

Re-Evaluating Microfinance

Evidence from Propensity Score Matching

Inna Cintina

Inessa Love



WORLD BANK GROUP

Development Economics Vice Presidency

Operations and Strategy Team

April 2017

Abstract

This paper evaluates effectiveness of microfinance using Propensity Score Matching (PSM) method applied to data collected in a recent randomized control trial. This method allows one to answer an additional set of questions not answered by the original study and provide more nuanced evidence by comparing Microfinance Institution (MFI) borrowers to those without any loans and those with prior

loans from other sources. It is argued that this unique setting with two comparison groups allows us to shed light on the unobservable entrepreneurial spirit bias and provides upper and lower bounds on the true microfinance impact. The results suggest that microfinance can make a modest difference for some households in several expenditure categories.

This paper is a product of the Operations and Strategy Team, Development Economics Vice Presidency. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at inna.cintina@lewin.com.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Re-evaluating Microfinance: Evidence from Propensity Score Matching

Inna Cintina and Inessa Love

JEL Classification codes: O12, O16, G21

Keywords: microfinance, poverty, propensity score matching.

Inna Cintina is a Senior Economist with The Lewin Group (inna.cintina@lewin.com); Inessa Love (contact author) is Professor at the University of Hawaii at Manoa; her email address is ilove@hawaii.edu, 808-956-7653. 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 or The Lewin Group.

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 overindebtedness, leading to illegal organ sales and suicides in extreme cases.¹ The research evidence also appears rather mixed. For example, Banerjee et al. (2015a) summarize the results of six randomized evaluations of microfinance and report “lack of evidence of transformative effects on the average borrower” while Bruhn and Love (2014) find significantly positive effects of introduction of microfinance-like product on income and labor market outcomes.

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 do not. 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, and the take-up is usually quite low, the outcomes are compared across areas (i.e., a treated area is compared with a nontreated 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 those 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 used. Thus, the RCT cannot 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 ATE). Depending on the policy question, different parameters will be of interest to policymakers. For example, if policymakers are interested in evaluating the average impact of credit introduction on the area as a whole, the ITT is the appropriate estimate. However, if policymakers want to know the impact of credit on individual borrowers, the ATE is the appropriate estimate.⁴ The latter estimates can be produced by the Propensity Score Matching (PSM) method, which is employed in this paper.

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 a 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 a loan by applying for one. Therefore, their results cannot be generalized to an average person. For further discussion see Banerjee et al. (2015a).

3. See Attanasio et al. (2015), Augsburg et al. (2015), Banerjee et al. (2015b), Beaman et al. (2014), Angelucci, Karlan, and Zinman (2015), Crepon et al. (2015), Tarozzi et al. (2015).

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

The PSM method creates a statistical comparison group of individuals without microfinance loans that has similar observable characteristics to the individuals with microfinance loans. While controlling on observables will reduce many of the significant differences between participants and nonparticipants, it cannot address the differences in unobservable characteristics such as the entrepreneurial spirit or “spunk” of the borrower. It is likely that such latent factors will affect the selection of people to obtain an MFI loan and the outcomes of interest such as poverty status, which will bias the results. 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 in the next section, 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). Most importantly, we argue that the unique set-up of our study that uses two different comparison groups (i.e., comparing MFI borrowers to those without any other loans and to those with other types of loans) allows us to evaluate the magnitude of the bias due to unobservable entrepreneurial spirit. We believe that, in this setting, PSM is 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 nonborrowers.

Specifically, we apply the PSM method on the data collected by a recent randomized experiment by Banerjee et al. (2015b). Our study addresses three main questions. First, what are the main characteristics of microfinance borrowers relative to nonborrowers versus those borrowing from other sources (such as family and friends or moneylenders)? 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). While MFI borrowers’ characteristics have been analyzed before (e.g., Crepon et al. 2015), to our knowledge this paper is the first to compare MFI borrowers to two distinct comparison groups: those who have no other loans versus those who borrow from other sources.

Second, what is the impact of microfinance on consumption and expenditures of average borrowers relative to nonborrowers, 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 versus those with other types of loans). To our knowledge, such comparison has not been done in the previous literature.

Third, 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 arising due to unobservable entrepreneurial spirit (or talent/motivation/spunk, etc.). We argue that the bias due to unobservable entrepreneurial spirit is likely to be positive in comparing MFI borrowers to those without any other loans (since relative to this group, MFI borrowers are likely to be more entrepreneurial), but is likely to be negative while comparing new MFI borrowers to those borrowing from other sources (since those households were arguably more entrepreneurial

increase power researchers often choose nonrandom sampling referred to as *convenience samples*. This further clouds the interpretation of the results. For example, according to Banerjee et al. (2015), 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 that take-up the credit when offered (i.e., those whose behavior is changed by the instrument, in this case increasing the availability of credit). In the case of heterogeneous impacts, the LATE estimator is not equal to the parameter of interest (i.e., ATE [see Ravallion, 2008]).

by taking advantage of various sources of credit that were available in a village prior to microfinance introduction). 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) and hence provide valuable information on the magnitude of entrepreneurial spirit bias for future studies of microfinance. Finally, we use data on self-reported loan purpose to identify entrepreneurial borrowers (those who borrow for new business or to repay old business loans) and nonentrepreneurial borrowers. We perform sample splits and compare expenditures among more entrepreneurial borrowers to nonentrepreneurial borrowers.

We have four main results. First, we present a profile of MFI borrowers, who are more likely to be middle-aged, have low education, be relatively poor (i.e., they have overcrowded living conditions), and have prior MFI experience. The characteristics of MFI borrowers are mostly similar when compared to households without any loans versus those with other types of loans.

Second, we find a significant increase in many of the expenditure categories: increased durable purchases, home repairs, festivals, and temptation goods. However, the categories of expenditure that are significantly increased represent a relatively minor share of total expenditures. Thus, food and nondurable expenditures, which are the largest shares of the total expenditure, show no significant changes. This explains why we do not find a significant increase in the total expenditure either.

Third, our results suggest that the entrepreneurial spirit bias is likely to be relatively small because the differences between our two comparison groups are relatively small. In addition, we find increased ‘nonproductive’ expenditures (such as festivals, temptation goods, and home repairs), which are unlikely to be subject to the entrepreneurial spirit bias. These expenditures, while improving utility in the short term, are unlikely to lead to any significant longterm transformation.

Fourth, we find that durable expenditures increase significantly more for entrepreneurial borrowers versus nonentrepreneurial borrowers. However, the temptation goods results are mixed: relative to no loan group entrepreneurial and nonentrepreneurial MFI borrowers increase temptation goods, while relative to other loans group only nonentrepreneurial group increases temptation goods purchases.

