et al. 2014). For example, among those who borrow from Spandana or another MFI family and friends loans also exist in about 30-31% of cases, while in the overall sample among those who borrow it is 38%. Similarly, money lenders represent a slightly smaller share among those with Spandana (43%) or another MFI (46%) versus general population of those with any loan

(48%).[16] Other common loan types, such as shopkeeper, chit fund, and commercial bank or finance company occur about at the same frequency among those with Spandana or other MFI loans, as among general population of borrowers.

4. Results

4.A. Estimation of the Propensity to Borrow

In this section we answer our first set of questions, such as what are the characteristics of MFI borrowers and how MFI borrowers are different from those borrowing from other informal sources. The estimation proceeds in several stages. First, we estimate the propensity to borrow model, which can be written as follows

$$T_i = \alpha + \beta X_i + \varepsilon_i \quad (1)$$

where T is a binary variable equal to one for a treatment group and zero for a comparison group, X is a vector of household characteristics and ε_i is an error term. As we discussed above, the variables in X vector are those we believe unlikely to be affected by the MFI borrowing.

We investigate three treatment/comparison group combinations. First, we compare Spandana borrowers to those without any loans. This treatment/comparison group combination is suggested by the set-up of the randomized experiment, which consisted of introduction of Spandana institution to the areas we study. Second, we combine Spandana borrowers and those who borrowed from another MFI – we refer to this as “Spandana/other MFI” or “any MFI” group - and compare them to those without any loans. Since our main interest is describing characteristics of microfinance borrowers and the impact of microfinance in general (rather than the impact of Spandana specifically), this combination is best suited to answer our main questions. Furthermore, there is no reason to believe that Spandana was different from other MFI operating in the area, and, even if it was, the difference is of little academic or policy relevance.[17] Third, we compare the Spandana/other MFI group to a comparison group of households with other types of loans. As Table 2 demonstrates, the most common other types of loans are

---

[16] In reality, the differences are more pronounced since the overall sample includes those who borrow from Spandana or other MFI.

[17] We have investigated the differences between Spandana and other MFI borrowers and found no differences; however, our sample of other MFI borrowers is rather small.

obtained from family and friends and moneylenders, followed by shopkeepers, chit funds and formal financial institutions. This treatment/comparison combination allows us to test whether those who borrow from MFIs are different from those that borrow from other sources and whether there are any differences in impact of microfinance loans vs. loans obtained from other sources. Our analysis in this section includes only 52 "treated" neighborhoods in which Spandana was randomly placed.

Figure 2 presents the density of propensity score estimated for the 3 models (with and without slum fixed effects). These graphs demonstrate that our model performs well in separating treatment and control groups, as the maximum density of propensity scores for the treatment group is always significantly higher than the maximum density of the control group. The separation is clearer for the first two models (i.e., comparing no loan to Spandana and Spandana/other MFI) than for the third one . This is not surprising, since these types of credit are likely to be close substitutes and our model is less able to discriminate between the two types of borrowers (MFI vs. other informal borrowing). In models with slum fixed effects the separation of two densities is even clearer for all three models. These graphs present a useful check on the ability of our model to predict the likelihood of using Spandana or other MFI vs. our two comparison groups (i.e. no loan and other loan).

Figure 2 also shows that there is sufficient common support (i.e., the area of overlap between two densities). The common support ensures that treatment observations have comparison observations "nearby" in the propensity score distribution. It is especially important that all treatment observations can be matched with comparison observations (i.e., no treatment observations are dropped due to lack of comparison units), and the graphs show that this is indeed the case.[18] We have also performed balancing tests to ensure that the treatment and comparison groups are balanced, meaning that similar propensity scores are based on similar observed X.[19] All the variables in our final model satisfy the balancing property.

Table 3 reports the results of propensity to borrow regressions. In this section, we focus on regressions with control households residing in the treatment slums and later we discuss

[18] There is no bias if some comparison units that fall outside of the common support are dropped; but if treatment observations are dropped due to lack of common support, this may create a bias because they might be systematically different and hence have not found comparable matching units. Fortunately, this is not the case in our situation since the range covered by the distribution of propensity scores of the treatment group completely overlaps with the range of the comparison group.

[19] Formally, balancing implies that $P(X | T = 1) = P(X | T = 0)$.

models where control households reside in control slums (i.e., columns 3, 6, and 9). Columns 1-2 report results of the selection model comparing Spandana borrowers to those without any loans, columns 4-5 compare those who borrow from any MFI (either Spandana or other MFI) to those without any loans, and columns 7-8 compare those who borrow from any MFI to those with other types of loans. We report results without and with slum fixed effects that control for neighborhood conditions (columns 1, 4, and 7 and columns 2, 5, and 8 respectively). Overall, the results with and without fixed effects are very similar, suggesting that the neighborhoods in our sample are fairly similar to each other.

Comparing the results across specifications, we find that results for Spandana vs. no loan (columns 1-2) and any MFI vs. no loan (columns 4-5) produce very similar result. This is not surprising since there is no reason to expect that Spandana would attract different types of borrowers relative to other MFIs. However, there are some minor differences in models that compare any MFI (columns 4-5) to other types of loans (columns 7-8) that we discuss below.

We find that MFI borrowers are more likely to be middle age (since the results on age and age squared show an inverse U-shape relationship) both relative to those without any loans and relative to those with other types of loans. The MFI borrowers are more likely to have low education for both male and female (the omitted category is low education).[20] We also find that females with stable employment are less likely to obtain MFI loans (although this variable is consistently negative in all models, it is only significant in models 1 and 4).

In terms of households' characteristics, we find that larger households are slightly more likely to borrow (the omitted category is seven or more people), while, surprisingly, the number of qualifying females is not statistically significantly associated with MFI borrowing. The number of young children and the presence of dependents are not significant. The dummy for English language is consistently positive, but only significant in models with slum fixed effect (5 and 8). This variable might proxy for the higher castes or more connected households.

We have several indicators for accidents and adverse shocks occurring in the past year. These are excellent predictors of MFI participation, as these random events are highly unlikely to be caused by MFI participation. We find that those that had a health related accident are more

[20] There are some differences in education impact for two comparison groups. For example, female education is a significant predictor of borrowing from MFI vs. no loan, but not significant in regressions comparing MFI to other types. Male education results show that middle and high education categories are less likely to borrow from MFI vs. no loan, while only males in the high education category is less likely to borrow from MFI vs. other types of loans. Despite these differences, the overall picture is that MFI borrowers have relatively low education.

likely to borrow from Spandana or another MFI relative to those without loans (columns 1, 2, 4 and 5), but are equally likely to borrow from other sources (columns 7-8). Similarly, we find that households that experienced lost property due to fire, theft, flood or accident are more likely to borrow from Spandana or other MFI, but not more likely to borrow from other sources. Both results suggest that when a health related accident or lost property happens, a household without a loan is likely to turn to an MFI, but if they already use other sources of credit they turn to these sources with equal probability. However, households that have a member who recently lost a job (due to layoff or company closure) are not more likely to turn to MFI vs. no loan or other loans.

We also find that households that receive government assistance are more likely to borrow from an MFI. This indicator is strongly significant in all specification, and is likely indicate more poverty or disability. However, the title on the house and ownership of large plots of land are not significant predictors of MFI borrowing.[21] Importantly, we find that an indicator of whether the household has repaid an MFI loan prior to 2006 is a strong indicator of Spandana or MFI borrowings. This suggests that those with prior familiarity with microfinance are more likely to borrow from this source again.

In terms of dwelling characteristics, we find that MFI borrowers have more overcrowded living conditions (i.e., they are more likely to have more than two persons per room). This variable is highly significant in all specifications (but only marginally significant in model 8). MFI borrowers are less likely to have a landline (this is significant only in columns 1-4) and marginally significant in column 5), when comparing to those without any loans. The quality of the dwelling is generally not significant (i.e., whether the structure is predominantly "pucca" type, whether a latrine is shared, and the water source within premises.)[22] MFI borrowers are more likely to not share a latrine (although this variable is only significant in specifications without fixed effects, so it appears to be neighborhood specific).

Overall, our results show that a number of variables are able to significantly predict MFI borrowings, and help differentiate between MFI vs. no loan and MFI vs. other loans. This reassures the validity of our methodology. The overall picture that emerges from these regressions is that MFI borrowers are more likely to be middle aged, have low education and be

[21] While in theory MFI borrowing can be used to purchase land, we control for large land ownership, which is unlikely to be endogenous to MFI participation because of a small size of MFI loans.

[22] The water source and distance characteristics are not significant in comparing Spandana and any MFI borrowers to those without loans, but tap water is positive in models 7-8 and negative in model 9.

relatively poor (i.e., overcrowded living conditions, no landline, receive government assistance), and have prior experience with MFI. They are more likely to have a health related accident or a loss of property via theft, fire, etc. The characteristics of MFI borrowers are mostly similar when compared to those without any loans vs. those with other types of loans. The fact that MFI borrowers appear to be *a priory* poorer implies that our outcome results (such as higher expenditures) are not likely to be attributed to preexisting differences in poverty levels.

4.B. The Impact of MFI borrowing

In this section we turn to our third question, specifically the differences in the impact of microfinance borrowing on household consumption and expenditures. As before, we run three models: (1) Spandana vs. no loans, (2) any MFI vs. no loans, and (3) any MFI vs. other types of loans. Because we are mostly interested in the impact of any MFI borrowing, rather than a specific impact of Spandana, we focus our attention on results of models (2) and (3). Since the results for models (1) and (2) are very similar, we don't report the results for model (1), they are available upon request.

As we discussed above, we report results with four matching methods: (1) nearest neighbor matching with replacement; (2) nearest neighbor matching without replacement; (3) Kernel matching; and (4) Stratification matching. We report 8 specifications for each outcome variable: four matching methods, each reported with and without slum fixed effects.

Table 4 reports the results for average treatment effects for comparing any MFI borrowers with those without any other loans. We find the following significant results: increase in home repairs, increase in durable goods purchases, increase in health expenditures, increase in temptation goods and festivals. The most significant of these are durable purchases, home repairs, and festivals; they are significant in 8 out of 8 models. Health expenditures and temptations goods are significant in 7 out of 8 models (both of these are insignificant in nearest neighbor matching methods). The non-durable expenditures, food expenditures, and education expenditures are generally not significant.

Interestingly, the total expenditures results are mixed. While all 8 results are positive, only 4 out of 8 are statistically significant. Thus, the evidence for total expenditures is not robust to matching methods. At a first glance it is surprising, given that several of the sub-categories of the total showed significant positive increase, but none showed a significant decline. This result is easily explained by the expenditure composition. The categories that show significant increases (durables, temptation goods, and health expenditures) represent a relatively small portion of overall household expenditures, while non-durables and food expenditures, which show no significant differences, represent a bulk of the total (see Table 1, Panel B).[23] Thus, averages for durables are only about 6% of total expenditures, health - 8%, and temptation goods - 5%. However, these are means, which are likely to be skewed by a few large outliers. Considering medians for these categories, durables represent 2%, health 5%, and temptation goods 3%.

Next we compare any MFI borrowers to those with other types of loans. Results are reported in Table 5. The two categories that show most significant and robust results are durables and home repairs (both are significant at 5% or better in all 8 out of 8 models). Again, columns 1 and 5 which are estimated by nearest neighbor with replacement algorithm have the lowest magnitude and weaker significance than other models. Expenditures on temptation goods are less robust and show significant increase in 7 out of 8 estimations, while festivals are only significant in 5 out of 8 categories. Again, the non-significant results are obtained in the nearest neighbor with replacement models. Interestingly, health expenditures are negative in all 8 estimations, but only significant in 4 out of 8 models. The total expenditures are not significant, along with non-durables, food and education.

To summarize, the comparison of MFI borrowers to those without loans and those borrowing from other sources yields some of the same results: increased durable purchases and home repairs are similarly significant in both cases, while the differences in festivals and temptation goods is slightly weaker when comparing MFI borrowers with other types of loans. These results make sense since MFI loans are often obtained to buy a small durable good, like an appliance (e.g., sewing machine, fridge, etc.) or can be used for small home repairs. In addition, MFI loans obtained for small business purposes are also likely to result in durables increase and home improvement. Health expenditures, however, show different patterns for these two comparison groups and a clear increase is only observable in the case of MFI borrowers vs. non-borrowers. This is consistent with our propensity score estimations, which showed that

[23] Home repairs and festivals/celebrations are not included in the total consumption expenditures consistent with Banerjee et al (2014). However, relative to totals these still are fairly small categories (see Table 1, Panel B).

households with health accident were equally likely to use MFI and other sources of informal lending.

The magnitudes of the effects appear fairly large. Since our outcome variables are in log form, the estimated coefficients show a percent increase in the variable. Thus, on average the home repairs are increased by 80-90% in a group of MFI borrowers compared with non-borrowers, but only increased by about 30-40% compared to those with other informal loans. Durable goods purchases increase on average by about 30-35% in both comparison groups. Festivals expenditures are increased by about 40-45% comparing to those without loans but only by about 13% relative to those with other sources of credit. Temptation goods expenditures also increase more relative to no loan group - by about 60% - while relative to other borrowers the increase is only about 20-25%.

Two points are worth noting here. First, except for durable expenditures, the increase in magnitude is larger in comparing MFI borrowers to non-borrowers vs. those borrowing from other sources. This is not surprising, since those borrowing from other sources are likely to be less constrained in their access and hence would experience a smaller change in their spending patterns relative to MFI borrowers. It is also possible that MFI borrowing substitutes for some other types of informal credit (weak evidence of this is discussed earlier). Second, while these appear as large numbers, recall that the categories increased are relatively small percent of total expenditures (e.g., durables, temptation goods).

Comparing the four matching methods, we find that nearest neighbor with replacement produces less significant results relative to other methods in several cases. The larger variance of this method has been reported before (eg. Caliendo and Kopeinig, 2008). This is because the same observation is sampled multiple times, which increases the variance. The good news is that in most cases all four methods produce very similar results.

4.C. Spillovers

In this section we turn to our fourth question and evaluate the presence of spillovers. Until now we have studied Spandana and other MFI borrowers relative to non-borrowers and those borrowing from other sources, all living in the same “treated” areas. There are two main advantages of limiting participants and non-participants to be located in the same neighborhood.

First, we can control for neighborhood specific fixed effects, which captures the common neighborhood characteristics that may influence both the propensity to participate in MFI borrowing and the outcome variables. Second, prior research established that PSM method performs better when participants and non-participants come from the same local labor markets (Heckman et al., 1997, 1998a, 1998b). Therefore, our main analysis up to now was limited to the 52 neighborhoods in which Spandana was randomly established.

However, the other 52 neighborhoods in which Spandana was not initially established (i.e. the RCT design control areas) can serve a useful role in our design. Since the initial allocation of Spandana branches was determined randomly, households residing in these neighborhoods should be similar to those residing in neighborhoods with Spandana. They should make a good match to the Spandana or other MFI borrowers. One advantage for considering these non-Spandana households as control group for the propensity score matching is that it allows us to test for the presence of spillovers from MFI borrowers to non-borrowers.

Spillovers (the consequences for households non-participating in MFI) are likely to occur in the same neighborhood. For example, if one household borrows from an MFI to purchase a TV, a DVD player or a satellite dish, the family, friends or the neighbors may decide to postpone their purchases of the same asset and instead just come over and watch TV at the MFI borrower's home. Some other small appliances, tools or even bicycles could also be shared with neighbors and friends. Similarly, spending on parties, festivals, eating out or alcohol purchases (i.e., "temptation goods" expenses) could also decline for non-participants if the MFI borrower chooses to "pick up the tab" more often (possibly reversing the roles later). In these examples, comparing MFI borrowers and non-borrowers in the same neighborhood will overestimate the true impact of MFI on the MFI borrowers (because non-borrowers are spending less and hence cannot be used as a valid counterfactual). The spillovers could go in the opposite direction if the spending of the MFI borrowers lead to increased spending by non-borrowers (in which cases the impact on borrowers is underestimated). For example, if an MFI borrower uses the loan to grow his business and hires someone in the village as an employee, the employee's household may also increase their spending. This, a priori it is not clear which direction the spillovers will go, but it is sensible to assume they are likely to have a local nature. Thus, by comparing the MFI participants from treated neighborhoods, i.e. those that got the Spandana random placement, with non-participants from control neighborhoods, i.e. the neighborhoods that did not get a Spandana branch, we should obtain the results free of spillover effects. In this section we report the results of these comparisons.

First, we re-estimate the PSM model using Spandana/other MFI borrowers from treated neighborhoods and non-borrowers or other informal borrowers from control neighborhoods only. Table 3 reports these results in columns 3, 6, and 9. Note that since our MFI participants and non-participants come from different neighborhoods, we are not able to include neighborhood specific fixed effects in these regressions (since they predict the outcome perfectly). Therefore, we compare the results to those obtained earlier without fixed effects (i.e., column 3 vs. 1, column 6 vs. 4, and column 9 vs. 7). Notably, most of the results predicting MFI participation are very similar when we consider these alternative control groups.[24] Thus, the same variables predicting any MFI participation relative to non-participants in the same neighborhood are the same when we consider non-participants in control neighborhoods. This is expected due to the RCT design, since the control and treatment households are expected to be similar.