Finally, we compare the results from the PSM method to RCT results obtained on the same dataset by Banerjee et al. (2015b). Two of our main results—the increase in durables and the lack of overall increase in total expenditures and food expenditures—are the same using both methods. However, some important results are different (such as expenditures on festivals and temptation goods), and we discuss potential reasons for these differences. We conclude with the discussion of the potential merits of the PSM method and the caveats that apply.

Some earlier papers used other nonexperimental 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 nonexperimental microfinance evaluation literature.

The PSM method has successfully been used in many different settings.⁵ To the best of our knowledge, only Floro and Swain (2012) have previously used the PSM method to study the

5. 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)

impact of microcredit on individual level. Due to the more extensive dataset and a unique setting, our paper offers a number of significant contributions relative to Floro and Swain (2012).⁶

The rest of the paper is organized as follows. Section I discusses PSM methodology, section II describes our data, section III presents our results, section IV contains a discussion and caveats, and section V concludes.

I. 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. 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 .⁷ 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 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).

A number of studies have established that PSM can provide fairly accurate estimates under certain conditions.⁸ They find that propensity score matching performs well if three conditions are met: (i) using a rich set of control variables; (ii) using the same survey instrument for treated and controls; and (iii) comparing participants and nonparticipants from the same local market.

Our data satisfies all three of these conditions. First, we have a rich set of control variables. As we describe below, the data come from 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–

study the impact of agricultural extension program in Peru, Chen et al. (2009) study the effects of the World Bank-financed Southwest China Poverty Reduction Project.

6. 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. Specifically, Floro and Swain’s (2012) control sample includes only 51 observations for nonparticipants, relative to nearly 700 participants. This implies that, on average, the same control observation has to be matched to nearly 14 treatment observations. In our sample, the treatment and control groups are much more balanced, which suggests likely smaller bias and variance. 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 methods, evaluate the presence of spillovers, and the magnitude of entrepreneurial bias. Finally, Floro and Swain (2012) study the impact of bank-connected Self Help Groups (SHG) rather than loans from a specialized microfinance institution.

7. 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. Rosenbaum and Rubin (1983) show that, under the two main assumptions: (i) conditional independence; and (ii) presence of a common support matching on $P(X)$ is as good as matching on X .

8. Heckman et al. (1997, 1998a, 1998b), Dehejia and Wahba (1999, 2002), Smith and Todd (2005), and Dehejia (2005) analyze performance of various matching schemes relative to experimental estimators.

59), the male head of household, household composition, and dwelling. Thus, we believe that our dataset contains ample covariates.⁹ There is very little previous work done on the selection into microfinance as most of the prior studies are more concerned with microfinance impact.¹⁰ Second, the same survey instrument was used for participants and control group. Third, participants and control group come from the same local markets. To satisfy this requirement, we only use slums in which microfinance was introduced and compare microfinance users to nonusers. 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, several studies argued that such biases are likely to be small.¹¹ Unfortunately, most of the previous studies were done on the labor markets, and the biases could be different in the context of microfinance. However, our unique setting allows us to use two comparison groups: those with no other loans and those with other types of loans. As we argue in the introduction, these groups are likely to have opposite biases due to omitted entrepreneurial spirit. Comparing the estimates for these two groups allows us to place upper and lower bounds on true estimates and shed light on the potential magnitude of entrepreneurial spirit bias. Consistent with previous studies, our results show that such bias is likely to be small.

One of the advantages of the PSM is its semiparametric nature, which imposes 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. We compare PSM performance with results obtained from several naïve regressions: (i) naïve regression run on the full sample; (ii) naïve regression run within common support; and (iii) naïve regression with inverse propensity score weighting. Our main results are produced using PSM methodology with nearest neighbor matching with replacement. This weighting minimizes bias because control units that are the closest to treatment units in terms of propensity scores can be used multiple times (see Dehejia and Wahba 2002).¹² The standard errors for PSM estimates are calculated using bootstrap simulation with 1,000 repetitions, which takes into account the fact that propensity scores are estimated.

II. DATA

Our data come from the randomized experiment of Banerjee et al. (2015b) and are described in more detail in their paper.¹³ Here we provide only a brief description. In 2005, 52 of 104 poor

9. Ideally, the control variables are observed preprogram. Unfortunately, we do not have preprogram data. Therefore, we are careful in selecting control variables that are unlikely to be affected by the program. In an earlier version, we have included a richer set of control variables with similar results.

10. Crepon et al. (2015) also estimate 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 nonparticipants as we do here.

11. 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). Glazer et al. (2003) find that bias of nonexperimental 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. Diaz and Handa (2006) argue that, in cases when the outcomes are measured using comparable surveys, the bias arising from PSM is negligible.

12. For sensitivity analysis, we also considered Kernel matching and Stratification matching which produce qualitatively similar results to the nearest neighbor method presented here (available on request).

13. We thank Esther Duflo for making the data generously available on her website.

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. 2015b). For our main analysis, we use data from the first wave of the household surveys, conducted about 15–18 months after Spandana openings in 52 neighborhoods where Spandana was opened to make sure our participants and nonparticipants come from the same local markets (which has been noted to improve PSM performance, as discussed in the previous section). 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. (2015b) as monthly adult-equivalent expenditures, adjusted for inflation. Because the distributions of expenditures in rupees have significantly long right tails, we use log 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 nondurable 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, and home repairs (the questionnaire only asked to report home repairs above 500 Rs).¹⁴

Table 1, panel B reports shares of expenditure categories as a percent of total expenditures. We first calculate shares of each category as a percent of total for each household and then report mean, median, etc., across all households. The average durable expenditures are only about 6%, while nondurable are 94%. Food is the largest category of nondurable 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 year-old range, we select the oldest eligible woman in the household and include her characteristics such as age and education. The woman’s age, education, and whether she is a head of household are clearly not affected by the MFI borrowing. We also include male education using either the head of the household (if male) or the oldest male permanently residing in the household.¹⁵ It is possible that MFI participation will affect some employment choices of female borrowers or their spouses (as also argued by Banerjee et al. 2015b). For example, self-employment can plausibly be affected by MFI. Thus, we cannot include indicators for female or male work, since these can be affected by MFI borrowings if one or both of them start their own business.

Household characteristics unlikely to be affected by MFI borrowing 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.

14. In our estimation, we use variables defined as in Banerjee et al. (2015b), but we have also done robustness tests for some alternatively defined variables and find results to be unaffected.

15. To preserve the sample size, we replace observations with missing values (i.e., those that answered “I don’t know” or refused to answer) with zero values and add dummies to capture the average impact of those with the missing education (for males and females). We do not report these dummies in regressions since their interpretation is unclear.

Importantly, the survey contains questions on whether a previous MFI loan has been repaid and the year when the household first borrowed 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.