Next, we perform the same outcome comparisons as before. First, we compare MFI borrowers (i.e., Spandana or other MFI) to those without other loans; results are reported in Table 6. These results should be compared to Table 4 left panel (i.e., models 1-4 estimated without fixed effects). We observe lower magnitude and less significant expenses on durable goods (only significant in column 2). The festivals expenditures are significant in 3 out of 4 models, but the magnitude is a little lower (i.e., about 25-38% in Table 6 vs. 33-49% in Table 4). Expenditures on temptation goods are strongly significant, but also a little lower in magnitude (change of about 45-55% in Table 6, while in Table 4 it was 56-77%). Home repairs are strongly significant in 4 out of 4 models; however, the magnitudes are again smaller: the estimates in Table 6 are in 60-70% range, while comparable estimates in Table 4 are in 80-90% range. Only health expenditures are about the same magnitude as before, at about 30% (and are larger and slightly more significant in models 1 and 2).

Next, in Table 7 we compare results for any MFI borrowers relative to those with other loans in control neighborhoods. The comparable results are in Table 5, models 1-4. Durable expenditures are again significant in 4 out of 4 models, but slightly lower in magnitude (20-24%

[24] One result that is different is significant effects on "latrine not shared with others" dummy variable. It is positive and significant in models 3, 6, and 9, while it was only marginally significant in models 1 and 4 and significant in model 7. This could be that this particular variable is a neighborhood-specific characteristic and the control neighborhoods just happened to have more shared latrines.

in Table 7 vs. 30-38% in Table 5). Interestingly, expenditures on festivals, home repairs, and temptation goods are much smaller in magnitude and in general no longer statistically significant. Health expenditures are now significantly negative and larger in magnitude.

Overall, these results are consistent with the hypothesis of negative spillovers: the non-borrowers spend less, since the borrowers are picking up some of the non-borrower's spending. These negative spillovers are pronounced for durable goods, festivals, temptation goods, and home repairs. For all these categories of expenditures it is plausible that households able to access the new source of credit (any MFI) would spend some of the borrowed money on their family, friends and neighbors, i.e. "pick up the tab" and share in their benefits of access to MFI loans.[25] However, health expenditures do not seem to have a negative spillover effect. While these spillover results are interesting, they should be treated with caution, since we are unable to control for the neighborhood-specific fixed effects and, according to previous research, the PSM method performs less reliably when participants and non-participants come from different labor markets.[26]

5. Discussion and Caveats

Relative to the original Banerjee et al. (2014) paper, our results are mixed. The two of the most important results of Banerjee et al. (2014) – the increase in durable purchases and the lack of increase in total expenditures – are confirmed in our paper. We add some nuances to these results since we are able to compare MFI borrowers relative to non-borrowers and relative to those borrowing from other sources. This is reassuring, and shows that despite of different methodologies, these two results prove to be very robust. The increase in durables also appears as the most logical consequence of microfinance, since it consists of relatively small loans, which are likely to be used by households to make lumpy purchases they would otherwise not be able to make.

[25] Although this reasoning does not work for home repairs as these are unlikely to generate negative spillovers of the sort we describe.
[26] However, it is plausible to consider all of the 104 neighborhoods in this study as representing one local labor market. Consistent with this argument, our PSM estimates rarely change when fixed effects are included.

However, there are some differences.[27] We find that relative to households without any other loans, microfinance borrowers have higher expenditures on health and house repairs. These can be considered "positive" results and suggest that microfinance can be used toward desirable ends, such as improvement in living conditions and in human capital. Contrary to Banerjee et al. (2014), we find an increase in "temptation goods" expenditures (such as eating out, alcohol/tobacco, and gambling), and festivals. These could be considered "negative" impacts, as such expenses, while possibly giving a utility boost in the short term, are not likely to have any positive long term effects. However, some of these results could be due to spillovers, i.e. the borrowers increasing their spending, while non-borrowers reducing it. We present some evidence consistent with such negative spillovers. Thus, the total outcome (i.e. taking into account negative spillovers) is consistent with Banerjee et al. (2014) who report reduced form estimates (i.e., the average for MFI borrowers and non-borrowers in the same neighborhood).

Nevertheless, it seems that, at least some MFI borrowers use their access to new source of credit toward seemingly fruitless choices. While somewhat disappointing, these results are not totally unexpected. For example, Banerjee and Duflo, 2011 suggest that poor people in a similar environment could spend up to 30% more on food if they cut the expenditures on alcohol, tobacco, and festivals . Previous research (cited in Banerjee and Duflo, 2011) has also found that of an additional 1% increase in income the poor spend only 0.67% on food. More importantly, even when the poor do spend more on food, they do not spend it on additional calories, but on better tasting and more interesting food- i.e. they are likely to buy more "junk food" (i.e., food with low nutritional value, such as sugary treats) or spend extra on more expensive food options to enhance variety and taste (Banerjee and Duflo, 2011). In another poignant example, Banerjee and Duflo (2011) tell a story of a man whose family did not have enough to eat, but had a TV, parabolic antenna, and a DVD player. Thus, it appears plausible that the poor, when given a chance for extra new credit, such as introduction of microfinance, are likely to make choices that make their lives a little more interesting/bearable, if only for a moment. This would explain a raise in expenditures on temptation goods, festivals and, even home repairs, as such could be pure cosmetic choices rather than improving quality (we don't have enough data to test this

[27] One possible reason for some of these differences in these results is the presence of outliers. As Figure 1 demonstrates, the distributions of various expenditure categories have some very large outliers. Because of this we use log transformation to reduce the influence of outliers. As an additional robustness check, we reproduced all outcome estimations after dropping 1% on the top tail of the distribution (after log-transformation). Our results remain largely unchanged (available on request).

hypothesis). Not surprising, therefore, is the lack of strong drastic and long-lasting positive impacts of microfinance recently documented by many experimental studies (cited in footnote 3).

The important caveat for our analysis, of course, is the ability of the PSM method to produce unbiased estimates of the true impacts. While previous research has shown that PSM can provide reasonably accurate estimates in comparison with experimental evidence, these simulations were done on labor market participants (Heckman et al. 1997, 1998a, 1998b) or participants in a conditional cash transfer program (Diaz and Handa 2006), rather than microfinance. These authors argued that unobservables result in a relatively minor bias, compared to bias stemming from lack of or incorrect use of observables. Arguably, the unobservable problem might be more severe in microcredit than in other settings. For example, entrepreneurial spirit, talent or motivation to escape poverty may be correlated with both self-selection into borrowing and better long run outcomes. Therefore, the estimates of the impact of microfinance are likely to be biased upward.[28]

To the best of our knowledge, this paper is the first one to make the evaluation of the impact of microfinance borrowers relative to two distinct comparison groups: non-borrowers and those borrowing from other sources. This distinction is important as it minimizes the expected role unobservable characteristics play in self-selection of MFI borrowers relative to those borrowing from other sources. In fact, it is highly plausible that those with higher entrepreneurial spirit, talent or motivation would have borrowed from other sources even prior to an MFI entering the area. The prevalence of a variety of informal and formal borrowing arrangements in the area of the study implies that credit was widely used in this sample of relatively poor (but not the poorest) households even prior to Spandana or other MFI's entering the area (as documented by Banerjee et al. 2014). Hence, when compared to those with other sources of credit, MFI borrowers are relative latecomers to the borrowing scene and could arguably have lower entrepreneurial spirit. This implies that the bias (due to unobservable entrepreneurial spirit, talent or motivation) can actually go the other way relative to those who already borrow from other sources. Thus, our estimates can be seen as straddling the upward biased results relative to those

[28] Angelucci et al. (2013) also discuss a second source of unobservable bias that steams from the decision of lenders to enter into areas with higher growth potential. However, since in our main estimations our participants and non-participants come from the same slums this source of bias is not a concern.

who do not borrow at all and downward biased results relative to those borrowing from other sources. Our estimates for home repairs, temptation goods, and festivals follow this hypothesis, as the impact relative to those who do not borrow is on average much larger than the impact relative to those that borrow from other sources. However, these are not the categories one would associate with the entrepreneurial spirit. As we discussed above, these are likely to be unproductive expenditure that while improving utility in the short term, are unlikely to lead to any significant long term changes. The fact that these show larger increases in MFI borrowers relative to those with no other loans is consistent with our previous discussion that given new sources of credit poor people would make choices to make their lives in the present a little more interesting or bearable.

The purchase of durables, however, can possibly be an investment that may lead to additional future benefits if these durables are used for a small business.[29] For this category of expenditures, the omitted entrepreneurial spirit could be a source of the bias. Interestingly, the magnitudes we find for durable expenditures using two comparison groups (i.e., those without any loans and those with other types of loans) are very similar in magnitude. Specifically, the average magnitude for durable purchases among the 8 models we estimate for those without loan is 0.35, while it is 0.31 for those with other loans. If our reasoning on the potential directions of bias in the two comparison groups is correct, these results suggest that omitted entrepreneurial characteristics are not likely to be causing significant bias in PSM estimates.

6. Conclusions

We employ PSM method to evaluate the impact of microfinance on various expenditure categories. While we use the data from a recent RCT experiment (Banerjee et al., 2014), our approach is able to answer a set of interesting and important questions unaddressed by the RCT design.

We contribute to existing literature on evaluation of the impact of microfinance in several important ways. First, we provide evidence on the impact of microfinance on individual level, which is not possible using RCT designs, which can only produce either ITT or LATE estimates (see footnote 4). While such parameters can be of interest, in most cases the policymakers would

[29] Of course it could also be used to purchases "unproductive" durables such as TV, DVD player, etc.

like to know the impact of microcredit on people who actually take it up (i.e., those that obtain a loan). Second, we describe characteristics of microfinance borrowers relative to those without loans and relative to those who also borrow from other sources. This is important for program targeting and allows for better understanding of factors influencing demand for microfinance. Third, we compare the outcomes of microfinance users to those without other loans and those who borrow from other sources. This allows us to test whether microfinance has same or different benefits relative to other sources of informal credit most often used by the poor. Fourth, we are able to evaluate the presence of spillovers and evaluate the magnitude of the biases commonly attributed to unobservable characteristics, such as entrepreneurial spirit or talent.

Finally, our results corroborate some of the key findings of the RCT that produced the dataset we use, notably the increase in durable purchases and lack of increase in total expenditures, thus confirming the robustness of these results using a completely different methodology. This also has implications for the validity of PSM methodology for estimating the impact of microfinance. Thus, our paper serves as a great complement to the recent emergence of RCT papers (cited in footnote 3).

References

Angelucci, M., D. Karlan, and J. Zinman (2013). Win some lose some? Evidence from a randomized microcredit program placement experiment by Compartamos Banco. *American Economic Journal: Applied Economics*, forthcoming

Attanasio, O., B. Augsburg, R. De Haas, E. Fitzsimons, and H. Harmgart (2013). Group lending or individual lending? Evidence from a randomised field experiment in Mongolia. Pub ref: MPRA Paper No. 35439.

Augsburg, B., R. D. Haas, H. Harmgart, and C. Meghir (2013). Microfinance, poverty and education. IFS working paper.

Beaman, Lori, Dean Karlan, Bram Thuysbaert, and Christopher Udry. 2014, Self-Selection into Credit Markets: Evidence from Agriculture in Mali. Mimeo. Yale University.

Banerjee, Abhijit and Duflo, Esther, 2011, "Poor Economics: A radical rethinking of the way to fight global poverty." Public Affairs, 250 W. 57 St, New York, NY.

Banerjee, Abhijit, Duflo, Esther, Rachel Glennerster, and Cynthia Kinnan (2014), "The Miracle of Microfinance? Evidence from a Randomized Evaluation," *American Economic Journal: Applied Economics*, forthcoming.

Caliendo, Marco and Sabine Kopeinig, 2008, "Some Practical Guidance for the Implementation of Propensity Score Matching,"*Journal of Economic Surveys* Vol. 22, No. 1, pp. 31–72.

Chen, Shaohua & Mu, Ren & Ravallion, Martin, 2009. "Are there lasting impacts of aid to poor areas?," *Journal of Public Economics*, Elsevier, vol. 93(3-4), pages 512-528, April.

Crépon, Bruno; Devoto, Florencia; Duflo, Esther and Pariente, William, 2014, "Estimating the Impact of Microcredit on Those who Take It Up: Evidence from a Randomized Experiment in Morocco," *American Economic Journal: Applied Economics*, forthcoming

Dehejia, Rajeev H. and Sadek Wahba, 2002, Propensity Score Matching Methods for Nonexperimental Causal Studies. The Review of Economics and Statistics, February 2002, 84(1): 151–161.

Desai, Jaikishan , Kristin Johnson and Alessandro Tarozzi, 2013. On the Impact of Micro-credit: Evidence from a Field Experiment in Rural Ethiopia. Duke University Mimeo.

Diaz J. J., and Handa, S. ,2006. An assessment of propensity score matching as a nonexperimental impact estimator: Evidence from Mexico's PROGRESA program. *Journal of Human Resources*, 41(2), 319-345.

Duflo, Esther & Glennerster, Rachel & Kremer, Michael, 2008. "Using Randomization in Development Economics Research: A Toolkit," Handbook of Development Economics, Elsevier.

Floro, Maria and Ranjula Bali Swain, 2012, Assessing the Effect of Microfinance on Vulnerability and Poverty among Low Income Households, The Journal of Development Studies, Volume 48, Issue 5, pages 605-618.

Hahn, Jinyong, Keisuke Hirano, and Dean Karlan. 2008. “Adaptive Experimental Design Using the Propensity Score.” Working Paper 969, Economic Growth Center, Yale University, New Haven, CT.

Heckman, James, Hidehiko Ichimura, and Petra Todd, 1997, “Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program,” *Review of Economic Studies*. 64: 605-654.

Heckman, James, Hidehiko Ichimura, and Petra Todd, 1998a, “Matching as an Econometric Evaluation Estimator.” *Review of Economic Studies*. 65: 261-294.

Heckman, James, Hidehiko Ichimura, Jeffrey Smith and Petra Todd, 1998b, “Charactrizing Selection Bias Using Experimental Data,” *Econometrica*. 66: 1017-1089.

Jalan, Jyotsna and Martin Ravallion, 2003a, “Does Piped Water Reduce Diarrhea for Children in Rural India?” *Journal of Econometrics* 112: 153-173.

___________ and _____________, 2003b, “Estimating Benefit Incidence for an Anti-poverty Program using Propensity Score Matching,” *Journal of Business and Economic Statistics*, 21(1): 19-30.

Glazerman, Steven, Dan Levy and David Myers, 2003, “Non-Experimental versus Experimental Estimates of Earnings Impacts,” *Annals of the American Academy of Political and Social Sciences* 589: 63-93.

Godtland, Erin, Elizabeth Sadoulet, Alain De Janvry, Rinku Murgai and Oscar Ortiz, 2004, “The Impact of Farmer Field Schools on Knowledge and Productivity: A Study of Potato Farmers in the Peruvian Andes,” Economic Development and Cultural Change, 53(1): 63-92.

Kaivan, Munshi and Mark Rosenzweig, 2006, Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy, *American Economic Review* 96(4): 1225-1252.

Karlan, Dean, and Jonathan Zinman. 2010. “Expanding Credit Access: Using Randomized Supply Decisions to Estimate the Impacts.” *Review of Financial Studies* 23 (1): 433–464.

———. 2011. "Microcredit in Theory and Practice: Using Randomized Credit Scoring for Impact Evaluation." *Science* 332 (6035) (June 10): 1278–1284.

Khandker, Shahidur R. 2006. "Microfinance and Poverty: Evidence Using Panel Data from Bangladesh." *World Bank Economic Review* 19 (2): 263–86.

Pitt, Mark M. 2014, "Response to 'The Impact of Microcredit on the Poor in Bangladesh: Revisiting the Evidence'," *Journal of Development Studies*. 50(4): 605-610.

Munshi, Kaivan and Mark Rosenzweig, 2006. Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy.American Economic Review, vol. 96 (4), pp. 1225-1252.

Pitt, Mark, and Shahidur Khandker. 1998. "The Impact of Group-Based Credit Programs on Poor Households in Bangladesh: Does the Gender of Participants Matter?" *Journal of Political Economy* 106 (5): 958–98.

Ravallion, Martin, 2003, "Assessing the Poverty Impact of an Assigned Program," in Bourguignon, F. and L. Pereira da Silva (eds.) *The Impact of Economic Policies on Poverty and Income Distribution*, New York: Oxford University Press.

Ravallion, Martin. 2008. "Evaluating Anti-poverty Programs." In *Handbook of Development Economics*, vol. 4, ed. T. Paul Schultz and John Strauss, 3787–846. Amsterdam: North-Holland.

Ravallion, Martin and Gaurav Datt, 1995. "Is Targeting through a Work Requirement Efficient? Some Evidence for Rural India," in D. van de Walle and K. Nead (eds) *Public Spending and the Poor: Theory and Evidence*, Baltimore: Johns Hopkins University Press. 91

Ravallion, Martin, Emanuela Galasso, Teodoro Lazo and Ernesto Philipp, 2005, "What Can Ex-Participants Reveal About a Program's Impact?" *Journal of Human Resources*, 40 (Winter): 208-230.

Ravallion, Martin, Dominique van de Walle and Madhur Gaurtam, 1995, "Testing a Social Safety Net," *Journal of Public Economics*, 57(2): 175-199.