Table 1, panel C presents summary statistics on our control variables.

We also examine the borrowing patterns of households in our sample. 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: nearly 70% of households have more than two loans. About 20% of households in our sample 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 lender loans), followed by family and friends (at 34%). Shopkeepers and chit funds are about 17% each. Also, 18% of these households have a commercial bank or other financial company loan. Those with Spandana loans or other MFI loans also have loans from all other types of borrowers. Thus, 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.

III. RESULTS

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

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

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.

Our treatment group includes all MFI borrowers (i.e., Spandana and other MFI). We refer to this group as MFI borrowers. 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.

We have two comparison groups. The first group is those without any other loans (i.e., nonborrowers). The second group is those with other types of loans (such as loans obtained from family and friends, moneylenders, shopkeepers, chit funds, and formal financial institutions). The first comparison group allows us to estimate the upper bound of the impact because those without any other loans are likely to be less entrepreneurial than those with MFI loans (and hence the estimates on the impact are likely to be biased upwards). The second comparison group allows us

to place a lower bound on the impact estimates because those who borrow from other sources are likely to be more entrepreneurial than those who borrow from the MFI's.

Figure 1 presents the density of propensity scores estimated for our two models. 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 visibly higher than the maximum density of the control group. We also report pseudo R square, chi-square statistics, and the area under ROC curve in table 2.¹⁶ These graphs and statistics provide a check on the ability of our model to predict the likelihood of using Spandana or other MFI versus our two comparison groups (i.e., no loan and other loan).

Figure 1 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. 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.¹⁷ All the variables in our final model satisfy the balancing property. Additionally, we report t-test between treatment and comparison groups before and after matching and calculate a standardized bias metric to assess whether the differences are eliminated. Both metrics for all considered models indicate balance. The standardized biases are relatively large for some variables before matching and relatively small (in some cases close to zero) after matching, indicating substantial reduction in bias (see appendix table A1).

Table 2 reports the results of propensity to borrow regressions: column 1 reports results of the selection model comparing MFI borrowers to those without any loans and column 2 compares MFI borrowers to those with other types of loans.

We find that MFI borrowers are more likely to be middle-aged (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).¹⁸

In terms of households' characteristics, we find that, surprisingly, the number of qualifying females and the size of the household are not statistically significantly associated with MFI borrowing. The number of young children is not significant, but the presence of dependents is significant in MFI versus other loans model. 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 both specifications.

Importantly, we also 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. Similarly, the number of pre-2006 businesses also has a strong positive effect. This suggests that those with prior familiarity

16. The ROC stands for Receiver Operating Characteristic analysis. The greater the predictive power of the model, the more bowed the curve, and hence the area beneath the curve is often used as a measure of the predictive power. A perfect model has area 1. Table 2 reports ROC of 0.71 and 0.64 in two models.

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

18. There are some differences in education impact for two comparison groups. For example, female education is a significant predictor of borrowing from MFI versus 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 versus no loan, while only males in the high education category are less likely to borrow from MFI versus other types of loans. Despite these differences, the overall picture is that MFI borrowers have relatively low education.

with microfinance and those with entrepreneurial experience are more likely to borrow from this source.

Overall, our results show that a number of variables are able to significantly predict MFI borrowing and help differentiate between MFI versus no loan and MFI versus 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, be relatively poor, and have prior experience with MFI and entrepreneurship. The characteristics of MFI borrowers are mostly similar when compared to those without any loans versus those with other types of loans. The fact that MFI borrowers appear to be *a priori* poorer implies that our outcome results (such as higher expenditures) are not likely to be attributed to pre-existing differences in poverty levels.

The Impact of MFI Borrowing

In this section, we turn to our second question, specifically the differences in the impact of microfinance borrowing on household expenditures. As before, we run two models: (i) MFI borrowers versus no loans; and (ii) MFI borrowers versus other types of loans.

We report results for four models: naïve regression, naïve regression with common support, naïve regression with inverse propensity score weighting, and PSM model using nearest neighbor matching with replacement.

Table 3 reports the results for average treatment effects for comparing MFI borrowers to those without any other loans. We find the following significant results: increase in home repairs, increase in durable goods purchases, increase in health expenditures, and increase in temptation goods and festivals. The magnitude of the effects is similar across three naïve regressions; however, the PSM model produces the most conservative estimates. Because of the advantages offered by matching over the naïve regressions, we use PSM as the more reliable estimates. The education expenditures are generally not significant. The naïve models indicate an increase in the nondurable expenditures and food expenditures, but the overall effect is small and not statistically significant in the PSM model.

Interestingly, the total expenditures results are mixed. While all estimates being positive, only naïve regressions (1)–(3) are statistically significant. In contrast, matching model produces an estimate half the size with relatively large standard error. The matching method provides the closest estimate to the RCT results on the overall expenditures, which also found an insignificant change.

At a first glance, the lack of increase in total expenditures may seem surprising given that several of the subcategories 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 nondurables and food expenditures, which show no significant differences, represent a bulk of the total (see table 1, panel B).¹⁹ Thus, averages for durables are only about 6% of total expenditures, health —8%, and temptation goods—5%. The medians for these categories are even lower (2% for durables, 5% for health, and 3% for temptation goods).

19. 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).

Next, we compare MFI borrowers to those with other types of loans and report results in table 4. The three categories that show most significant and robust results are durables, home repairs, and temptation goods (all are significant at 1% and are of relatively large magnitude). Expenditures on festivals also show significant increase but of smaller magnitude. Interestingly, health expenditures are negative in all estimations, but only significant in naïve models. The total expenditures are not significant, as are nondurables 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 small durable goods, like an appliance (e.g., sewing machine, refrigerator, 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 versus nonborrowers.²⁰

The magnitudes of the effects appear large. We base our discussion on PSM estimates that we believe provide more conservative and better controlled estimates for results. Since our outcome variables are in log form, the estimated coefficients show a percentage increase in the variable. Thus, we find that durable goods purchases increase on average by 42% compared to those without loans and 41% compared to those with informal loans. The home repairs are increased by 90% in a group of MFI borrowers compared with nonborrowers, but only increased by about 50% compared to those with other informal loans. Festivals expenditures are increased by about 35% comparing to those without loans but only by about 20% 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 30%.

Two points are worth noting here. First, the increase in magnitude is larger in comparing MFI borrowers to nonborrowers versus those borrowing from other sources. This is consistent with our argument that the two estimates (i.e., group with no loans and group with other loans) can be interpreted as the lower and upper bounds on the true effect because of the omitted variable bias due to unobservable entrepreneurial spirit or “spunk.” Second, while these appear as large numbers, recall that the categories increased are a relatively small percentage of total expenditures (e.g., durables, temptation goods).