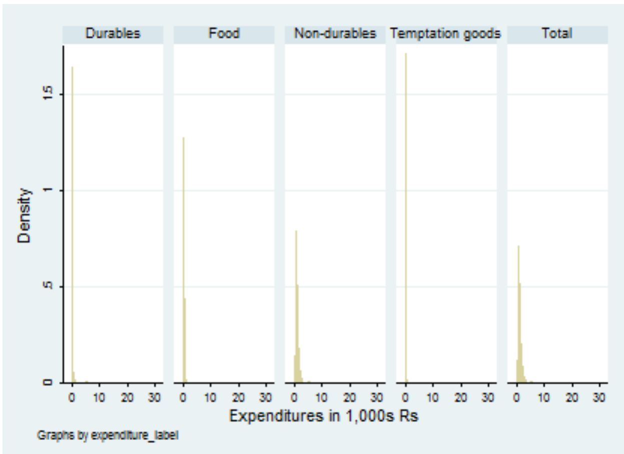
Ravallion, Martin and Quentin Wodon, 2000, "Does Child Labor Displace Schooling? Evidence on Behavioral Responses to an Enrolment Subsidy," *Economic Journal* 110: C158-C176.

Roodman, David and J. Morduch. 2014. The Impact of Microcredit on the Poor in Bangladesh: Revisiting the Evidence, *Journal of Development Studies*, 50(4): 583-604.

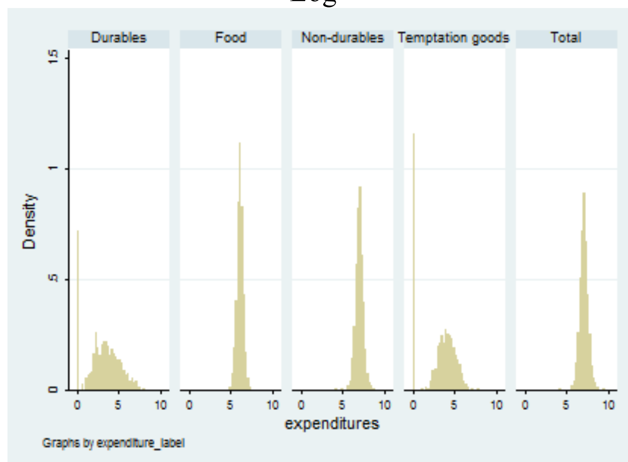
Smith, Jeffrey and Petra Todd, 2005, "Does Matching Overcome LaLonde's Critique of Non-experimental Estimators?" *Journal of Econometrics*, 25 (2005) 305–353.

Figure 1: Spending by expenditure category

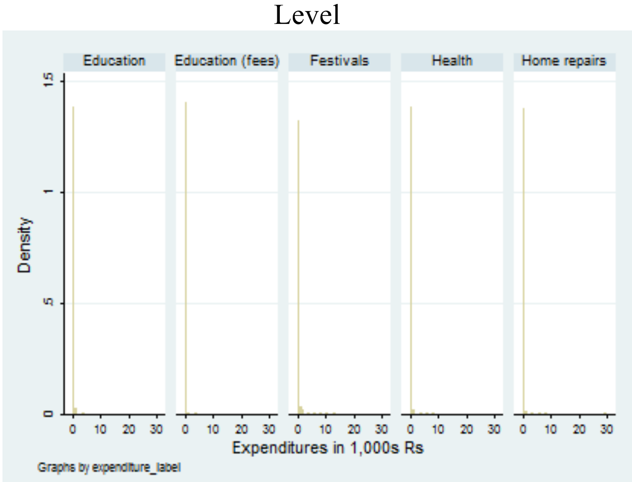
Level



Log



Level



Log

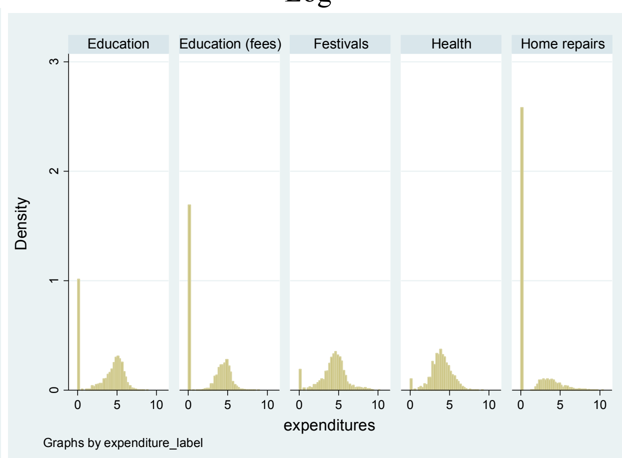
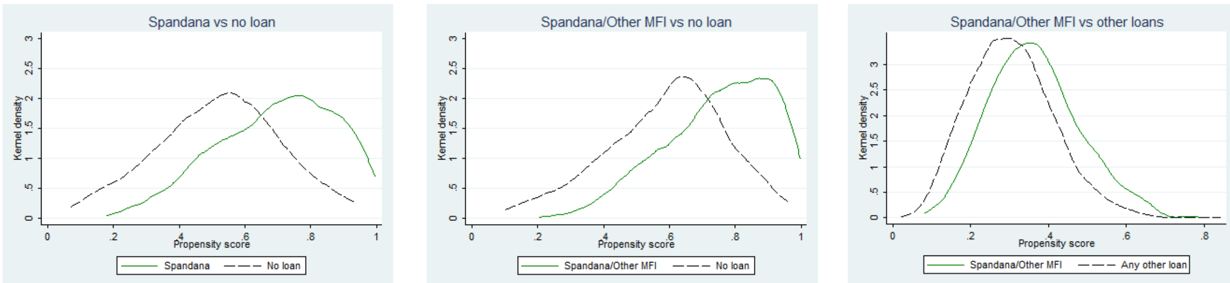


Figure 2: Propensity score by outcome

Panel A: Without slum fixed effects



Panel B: With slums fixed effects

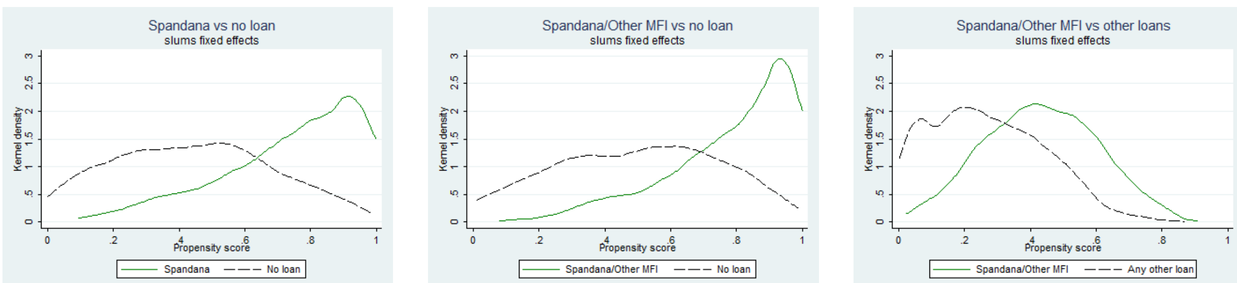


Table 1: Summary statistics

Panel A. Outcome variables

Variable	Mean	Std. Dev.	Min	Max
Expenditure measures (monthly per capita)				
Log total household expenditures, Rs2007 (N=3308)	7.10	0.54	4.22	10.10
Log non-durable expenditures, Rs2007 (N=3284)	7.03	0.50	4.22	10.10
Log durable expenditures, Rs2007 (N=3284)	3.11	1.99	0.00	9.29
Log food items, Rs2007 (N=3249)	6.12	0.39	4.52	9.28
Log temptation goods, Rs2007 (N=3258)	2.96	2.05	0.00	7.83
Log festivals, Rs2007 (N=3303)	4.35	1.70	0.00	10.13
Log health, Rs2007 (N=3299)	3.96	1.37	0.00	9.02
Log education (total), Rs2007 (N=3275)	3.51	2.30	0.00	8.88
Log education (fees), Rs2007 (N=3249)	2.59	2.33	0.00	8.83
Log home repairs >500, Rs2007 (N=2700)	1.45	2.17	0.00	10.27

Note: Number of observations is 3,318, unless indicated otherwise. The summary statistics is limited to treated slums. All expenditures are adjusted for inflation (Rs 2007).

Panel B. Expenditure shares by category (average and median)

Variable	Obs	Mean	Std. Dev.	Min	Max	Median
Non-durable expenditures, Rs2007	3284	0.94	0.11	0.02	1.00	0.98
Durable expenditures, Rs2007	3284	0.06	0.11	0.00	0.98	0.02
Food items, Rs2007	3249	0.39	0.12	0.02	0.96	0.39
Temptation goods, Rs2007	3258	0.05	0.07	0.00	0.56	0.03
Festivals, Rs2007	3301	0.22	0.70	0.00	20.99	0.08
Health, Rs2007	3299	0.08	0.11	0.00	0.94	0.05
Education (total), Rs2007	3275	0.10	0.11	0.00	0.92	0.07
Education (fees), Rs2007	3249	0.05	0.08	0.00	0.87	0.03
Home repairs >500, Rs2007	2698	0.09	0.53	0.00	15.22	0.00

Note: Each variable is calculated as share of total for the household and average/median of the household level shares are reported in the table. Home repairs and expenditures on festivals and celebrations are not included in the total consumption expenditures consistent with Banerjee et al (2014).

Panel C. Characteristics of household (hhd) and its members

Head of hhd is a woman	0.10	0.29	0	1
Age of oldest qualifying female	34.46	8.79	18	55
Female: education 5-10 standard	0.40	0.49	0	1
Female: education 10+ standard	0.05	0.22	0	1
Female: education is unknown	0.49	0.50	0	1
Female: stable employment (4y+) (primary job)	0.16	0.37	0	1
Female: stable employment status is unknown (primary job)	0.73	0.44	0	1
Male: education 5-10 standard	0.47	0.50	0	1
Male: education 10+ standard	0.11	0.32	0	1
Male: education is unknown	0.35	0.48	0	1
Persons per room > 2	0.60	0.49	0	1
There are dependents	0.70	0.46	0	1
Have children age 0-2	0.20	0.40	0	1
At least 1 kid age 0-16 is currently studying in English	0.48	0.50	0	1
Language of instruction is unknown for all curr studying kids 0-16	0.08	0.28	0	1
50% or more are pucca rooms	0.42	0.49	0	1
Have a landline in hhd	0.13	0.34	0	1
Latrine is not shared with other hhd	0.60	0.49	0	1
Owner with the title	0.79	0.41	0	1
Water source is tap	0.90	0.30	0	1
Water source within premises	0.58	0.49	0	1
Water source distance unknown	0.33	0.47	0	1
Big plot owner in city or village (> 1 acre)	0.10	0.30	0	1
Repaid MFI loan before 2006	0.07	0.26	0	1
Hhd receives government assistance	0.09	0.28	0	1
Accident in the hhd	0.08	0.27	0	1
Lost job in the hhd	0.01	0.12	0	1
Lost property (>500Rs) due to fire, theft, flood or accident last year	0.10	0.30	0	1
Number of people in the hhd is between 1 and 4	0.31	0.46	0	1
Number of people in the hhd is 5-6	0.43	0.50	0	1
Number of qualifying women permanently residing in hhd is 2+	0.34	0.47	0	1

Note: Number of observations is 3,318, unless indicated otherwise. The summary statistics is limited to treated slums. All expenditures are adjusted for inflation (Rs 2007).

Table 2: Loan Characteristics

Panel A. Household breakdown by number of loans

	# of households	%
no loans	404	12.18
1	632	19.05
2	570	17.18
3	517	15.58
4	415	12.51
5+	780	23.5
Total	3,318	100.00

Panel B. Household breakdown by the loan type

At least 1 loan from	# of households	% of total (N=3,318)	% of those with any loan
Spandana	687	21%	24%
Other MFI	435	13%	15%
Family/neighbor/friend	1,116	34%	38%
Money lender	1,236	37%	42%
Shopkeeper	584	18%	20%
Chit fund	556	17%	19%
Commercial bank or financial company fund	618	19%	21%
Other (not specified)	1,043	31%	36%

Panel C. Combination of loans

Combination of loans		% of Spandana loans
Spandana AND other MFI	178	26%
Spandana AND family/neighbor/friend	215	31%
Spandana AND money lender	292	43%
Spandana AND shopkeeper	127	18%
Spandana AND chit fund	134	20%
Spandana AND commercial bank/fin company	143	21%
Spandana AND other unspecified	206	30%
		% of those with other MFI
Other MFI AND family/neighbor/friend	132	30%
Other MFI AND money lender	198	46%
Other MFI AND shopkeeper	87	20%
Other MFI AND chit fund	87	20%
Other MFI AND commercial bank/fin company	98	23%
Other MFI AND other unspecified	145	33%
		% of those with either Spandana or other MFI
Spandana/Other MFI AND family/neighbor/friend	294	31%
Spandana/Other MFI AND money lender	414	44%
Spandana/Other MFI AND shopkeeper	181	19%
Spandana/Other MFI AND chit fund	177	19%
Spandana/Other MFI AND com. bank/fin.comp	196	21%
Spandana/Other MFI AND other unspecified	298	32%

Table 3: Propensity score estimation

	Dependent variable								
	Spandana =1; No loan=0			Spandana/Other MFI=1; No loan=0			Spandana/Other MFI=1; Other loans=0		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Head of household (hhd) is women	0.270	0.133	-0.068	0.204	0.040	-0.091	0.021	-0.075	-0.131
	(0.17)	(0.19)	(0.17)	(0.16)	(0.18)	(0.16)	(0.10)	(0.10)	(0.09)
Age of oldest qualifying female	0.113**	0.134**	0.079*	0.099**	0.115**	0.061+	0.063**	0.055**	0.023
	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Age of oldest qualifying female squared	-0.002**	-0.002**	-0.001*	-0.001**	-0.002**	-0.001*	-0.001**	-0.001**	-0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Female: education 5-10 standard	-0.401+	-0.326	-0.089	-0.397+	-0.332	-0.078	0.079	0.082	0.065
	(0.22)	(0.23)	(0.19)	(0.20)	(0.22)	(0.18)	(0.11)	(0.11)	(0.11)
Female: education 10+ standard	-0.742**	-0.655*	-0.676**	-0.737**	-0.671*	-0.619**	-0.035	-0.077	-0.114
	(0.28)	(0.31)	(0.26)	(0.26)	(0.29)	(0.24)	(0.16)	(0.17)	(0.16)
Female: stable employment (4y+) (primary job)	-0.386*	-0.241	-0.260	-0.301+	-0.238	-0.195	-0.056	-0.005	-0.003
	(0.19)	(0.22)	(0.18)	(0.18)	(0.20)	(0.17)	(0.09)	(0.10)	(0.09)
Male: education 5-10 standard	-0.429*	-0.476*	-0.398*	-0.370*	-0.434*	-0.331+	-0.071	-0.107	0.025
	(0.19)	(0.21)	(0.20)	(0.18)	(0.20)	(0.18)	(0.10)	(0.11)	(0.10)
Male: education 10+ standard	-0.851**	-0.990**	-0.717**	-0.859**	-0.964**	-0.719**	-0.551**	-0.574**	-0.402**
	(0.23)	(0.25)	(0.24)	(0.21)	(0.23)	(0.22)	(0.13)	(0.14)	(0.13)
Persons per room > 2	0.305**	0.246*	0.539**	0.308**	0.278**	0.550**	0.150*	0.094	0.201**
	(0.10)	(0.12)	(0.11)	(0.10)	(0.11)	(0.10)	(0.06)	(0.07)	(0.06)
There are dependents	0.016	-0.025	-0.126	-0.024	-0.041	-0.154	0.150*	0.143+	0.069
	(0.12)	(0.13)	(0.13)	(0.11)	(0.12)	(0.12)	(0.07)	(0.08)	(0.07)
Have children age 0-2	-0.008	0.039	0.108	0.023	0.047	0.105	0.074	0.069	0.042
	(0.12)	(0.13)	(0.13)	(0.11)	(0.12)	(0.12)	(0.07)	(0.07)	(0.07)
At least 1 kid age 0-16 is currently studying in English	0.012	0.201	0.110	0.034	0.221*	0.106	0.062	0.124+	0.047
	(0.11)	(0.12)	(0.11)	(0.10)	(0.11)	(0.10)	(0.06)	(0.07)	(0.06)
50% or more are pucca rooms	0.054	-0.082	0.143	0.053	-0.086	0.157+	-0.080	-0.227**	0.161**
	(0.09)	(0.11)	(0.10)	(0.08)	(0.10)	(0.09)	(0.05)	(0.06)	(0.05)
Have a landline in hhd	-0.383**	-0.262+	-0.285*	-0.267*	-0.201	-0.164	0.057	0.110	0.172*
	(0.13)	(0.14)	(0.14)	(0.12)	(0.13)	(0.13)	(0.08)	(0.08)	(0.08)