Finally, we compare the magnitude of expenditure increase produced by our estimates to the average loan size. Appendix table A2 shows the average loan amounts for different categories of borrowers. For example, the median (mean) amount of outstanding Spandana or other MFI loan is Rs 10,000 (9,759). The amounts vary slightly with the stated loan purpose but are generally in the same range. It is clear that the MFI loan amounts are constrained by the maximum loan ceiling of Rs 10,000, which was the institutional constraint from Spandana. Thus, the non-MFI loan amounts are generally larger than MFI loans for various categories of expenditures. The difference is the largest for starting a new business: while the MFI loans are in the Rs 10,000 ballpark, the non-MFI loans are, on average, Rs 41,000 and median of Rs 18,000.

Next, we use the estimates from the PSM model and calculate the average change in annual expenditures implied by our estimates. Table A3 shows that the total increase in expenditures among the statistically significant categories adds up to about Rs 5,250. The difference with the average loan amounts is likely due to the other categories of expenditures that are not statistically

20. This could indicate reverse causality (i.e., those with a health crisis may turn to the MFIs for credit).

significant in our estimates. Thus, the total increase in expenditures is in line with the average loan amounts and serves as a useful validity check.

IV. DISCUSSION AND SAMPLE SPLITS

Relative to the original Banerjee et al. (2015b) paper, our results are mixed. The two of the most important results of Banerjee et al. (2015b)—the increase in durable purchases and the lack of increase in total expenditures—are confirmed in our paper. This shows that, despite different methodologies, these two results prove to be very robust. However, there are some important differences. Contrary to Banerjee et al. (2015b), 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.

There could be three main reasons for the differences in our estimates. The first reason could be the presence of spillovers (i.e., the borrowers increasing their spending while nonborrowers reduce it. It is possible that borrowers are more likely to pick up the tab for festivals, temptation goods, health expenditures, and other categories of spending, thus resulting in negative spillovers (i.e., nonborrowers spending less on these categories). This is at least consistent with anecdotal evidence that new microfinance borrowers often have to share the “windfall” with their families and friends. Such spillovers would render the ITT estimates insignificant because they are averaged over the whole population in an area, and, hence, any increases in such expenditures among the borrowers are masked by a corresponding decrease among the nonborrowers.

The second reason could be differences in methodology: we compare actual MFI borrowers to nonborrowers or borrowers from other source in the same areas while Banerjee et al. (2015b) compare the average for the treated area (i.e., average over all borrowers and nonborrowers) with the average for the control area. In other words, we estimate the individual ATE while Banerjee et al. (2015b) estimate the ITT (as discussed in the introduction). In the presence of heterogeneous impacts these estimates will differ.

The third reason could be due to the omitted variable bias due to unobservables. Recall that PSM allows us to condition only on observable variables, and in the presence of important unobservables such as entrepreneurial spirit or “spunk,” the PSM results are going to be biased. While the entrepreneurial spirit cannot be observed directly, our methodology can shed some light on its magnitude. It is plausible that those with higher entrepreneurial spirit or “spunk” 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 households even prior to Spandana or other MFIs entering the area (as documented by Banerjee et al. 2015b). 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) can actually be negative relative to those who already borrow from other sources. Thus, our estimates can be seen as straddling the upward biased results relative to those without loans and downward biased results relative to those already borrowing from other sources.

It is important to note that the entrepreneurial spirit bias is more likely to be present in some categories of expenditures and not others. For example, purchase of durables that could be used for a small business is likely to be more affected by this bias than expenditures on festivals or

temptation goods.²¹ Our estimates on durables are around 40% for both groups (i.e., no loan and other loan groups), which suggest that the omitted variable bias is not very large. Alternatively, the two comparison groups may not actually capture the differences in entrepreneurial spirit. We investigate this further with our sample splits below. The results on increased festivals or temptation goods expenditures suggest 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 and in line with some previous literature. 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.²² Thus, it appears plausible that at least some of the poor, when given a chance for extra new credit such as 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 do not have enough data to test this hypothesis).

Finally, there could be other unobservables besides entrepreneurial spirit that would render PSM results invalid. For example, Anderson and Baland (2002) find that, in Kenya, married women participate in ROSCAs as a strategy to shelter savings against claims by their husbands for immediate consumption. Such commitment problems may also lead some women to participate in a group microfinance product, and it is exactly the households with commitment problems that would spend more on temptation goods. Others without that problem either borrow from other sources if they are entrepreneurial or do not if they are not. Unfortunately, we do not have the data to test the commitment hypothesis directly. However, the existing data allows us to take a deeper look at the relationship between entrepreneurial spirit and expenditures.

Specifically, the questionnaire asks borrowers from any source to state the intended use for a loan. Two particular uses stand out as possibly capturing those borrowers who are more likely to be entrepreneurial: borrowing for a new business and repayment of old business loans. We use such self-disclosed classification to split our sample on the entrepreneurial borrowers (i.e., those that state the loan purpose is either new business or repayment of business loans) and nonentrepreneurial borrowers (those with loans for other purposes). The sample sizes are as follows: 174 entrepreneurial MFI, 76 entrepreneurial other loans, 772 nonentrepreneurial MFI, 1,902 nonentrepreneurial other loans.

In table 5, we present our sample splits results. In column 1, we present PSM estimates for entrepreneurial MFI borrowers compared to those without any loans; in column 2, we compare nonentrepreneurial MFI borrowers to those without any loans; in column 3, we compare entrepreneurial MFI borrowers to entrepreneurial other loans borrowers, and in column 4, we compare nonentrepreneurial MFI borrowers to nonentrepreneurial other loans borrowers. The results show that expenditures on durable goods are twice as high in the entrepreneurial groups versus nonentrepreneurial groups. Specifically, the effects are much larger in column 1 relative to column 2 (56% for entrepreneurial MFI versus 32% for nonentrepreneurial MFI compared with no loans group) and, in column 3 relative to column 4 (56% for entrepreneurial MFI and 28% for

21. Of course it could also be used to purchase “unproductive” durables such as TV, DVD player, etc.

22. 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” or 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.

nonentrepreneurial MFI compared with other loans). These results make sense since entrepreneurial MFI borrowers are more likely to invest in durable goods than nonentrepreneurial borrowers.

Looking at the temptation goods results, the results are mixed. Thus, comparing MFI borrowers to those without any loans, we still find large and significant expenditures on temptation goods in both groups (columns 1 and 2). However, comparing MFI borrowers to those with other loans, the temptation goods expenditures are close to zero and not significant in the entrepreneurial group (3% column 3), but about ten times as large and significant in the nonentrepreneurial group (39% column 4). The mixed results could be due to the volatile nature of temptation goods data, but, at a minimum, they show that relative to those with other loans, MFI does not increase temptation goods expenditures for entrepreneurial borrowers. Despite small sizes of entrepreneurial borrowers, these results are at least suggestive of the possibility that temptation and commitment problems are more present among the nonentrepreneurial borrowers. While these results do not represent conclusive evidence, such sample splits demonstrate that the PSM method has an edge in unpacking the heterogeneous impacts for various groups of borrowers, which is not possible with the intention to treat RCT designs that estimate the average for the treated versus nontreated areas.