Latrine is not shared with other hhd	0.146	-0.062	0.359**	0.136	-0.055	0.350**	0.111*	-0.033	0.218**
	(0.09)	(0.10)	(0.10)	(0.08)	(0.10)	(0.09)	(0.05)	(0.06)	(0.05)
Owner with the title	0.050	0.163	0.105	0.029	0.131	0.088	-0.017	-0.003	0.156**
	(0.11)	(0.13)	(0.11)	(0.10)	(0.12)	(0.10)	(0.06)	(0.07)	(0.06)
Water source is tap	0.123	0.096	-0.080	0.196	0.194	0.000	0.328**	0.234*	-0.427**
	(0.15)	(0.20)	(0.18)	(0.15)	(0.19)	(0.17)	(0.09)	(0.11)	(0.11)
Water source within premises	-0.154	-0.078	-0.028	-0.198	-0.085	-0.059	-0.096	-0.048	-0.223*
	(0.17)	(0.20)	(0.17)	(0.16)	(0.18)	(0.16)	(0.10)	(0.11)	(0.09)
Big plot owner in city or village	0.181	-0.019	0.261	0.154	0.058	0.247	0.030	-0.007	0.140
(> 1 acre)	(0.15)	(0.17)	(0.17)	(0.14)	(0.16)	(0.15)	(0.08)	(0.09)	(0.09)
Repaid MFI loan before 2006	0.513**	0.673**	0.852**	0.597**	0.623**	0.956**	0.276**	0.183+	0.384**
	(0.18)	(0.21)	(0.24)	(0.17)	(0.19)	(0.22)	(0.09)	(0.10)	(0.09)
Hhd receives government	0.741**	0.738**	0.941**	0.656**	0.762**	0.831**	0.409**	0.417**	0.454**
assistance	(0.17)	(0.20)	(0.19)	(0.16)	(0.19)	(0.18)	(0.09)	(0.09)	(0.09)
Accident in the hhd	0.554**	0.592**	0.451*	0.579**	0.669**	0.494**	-0.049	0.042	0.039
	(0.19)	(0.21)	(0.20)	(0.18)	(0.20)	(0.19)	(0.09)	(0.10)	(0.09)
Lost job in the hhd	0.053	-0.105	0.359	0.033	-0.092	0.353	0.228	0.039	-0.006
	(0.34)	(0.38)	(0.41)	(0.32)	(0.35)	(0.39)	(0.20)	(0.21)	(0.19)
Lost property (>500Rs) due to fire,	0.474**	0.532**	0.248	0.508**	0.601**	0.300*	0.098	0.087	0.003
theft, flood or accident last year	(0.16)	(0.18)	(0.16)	(0.15)	(0.17)	(0.14)	(0.08)	(0.08)	(0.08)
Number of qualifying women	-0.084	-0.091	-0.009	-0.114	-0.054	-0.054	0.050	0.036	0.089
permanently residing in hhd=2+	(0.10)	(0.11)	(0.11)	(0.09)	(0.10)	(0.10)	(0.06)	(0.06)	(0.06)
Number of people in hhd is 1-4	-0.218	-0.293+	-0.151	-0.224+	-0.261+	-0.175	-0.092	-0.181*	0.059
	(0.15)	(0.16)	(0.15)	(0.13)	(0.15)	(0.14)	(0.08)	(0.09)	(0.08)
Number of people in hhd is 5-6	-0.034	-0.019	-0.149	-0.053	-0.034	-0.160	-0.084	-0.120+	-0.020
	(0.11)	(0.13)	(0.13)	(0.11)	(0.12)	(0.11)	(0.07)	(0.07)	(0.06)
Constant	-0.233	-1.098	-0.539	0.060	-0.779	-0.168	-1.777**	-1.874**	-0.893*
	(0.75)	(0.93)	(0.81)	(0.69)	(0.80)	(0.75)	(0.44)	(0.56)	(0.44)
Observations	1,087	1,064	1,027	1,343	1,324	1,283	2,916	2,915	3,087
Control households reside in	T	T	C	T	T	C	T	T	C
Slum fixed effects	no	yes	no	no	yes	no	no	yes	no

Note: ** $p<0.01$, * $p<0.05$, + $p<0.1$. Control households can reside in treated slums (T) or control slums (C). Additionally included, but not reported, dummies for female or male education is unknown, female stable job status is unknown, language of instruction for all currently studying kids 0-16 is unknown, distance to the water source is unknown.

Table 4: Average treatment effect on the treated – Spandana/Other MFI vs No loan

	Without slums fixed effects				With slums fixed effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log monthly durable goods	0.294+	0.472***	0.369**	0.366*	0.313**	0.310**	0.341**	0.266+
	(0.182)	(0.136)	(0.147)	(0.193)	(0.138)	(0.138)	(0.152)	(0.173)
Log monthly education (fees)	-0.439**	0.116	-0.147	-0.140	-0.271+	0.024	0.007	-0.107
	(0.201)	(0.130)	(0.158)	(0.170)	(0.175)	(0.160)	(0.212)	(0.213)
Log monthly education (total)	-0.381*	0.071	-0.181	-0.228	-0.094	0.066	0.042	-0.063
	(0.207)	(0.185)	(0.166)	(0.178)	(0.245)	(0.168)	(0.230)	(0.209)
Log monthly festivals	0.331**	0.499***	0.467***	0.437***	0.509***	0.542***	0.442***	0.466***
	(0.158)	(0.117)	(0.117)	(0.135)	(0.160)	(0.108)	(0.147)	(0.138)
Log monthly food items	0.008	0.030	0.031	0.042	0.065	0.012	0.073**	0.049
	(0.037)	(0.034)	(0.031)	(0.031)	(0.054)	(0.029)	(0.037)	(0.046)
Log monthly health	0.162	0.193+	0.288**	0.293*	0.288*	0.243**	0.277*	0.235+
	(0.137)	(0.121)	(0.132)	(0.157)	(0.155)	(0.100)	(0.145)	(0.146)
Log monthly home repairs >Rs500	0.954***	0.901***	0.865***	0.836***	1.039***	0.912***	0.849***	0.895***
	(0.199)	(0.174)	(0.144)	(0.190)	(0.199)	(0.157)	(0.158)	(0.193)
Log monthly non-durable expenditures	-0.014	0.038	0.010	-0.004	0.088+	0.038	0.088*	0.057
	(0.042)	(0.036)	(0.032)	(0.045)	(0.061)	(0.035)	(0.047)	(0.043)
Log monthly temptation goods	0.568***	0.774***	0.681***	0.599***	0.280	0.766***	0.656***	0.478***
	(0.173)	(0.137)	(0.155)	(0.159)	(0.196)	(0.177)	(0.208)	(0.184)
Log monthly total hhd expenditures	0.006	0.070*	0.042	0.021	0.120**	0.063	0.118**	0.086**
	(0.046)	(0.038)	(0.044)	(0.052)	(0.057)	(0.049)	(0.050)	(0.044)

Note: *** p<0.01, ** p<0.05, * p<0.1, + p<0.15. Bootstrap standard errors with 100 repetitions are in parentheses. All variables are per capita adult equivalent. Specification (1) and (5) corresponds to the nearest neighbor matching with replacement; (2) and (6) – the nearest neighbor matching without replacement; (3) and (7) Kernel matching; (4) and (8) Stratification matching.

Table 5: Average treatment effect on the treated – Spandana/Other MFI vs Other loans

	Without slums fixed effects				With slums fixed effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log monthly durable goods	0.383***	0.382***	0.312***	0.310***	0.338***	0.392***	0.301***	0.282***
	(0.114)	(0.076)	(0.072)	(0.083)	(0.106)	(0.091)	(0.081)	(0.093)
Log monthly education (fees)	-0.093	0.021	0.056	0.042	-0.154	0.068	-0.003	-0.027
	(0.117)	(0.102)	(0.087)	(0.107)	(0.159)	(0.119)	(0.092)	(0.112)
Log monthly education (total)	-0.268**	-0.085	-0.063	-0.078	-0.254**	-0.035	-0.097	-0.131
	(0.116)	(0.109)	(0.085)	(0.084)	(0.119)	(0.095)	(0.086)	(0.101)
Log monthly festivals	0.124	0.131+	0.143**	0.132*	0.070	0.077	0.129*	0.137*
	(0.101)	(0.080)	(0.065)	(0.070)	(0.113)	(0.075)	(0.069)	(0.075)
Log monthly food items	0.011	0.022	0.029*	0.035**	0.022	0.026	0.035*	0.034**
	(0.023)	(0.018)	(0.016)	(0.016)	(0.024)	(0.019)	(0.018)	(0.017)
Log monthly health	-0.109	-0.114+	-0.106**	-0.109*	-0.037	-0.083+	-0.072	-0.070
	(0.099)	(0.072)	(0.053)	(0.061)	(0.076)	(0.057)	(0.060)	(0.065)
Log monthly home repairs >Rs500	0.384***	0.369***	0.397***	0.387***	0.271*	0.263**	0.358***	0.348***
	(0.146)	(0.111)	(0.089)	(0.100)	(0.164)	(0.123)	(0.100)	(0.106)
Log monthly non-durable	-0.016	-0.017	-0.015	-0.014	-0.072**	-0.025	-0.015	-0.018
expenditures	(0.038)	(0.025)	(0.017)	(0.019)	(0.035)	(0.020)	(0.019)	(0.020)
Log monthly temptation goods	0.297***	0.269***	0.288***	0.273***	-0.074	0.186*	0.176**	0.152+
	(0.107)	(0.096)	(0.075)	(0.087)	(0.132)	(0.102)	(0.082)	(0.099)
Log monthly total hhd	-0.003	0.000	-0.005	-0.006	-0.052	-0.012	-0.004	-0.006
expenditures	(0.028)	(0.020)	(0.021)	(0.019)	(0.038)	(0.024)	(0.022)	(0.020)

Note: *** $p<0.01$, ** $p<0.05$, * $p<0.1$, + $p<0.15$. Bootstrap standard errors with 100 repetitions are in parentheses. All variables are per capita adult equivalent. Specification (1) and (5) corresponds to the nearest neighbor matching with replacement; (2) and (6) – the nearest neighbor matching without replacement; (3) and (7) Kernel matching; (4) and (8) Stratification matching.