V. CONCLUSIONS

We employ the 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. 2015b), 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 an individual level, which is not possible using RCT designs that can only produce either ITT or LATE estimates (see footnote 3). While such parameters can be of interest in answering some policy questions (such as estimating the impact of introduction of microfinance on the total expenditures in the area), in other cases the policymakers would 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 any loans and those who borrow from other sources. This allows us to evaluate the extent of the omitted variable bias due to entrepreneurial spirit and place upper and lower bounds on the true impacts. Thus, our paper serves as a complement to the recent emergence of RCT papers (cited in footnote 2) and shows how using PSM methodology can be useful in answering questions that RCTs are unable to address. Specifically, the PSM has an edge over the RCT in investigating the heterogeneous impacts, such as the entrepreneurial ability samples splits that we perform in this paper.

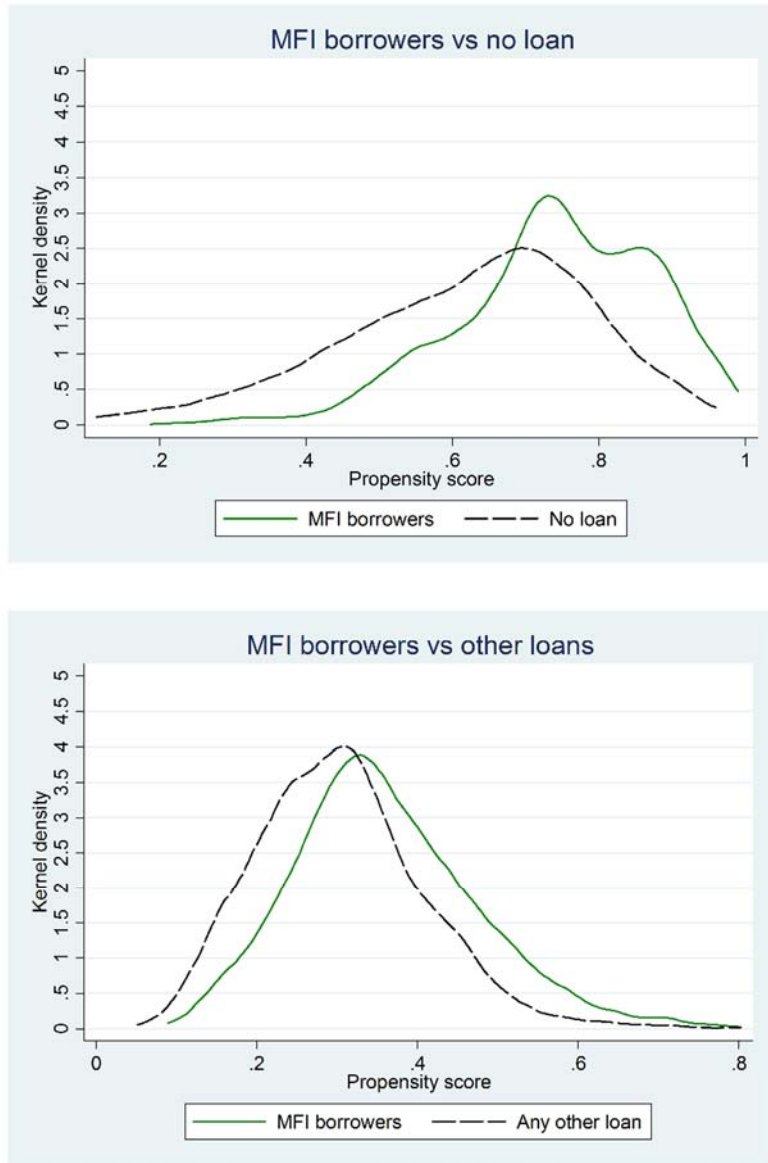
Finally, we believe that while the PSM method can be useful for investigating the impact of microfinance on poverty, the results should be treated with caution. Unobservables, such as entrepreneurial spirit and commitment problems, should be explicitly considered in the research design. Shedding more light on the extent of these biases is a fruitful avenue for future research.

REFERENCES

- Anderson, S., and J. Baland. 2002. The Economics of Roscas and Intrahousehold Resource Allocation, *The Quarterly Journal of Economics* 117 (3): 963–95.
- Angelucci, M., D. Karlan, and J. Zinman. 2015. “Microcredit Impacts: Evidence from a Randomized Microcredit Program Placement Experiment by Compartamos Banco.” *American Economic Journal: Applied Economics* 7 (1): 151–82.
- Attanasio, O., B. Augsburg, R. De Haas, E. Fitzsimons, and H. Harmgart. 2015 “The Impacts of Microfinance: Evidence from Joint-Liability Lending in Mongolia.” *American Economic Journal: Applied Economics* 7 (1), 90–122.
- Augsburg, B., R. De Haas, H. Harmgart, and C. Meghir. 2015. “The Impacts of Microcredit: Evidence from Bosnia and Herzegovina.” *American Economic Journal: Applied Economics* 7 (1), 183–203.
- Banerjee, A., D. Karlan, and J. Zinman. 2015a. "Six Randomized Evaluations of Microcredit: Introduction and Further Steps." *American Economic Journal: Applied Economics* 7 (1): 1–21.
- Banerjee, A., E. Duflo, R. Glennerster, and C. Kinnan. 2015b, “The Miracle of Microfinance? Evidence from a Randomized Evaluation.” *American Economic Journal: Applied Economics* 7 (1): 22–53.
- Banerjee, A. and E. Duflo. 2011, “Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty.” *Public Affairs*, New York, NY.
- Beaman, L., D. Karlan, B. Thuysbaert, and C. Udry, 2014, “Self-Selection into Credit Markets: Evidence from Agriculture in Mali.” Mimeo. Yale University.
- Bruhn, M. and I. Love. 2014, “The Real Impact of Improved Access to Finance: Evidence from Mexico.” *The Journal of Finance* 69 (3): 1347–76.
- Caliendo, M. and S. Kopeinig, 2008, “Some Practical Guidance for the Implementation of Propensity Score Matching,” *Journal of Economic Surveys* 22 (1): 31–72.
- Chen, S., R. Mu, and M. Ravallion. 2009. “Are There Lasting Impacts of Aid to Poor Areas?,” *Journal of Public Economics* 93 (3–4): 512–28.
- Crépon, B., F. Devoto, E. Duflo, and W. Pariente. 2015, “Estimating the Impact of Microcredit on Those who Take It Up: Evidence from a Randomized Experiment in Morocco,” *American Economic Journal: Applied Economics* 7 (1): 90–122.
- Dehejia, R.H., 2005, “Practical Propensity Score Matching: A Reply to Smith and Todd.” *Journal of Econometrics* 125 (1): 355–64.
- Dehejia, R.H., and S. Wahba, 1999, “Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs.” *Journal of the American Statistical Association* 94 (448): 1053–62.
- Dehejia, R.H. and S. Wahba, 2002, “Propensity Score Matching Methods for Nonexperimental Causal Studies.” *The Review of Economics and Statistics* 84 (1): 151–61.
- Diaz J.J., and S. Handa. 2006. “An Assessment of Propensity Score Matching as a Nonexperimental Impact Estimator: Evidence from Mexico's PROGRESA Program.” *Journal of Human Resources* 41 (2), 319–45.
- Duflo, E., R. Glennerster, and M. Kremer. 2008. ”Using Randomization in Development Economics Research: A Toolkit.” *Handbook of Development Economics*, Elsevier.