Table 6: Average treatment effect on the treated – Spandana/Other MFI vs No loan - in alternative control areas.

	Without slums fixed effects			
	(1)	(2)	(3)	(4)
Log monthly durable goods [a]	-0.017	0.415***	0.219	0.062
	(0.178)	(0.145)	(0.154)	(0.178)
Log monthly education (fees)	0.045	-0.124	0.101	0.050
	(0.306)	(0.169)	(0.182)	(0.205)
Log monthly education (total)	-0.197	-0.276	-0.051	-0.091
	(0.287)	(0.196)	(0.204)	(0.215)
Log monthly festivals	0.285+	0.389***	0.264*	0.154
	(0.193)	(0.129)	(0.140)	(0.137)
Log monthly food items	0.010	0.063*	0.030	-0.003
	(0.040)	(0.032)	(0.036)	(0.039)
Log monthly health	0.308***	0.327***	0.286**	0.308***
	(0.091)	(0.126)	(0.115)	(0.101)
Log monthly home repairs >Rs500	0.681***	0.660***	0.709***	0.597***
	(0.206)	(0.175)	(0.188)	(0.187)
Log monthly non-durable expenditures [a]	0.014	0.090*	0.062	0.031
	(0.047)	(0.046)	(0.050)	(0.049)
Log monthly temptation goods	0.534**	0.453***	0.541***	0.463***
	(0.225)	(0.145)	(0.165)	(0.175)
Log monthly total hhd expenditures [a]	0.023	0.122***	0.075	0.035
	(0.050)	(0.043)	(0.054)	(0.055)

Note: *** p<0.01, ** p<0.05, * p<0.1, + p<0.15. Bootstrap standard errors with 100 repetitions are in parentheses. All variables are per capita adult equivalent. Specification (1) corresponds to the nearest neighbor matching with replacement; (2) – the nearest neighbor matching without replacement; (3) Kernel matching; (4) Stratification matching. [a] Generally, the results are qualitatively similar for the calculated version of this variable.

Table 7: Average treatment effect on the treated – Spandana/Other MFI vs Other loans – in alternative control areas.

	Without slums fixed effects			
	(1)	(2)	(3)	(4)
Log monthly durable goods [a]	0.242*	0.234***	0.222***	0.208**
	(0.133)	(0.090)	(0.082)	(0.088)
Log monthly education (fees)	0.015	0.087	0.117	0.096
	(0.130)	(0.105)	(0.098)	(0.082)
Log monthly education (total)	-0.063	-0.015	0.009	-0.010
	(0.129)	(0.101)	(0.101)	(0.077)
Log monthly festivals	0.063	0.040	0.053	0.052
	(0.097)	(0.084)	(0.056)	(0.071)
Log monthly food items	-0.003	0.009	0.014	0.012
	(0.024)	(0.016)	(0.013)	(0.016)
Log monthly health	-0.242***	-0.154**	-0.122**	-0.128**
	(0.073)	(0.072)	(0.056)	(0.058)
Log monthly home repairs >Rs500	0.035	0.111	0.149+	0.130
	(0.161)	(0.122)	(0.098)	(0.115)
Log monthly non-durable expenditures [a]	-0.040	-0.028	-0.012	-0.017
	(0.034)	(0.024)	(0.019)	(0.020)
Log monthly temptation goods	0.178+	0.049	0.036	0.044
	(0.119)	(0.094)	(0.077)	(0.079)
Log monthly total hhd expenditures [a]	-0.027	-0.011	0.003	-0.001
	(0.030)	(0.022)	(0.020)	(0.022)

Note: *** p<0.01, ** p<0.05, * p<0.1, + p<0.15. Bootstrap standard errors with 100 repetitions are in parentheses. All variables are per capita adult equivalent. Specification (1) corresponds to the nearest neighbor matching with replacement; (2) – the nearest neighbor matching without replacement; (3) Kernel matching; (4) Stratification matching. [a] Generally, the results are qualitatively similar for the calculated version of this variable.

Appendix A1: Description of variables

Our construction of expenditure variables closely follows the description in Banerjee et al. (2014).

Survey questions were formulated to inquire information about the expenditures on food items, fuel, and miscellaneous goods and services (such as utility payments, entertainment, school fees, media, medical expenses, rent, etc.) in the last 30 days, as well as the total expenditures on school supplies, hospitals and nursing homes (institutional), clothing, and durable goods in the last 365 days. We consider three measures of expenditure at a monthly frequency and seven measures at an yearly frequency. Monthly expenditures are adjusted per capita in adult equivalent. Weights for conversion are: for adult males and females 1.0 and 0.9, respectively; for males and females aged 13-18, 0.94 and 0.83, respectively; for children aged 7-12, 0.67 regardless of gender; for children 4-6, 0.52; for toddlers 1-3, 0.32; and for infants 0.05 (Banerjee et al. 2014, Appendix 1). All expenditures are adjusted for inflation using consumer price index (CPI) and are expressed in 2007 rupees. Source for CPI is the Labor Bureau of India (website http://labourbureau.nic.in/ last accessed on 3/28/2014).

All expenditure categories are per capita and their descriptions are summarized below.

Expenditure	Description
Monthly total consumer expenditures	Monthly total consumer expenditure (Question B8.0_19)
Monthly non-durable expenditures	Difference between total consumer expenditure and total durable goods divided by 12 ((Question B8.0_19*12 – Question B8.0_16)/12)
Monthly food items	Sum of monthly spending on food items Question B8.0_1 to Question B8.0_8 and Question B8.0_11
Monthly temptation goods	Sum of monthly spending on meals or snacks consumed outside the home; pan, tobacco and intoxicants; and lottery tickets/gambling (Question B8.0_9 + Question B8.0_10 + Question B8.0_14.3)
Monthly durable goods	Yearly total durable goods divided by 12 (Question B8.0_16/12)
Monthly home repairs (any>Rs500)	Expenditure on repairs to the house over the last year if the amount of expenditures exceeds Rs 500 divided by 12 (Question B1.22/12)
Monthly festivals	Sum of various household expenditures on festivals and celebrations in the past year divided by 12 (Question B8.1_a1 + Question B8.1_a2 + Question B8.1_a3 +Question B8.1_a4 + Question B8.1_a5+ Question B8.1_a6)/12
Monthly health	Sum of monthly medical non-institutional expenditures and annual institutional medical fees divided by 12 (Question B8.0_14.7 + Question B8.0_15.2/12)
Monthly education (total)	Sum of monthly school fees, monthly school tuition fees, and annual total expenditures on school supplies divided by 12 (Question B8.0_14.4 + Question B8.0_14.5 + Question B8.0_15.1/12)
Monthly education (fees)	Monthly school fees and monthly school tuition fees (Question B8.0_14.4+ Question B8.0_14.5)

Quality check: Total yearly expenditure on durable goods is referenced in two places in the survey. By construction and according to the survey's instruction both numbers should match, as one is a copy of the other. We compared two instances and concluded that these numbers do not match. The precision of this expenditure category affects two additional variables: non-durable goods and total consumption expenditures, calculation of which relies on the reported expenditures on durable goods. For our analysis we use variables constructed based on the durable expenditure number reported in the portion of the survey devoted to "expenditure of household" (Questions B8.0).

However, to verify whether our results are sensitive to the chosen measure of expenditures on durable goods, we recalculate total expenditures on durable goods using information from Question B2.0 that asks to report any purchases of listed items in the past year and the amount paid for them. Based on a new number, we also recalculate expenditures on non-durable goods and total consumption. Then we repeat the ATT estimation for these variables. For all three expenditure categories both methods of calculation produce qualitatively similar results, therefore we do not report the results for calculated measures (results are available upon request).