- Floro, M., and R.B. Swain. 2012. "Assessing the Effect of Microfinance on Vulnerability and Poverty among Low Income Households." *The Journal of Development Studies* 48 (5): 605–18.
- Glazerman, S., D. Levy, and D. Myers. 2003. "Non-Experimental versus Experimental Estimates of Earnings Impacts." *Annals of the American Academy of Political and Social Sciences* 589: 63–93.
- Godtland, E., E. Sadoulet, A. De Janvry, R. Murgai, and O. 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.
- Heckman, J., H. Ichimura, and P. Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program." *Review of Economic Studies* 64: 605–54.
- . 1998a. "Matching as an Econometric Evaluation Estimator." *Review of Economic Studies* 65: 261–94.
- Heckman, J., H. Ichimura, J. Smith, and P. Todd. 1998b. "Characterizing Selection Bias Using Experimental Data." *Econometrica* 66: 1017–89.
- Jalan, J., and M. Ravallion. 2003a. "Does Piped Water Reduce Diarrhea for Children in Rural India?" *Journal of Econometrics* 112: 153–73.
- . 2003b. "Estimating Benefit Incidence for an Anti-poverty Program using Propensity Score Matching," *Journal of Business and Economic Statistics* 21 (1): 19–30.
- Karlan, D., and J. Zinman. 2010. "Expanding Credit Access: Using Randomized Supply Decisions to Estimate the Impacts." *Review of Financial Studies* 23 (1): 433–64.
- . 2011. "Microcredit in Theory and Practice: Using Randomized Credit Scoring for Impact Evaluation." *Science* 332 (6035) (June 10): 1278–84.
- Khandker, S.R. 2006. "Microfinance and Poverty: Evidence Using Panel Data from Bangladesh." *World Bank Economic Review* 19 (2): 263–86.
- Pitt, M.M. 2014. "Response to 'The Impact of Microcredit on the Poor in Bangladesh: Revisiting the Evidence'." *Journal of Development Studies* 50 (4): 605–10.
- Pitt, M., and S. 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, M.. 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, M.. 2008. "Evaluating Anti-poverty Programs." In T. P. Schultz and J. Strauss, eds. *Handbook of Development Economics* 4: 3787–846. Amsterdam: North-Holland.
- Roodman, D. and J. Morduch. 2014. "The Impact of Microcredit on the Poor in Bangladesh: Revisiting the Evidence." *Journal of Development Studies* 50 (4): 583–604.
- Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The central role of the propensity score in observational studies for causal effects." *Biometrika*: 41-55.
- Smith, J. and P. Todd. 2005. "Does Matching Overcome LaLonde's Critique of Non-experimental Estimators?" *Journal of Econometrics* 25: 305–53.
- Tarozzi, A., D. Jaikishan, and K. Johnson. 2015. "The Impacts of Microcredit: Evidence from Ethiopia." *American Economic Journal: Applied Economics* 7 (1): 54–89.

Figure 1. Propensity score by outcome



Note: We also considered Spandana borrowers only vs. no loan and the propensity score distribution looks similar to MFI borrowers vs. no loan graph.

Source: Authors' analysis based on data described in the text.

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=3,308)	7.10	0.54	4.22	10.10
Log non-durable expenditures, Rs2007 (N=3,284)	7.03	0.50	4.22	10.10
Log durable expenditures, Rs2007 (N=3,284)	3.11	1.99	0.00	9.29
Log food items, Rs2007 (N=3,249)	6.12	0.39	4.52	9.28
Log temptation goods, Rs2007 (N=3,258)	2.96	2.05	0.00	7.83
Log festivals, Rs2007 (N=3,303)	4.35	1.70	0.00	10.13
Log health, Rs2007 (N=3,299)	3.96	1.37	0.00	9.02
Log education (total), Rs2007 (N=3,275)	3.51	2.30	0.00	8.88
Log education (fees), Rs2007 (N=3,249)	2.59	2.33	0.00	8.83
Log home repairs >500, Rs2007 (N=2,700)	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).

Source: Authors' analysis based on data described in the text.

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

Variable	Obs	Mean	Std. Dev.	Min	Max	Median
Non-durable expenditures, Rs2007	3,284	0.94	0.11	0.02	1.00	0.98
Durable expenditures, Rs2007	3,284	0.06	0.11	0.00	0.98	0.02
Food items, Rs2007	3,249	0.39	0.12	0.02	0.96	0.39
Temptation goods, Rs2007	3,258	0.05	0.07	0.00	0.56	0.03
Festivals, Rs2007	3,301	0.22	0.70	0.00	20.99	0.08
Health, Rs2007	3,299	0.08	0.11	0.00	0.94	0.05
Education (total), Rs2007	3,275	0.10	0.11	0.00	0.92	0.07
Education (fees), Rs2007	3,249	0.05	0.08	0.00	0.87	0.03
Home repairs >500, Rs2007	2,698	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 (2015b).

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

Variable	Mean	Std. Dev.	Min	Max
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
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
Repaid MFI loan before 2006	0.07	0.26	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
Number of pre-2006 businesses (opened 1 year or more before endline 1)	0.39	0.65	0	5

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).

Source: Authors' analysis based on data described in the text.

Table 2. Propensity score estimation

	MFI borrowers vs. no loan	MFI borrowers vs. other loans
	(1)	(2)
Head of Household is a woman	0.267* (0.16)	-0.003 (0.09)
Age of oldest qualifying female	0.090*** (0.03)	0.076*** (0.02)
Age of oldest qualifying female squared	-0.001*** (0.00)	-0.001*** (0.00)
Female: education between 5 and 10 standard or more	-0.415** (0.21)	0.052 (0.11)
Female: education is 10 standard or more	-0.899*** (0.26)	-0.063 (0.16)
Male: education between 5 and 10 standard or more	-0.364** (0.17)	-0.081 (0.10)
Male: education is 10 standard or more	-0.876*** (0.20)	-0.551*** (0.13)
Persons per room > 2	0.357*** (0.09)	0.195*** (0.06)
There are dependents	0.112 (0.10)	0.209*** (0.07)
Have children age 0-2	-0.068 (0.10)	0.038 (0.07)
Repaid MFI loan before 2006	0.491*** (0.16)	0.290*** (0.09)
Number of qualifying women permanently residing in hhd is 2+	-0.048 (0.09)	0.056 (0.06)
Number of people in the hhd is between 1 and 4	-0.132 (0.13)	-0.043 (0.08)
Number of people in the hhd is 5-6	-0.036 (0.10)	-0.040 (0.07)
Number of pre-2006 businesses	0.407*** (0.06)	0.306*** (0.04)
Constant	-0.486 (0.70)	-2.009*** (0.46)
Observations	1,319	2,894
Pseudo R2	0.1011	0.0461
Chi-square statistics (p-value)	161.40 (0.000)	167.56 (0.000)
Area under ROC curve	0.7100	0.6446

Note: *** p<0.01, ** p<0.05, * p<0.1.

Source: Authors' analysis based on data described in the text.

Table 3. Average treatment effect on the treated: MFI borrowers vs. no loan

	Naïve OLS without common support	Naïve OLS with common support	OLS with inverse probability weights	Nearest neighbor with replacement
Log monthly durable goods	0.454*** (0.137)	0.452*** (0.137)	0.505*** (0.142)	0.420* (0.173)
Log monthly education (fees)	-0.026 (0.142)	-0.033 (0.142)	-0.042 (0.143)	-0.233 (0.199)
Log monthly education (total)	-0.024 (0.150)	-0.031 (0.151)	-0.097 (0.161)	-0.371* (0.187)
Log monthly festivals	0.561*** (0.103)	0.558*** (0.103)	0.529*** (0.103)	0.342* (0.143)
Log monthly food items	0.039* (0.021)	0.039* (0.021)	0.043* (0.022)	0.051 (0.036)
Log monthly health	0.275*** (0.099)	0.266*** (0.099)	0.266** (0.111)	0.251* (0.124)
Log monthly home repairs >Rs500	0.809*** (0.143)	0.806*** (0.143)	0.834*** (0.137)	0.900** (0.157)
Log monthly non-durable expenditures	0.067** (0.026)	0.066** (0.026)	0.052 (0.032)	0.023 (0.041)
Log monthly temptation goods	0.807*** (0.147)	0.802*** (0.147)	0.813*** (0.146)	0.614** (0.173)
Log monthly total hhd expenditures	0.095*** (0.031)	0.094*** (0.031)	0.083** (0.034)	0.045 (0.046)

Note: *** p<0.01, ** p<0.05, * p<0.1. For OLS, standard errors reported in parentheses are clustered on slums; for PSM nearest neighbor with replacement bootstrap standard errors with 1,000 repetitions are in parentheses. All variables are per capita adult equivalent. Results for Kernel matching and Stratification matching are not materially different from the nearest neighbor matching.

Source: Authors' analysis based on data described in the text.

Table 4. Average treatment effect on the treated: MFI borrowers vs. other loans

	Naïve OLS without common support	Naïve OLS with common support	OLS with inverse probability weights	Nearest neighbor with replacement
Log monthly durable goods	0.321*** (0.100)	0.322*** (0.100)	0.356*** (0.102)	0.407*** (0.113)
Log monthly education (fees)	0.064 (0.087)	0.060 (0.088)	0.054 (0.087)	0.080 (0.134)
Log monthly education (total)	-0.048 (0.088)	-0.043 (0.089)	-0.057 (0.090)	-0.030 (0.125)
Log monthly festivals	0.145** (0.064)	0.145** (0.064)	0.148** (0.063)	0.243** (0.099)
Log monthly food items	0.029* (0.015)	0.030* (0.015)	0.033** (0.015)	0.026 (0.024)
Log monthly health	-0.121* (0.065)	-0.122* (0.065)	-0.120* (0.068)	-0.074 (0.076)
Log monthly home repairs >Rs500	0.354*** (0.117)	0.359*** (0.116)	0.363*** (0.114)	0.521*** (0.140)
Log monthly non-durable expenditures	-0.013 (0.019)	-0.012 (0.019)	-0.010 (0.019)	-0.019 (0.030)
Log monthly temptation goods	0.318*** (0.106)	0.314*** (0.106)	0.291** (0.111)	0.277** (0.119)
Log monthly total hhd expenditures	-0.003 (0.017)	-0.002 (0.017)	0.003 (0.018)	-0.012 (0.031)

Note: *** p<0.01, ** p<0.05, * p<0.1. For OLS, standard errors reported in parentheses are clustered on slums; for PSM nearest neighbor with replacement bootstrap standard errors with 1,000 repetitions are in parentheses. All variables are per capita adult equivalent. Results for Kernel matching and Stratification matching are not materially different from the nearest neighbor matching.

Source: Authors' analysis based on data described in the text.

Table 5. Subgroup analysis – entrepreneurial spirit sample splits

	Entrepreneurial MFI vs. no loan	Non- Entrepreneurial MFI vs. no loan	Entrepreneurial MFI vs. entrepreneurial other loan	Non- Entrepreneurial MFI vs. non-entrepreneurial other loan
	(1)	(2)	(3)	(4)
Log monthly durable goods	0.563* (0.290)	0.325* (0.188)	0.558* (0.334)	0.279** (0.123)
Log monthly education (fees)	-0.035 (0.307)	-0.195 (0.207)	-0.303 (0.401)	-0.047 (0.142)
Log monthly education (total)	0.188 (0.318)	-0.168 (0.207)	-0.153 (0.339)	-0.035 (0.140)
Log monthly festivals	0.412* (0.236)	0.545*** (0.161)	0.289 (0.253)	0.185* (0.107)
Log monthly food items	0.119* (0.068)	0.027 (0.042)	-0.021 (0.075)	0.027 (0.024)
Log monthly health	0.237 (0.207)	0.393*** (0.148)	0.043 (0.221)	-0.083 (0.085)
Log monthly home repairs >Rs500	0.585* (0.329)	1.104*** (0.164)	0.714* (0.389)	0.600*** (0.148)
Log monthly non-durable expenditures	0.102 (0.072)	0.016 (0.041)	0.018 (0.097)	-0.022 (0.030)
Log monthly temptation goods	1.025*** (0.303)	0.740*** (0.180)	0.032 (0.377)	0.395*** (0.128)
Log monthly total hhd	0.122 (0.079)	0.064 (0.050)	0.015 (0.106)	-0.001 (0.033)

Note: *** p<0.01, ** p<0.05, * p<0.1. Reported coefficients are PSM nearest neighbor with replacement matching, bootstrap standard errors with 1,000 repetitions are in parentheses. All variables are per capita adult equivalent. Group sizes: entrepreneurial MFI=174, entrepreneurial other loans =76; non-entrepreneurial MFI =772; non-entrepreneurial other loans =1,902.

Source: Authors' analysis based on data described in the text.

Table A1. Differences between treatment and control before and after matching

Panel A: MFI borrowers vs. no loan

	BEFORE MATCHING			AFTER MATCHING		
	Mean		% Bias	Mean		% Bias
	T	C		T	C	
Head of Household is a woman	0.09	0.08	3.8	0.09	0.08	2.8
Age of oldest qualifying female	34.43	34.69	-2.8	34.32	33.97	4.0
Age of oldest qualifying female squared	1256.90	1296.10	-5.9	1249.00	1222.40	4.0
Female: education 5-10 standard	0.44	0.42	4.3	0.45	0.44	2.5
Female: education 10+ standard	0.04	0.10 ***	-24.6	0.04	0.03	2.5
Female: education is unknown	0.46	0.46	0.0	0.46	0.48	-5.2
Male: education 5-10 standard	0.51	0.49	5.3	0.52	0.53	-2.1
Male: education 10+ standard	0.07	0.17 ***	-30.2	0.07	0.06	2.5
Male: education is unknown	0.34	0.31	6.5	0.34	0.35	-2.1
Persons per room > 2	0.67	0.50 ***	34.6	0.67	0.68	-3.1
There are dependents	0.76	0.69 ***	15.6	0.76	0.78	-5.0
Have children age 0-2	0.21	0.23	-5.7	0.21	0.23	-3.6
Repaid MFI loan before 2006	0.10	0.04 ***	24.2	0.10	0.09	5.7
Number of qualifying women permanently residing in hhd is 2+	0.34	0.38	-6.8	0.34	0.35	-0.4
Number of people in the hhd is 1-4	0.26	0.38 ***	-24.2	0.27	0.28	-3.2
Number of people in the hhd is 5-6	0.44	0.38 **	12.1	0.44	0.42	3.3
Number of pre-2006 businesses	0.55	0.28 ***	40.4	0.51	0.49	2.4

Panel B: MFI borrowers vs. other loans

	BEFORE MATCHING			AFTER MATCHING		
	Mean		% Bias	Mean		% Bias
	T	C		T	C	
Head of Household is a woman	0.09	0.10	-5.1	0.09	0.09	-0.8
Age of oldest qualifying female	34.43	34.70	-3.0	34.35	34.29	0.7
Age of oldest qualifying female squared	1256.90	1285.40	-4.4	1250.90	1246.30	0.7
Female: education 5-10 standard	0.44	0.39 ***	10.9	0.44	0.44	0.8
Female: education 10+ standard	0.04	0.05	-6.1	0.04	0.03	2.0
Female: education is unknown	0.46	0.51 **	-9.8	0.45	0.46	-2.2
Male: education 5-10 standard	0.51	0.45 ***	13.5	0.52	0.51	1.2
Male: education 10+ standard	0.07	0.12 ***	-18.8	0.07	0.07	0.2
Male: education is unknown	0.34	0.37	-6.4	0.34	0.34	-1.0

Persons per room > 2	0.67	0.58 ***	18.5	0.67	0.67	0.9
There are dependents	0.76	0.68 ***	17.8	0.76	0.76	-0.9
Have children age 0-2	0.21	0.19	4.4	0.21	0.21	-1.1
Repaid MFI loan before 2006	0.10	0.06 ***	13.4	0.10	0.09	1.1
Number of qualifying women permanently residing in hhd is 2+	0.34	0.33	2.0	0.34	0.35	-1.0
Number of people in the hhd is 1-4	0.26	0.32 ***	-13.2	0.26	0.27	-2.1
Number of people in the hhd is 5-6	0.44	0.44	-0.4	0.44	0.44	0.6
Number of pre-2006 businesses	0.55	0.33 ***	31.8	0.55	0.53	3.2

Note: *** p<0.01, ** p<0.05. T denotes treated group; C denotes control group.

Source: Authors' analysis based on data described in the text.

Table A2. Loans MFI borrowers vs Borrowers from Other Sources

	MFI borrowers		Non MFI borrowers	
	Mean	Median	Mean	Median
Number of loans	4.87	4	4.77	4
Amount of loans (any purpose)	9,758.92	10,000	23,836.51	10,000
Purpose: Health	9,783.07	10,000	14,753.32	6,000
Purpose: Durables	8,256.25	8,000	8,727.89	2,000
Purpose: Home improvements	11,259.42	10,000	28,885.35	10,000
Purpose: Regular consumption	9,036.99	8,000	5,280.70	1,800
Purpose: Start a business	9,888.02	10,000	41,053.23	18,000
Purpose: Start a business or Repay old business loan	9,868.87	8,000	3,378.79	15,000

Note: The purpose of the loan identified based on a question about the “primary purpose.”

Source: Authors’ analysis based on data described in the text.

Table A3. Expenditures and corresponding change

Variable	Obs	Mean	Std. Dev.	Min	Max	Annual mean expenditures	Coef. NN	Increase in monthly expenditures	Increase in annual expenditures
Monthly total household expenditures	943	1,378.2	884.6	74	9,296	16,539			
Real Monthly non-durable expenditures	939	1,237.7	688.3	72	7,841	14,852			
Real monthly temptation goods	924	85.0	124.4	1	1,319	1,020	0.614	52.17	626.06
Real monthly food items	927	488.8	193.2	102	1,503	5,865			
Real monthly exp on durables	939	143.7	468.2	1	9,070	1,724	0.420	60.36	724.26
Real monthly exp on home repair	757	196.0	1,396.6	1	28,946	2,352	0.900	176.39	2,116.73
Real monthly exp on festivals	942	348.6	1,250.0	1	25,150	4,183	0.342	119.22	1,430.62
Real monthly exp on health	942	118.1	258.0	1	4,182	1,417	0.251	29.65	355.78
Real monthly exp on education	935	142.2	307.1	1	7,182	1,707			
Real monthly exp on education fees	925	88.4	275.0	1	6,817	1,060			
								TOTAL CHANGE	5,253.46

Note: All expenditures are per capita and expressed in Rs 2007.

Source: Authors' analysis based on data described in the text.