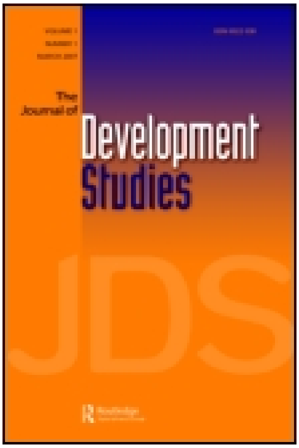


Journal of Development Studies
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/fjds20

The Benefits and Costs of Microfinance: Evidence from Bangladesh

Matthieu Chemin [a]

[a] Department of Economics, University of Quebec at Montreal, Canada

Available online: 28 Apr 2008

To cite this article: Matthieu Chemin (2008): The Benefits and Costs of Microfinance: Evidence from Bangladesh, Journal of Development Studies, 44:4, 463-484

To link to this article: http://dx.doi.org/10.1080/00220380701846735

Journal of Development Studies,
Vol. 44, No. 4, 463–484, April 2008

The Benefits and Costs of Microfinance: Evidence from Bangladesh

MATTHIEU CHEMIN
Department of Economics, University of Quebec at Montreal, Canada

Final version received September 2007

ABSTRACT *Using the latest developments from the evaluation literature, namely the technique of matching, this paper shows a positive, but lower than previously thought, effect of microfinance on expenditure per capita, supply of labour, and level of school enrolment for boys and girls. For instance, participants spend 3 per cent more on average than non-participants in control villages. This paper also takes into account repayment delays to calculate the cost of credit provision. It shows how a better investigation at the individual level of the benefits brought and the cost borne could help microfinance institutions to better select their customers.*

JEL Classification: C14, D10, G21, I38, O12, O16

I. Introduction

Advocates of microfinance support the view that it could break the vicious cycle of poverty. The Grameen Bank in Bangladesh, one of the flagship programmes, reports that:

> It is estimated that the average household income of Grameen Bank members is about 50 per cent higher than the target group in the control village, and 25 per cent higher than the target group non-members in Grameen Bank villages.[1]

This statement is, however, subject to two important criticisms concerning programme evaluation. First, the placement of the programme is non-random. Comparing Grameen Bank members to the target group in control villages would isolate the impact of microfinance as well as systematic differences between villages.[2] Second, the attribution of loans is on a voluntary basis. People self-select in microfinance. It might be that people who try to obtain a loan differ in a systematic way from people who do not attempt to get a loan. Comparing Grameen Bank members to the target group in Grameen Bank villages would isolate the impact of

Correspondence Address: Matthieu Chemin, University of Quebec at Montreal, CP 8888 Succursalle Centre Ville, Montreal, Quebec, Canada H3C 3P8. Email: chemin.matthieu@uqam.ca

ISSN 0022-0388 Print/1743-9140 Online/08/040463-22 © 2008 Taylor & Francis
DOI: 10.1080/00220380701846735

microfinance as well as systematic differences between these two groups (such as entrepreneurial skills for example). This paper uses data from a large survey collected in Bangladesh in 1991–1992, containing information on ineligible as well as eligible individuals, participants and non-participants, treatment and control villages. A matching technique is used to adjust for differences in pre-treatment characteristics between treatment and non-treatment groups by pairing each treated individual to a non-treated unit with the 'same' observable characteristics. Given the assumption that relevant differences between two groups are captured by their observable characteristics, the average outcome experienced by the matched pool of non-treated individuals identifies the counterfactual outcome the treated units would have experienced had they not been treated. This technique is employed to compare Grameen Bank members to the target group in the control village and to the target group non-members in Grameen Bank villages. Considering the absence of a randomised experiment in the data, this non-experimental technique attempts to mimic one. Matching deals with non-random programme placement as participants in treated villages are compared with the 'same' non-participant in control villages. Additionally, this paper will compare the spread of outcomes to their village average in order to net out any village effects. Matching also deals with self-selection for the same reason: participants in treated villages are compared with the 'same' non-participant in treated and control villages. Given the assumption that selection is based on observables, the 'same' individuals in control villages would have participated, had they had access to microfinance.

Microfinance has often been described as a win–win programme. Dr Muhammad Yunus, the founder of the Grameen Bank, has recently received the Nobel Peace Prize for 'efforts to create economic and social development from below'. Evaluating the benefits of microfinance is therefore a topic of first order importance. Theoretically, it is quite clear that access to credit for the poor is key to economic development. The reluctance of banks to lend to people without collateral could cause poverty traps (Banerjee and Newman, 1993). Finding ways to give loans to poor people without collateral could lift a country out of poverty. Microfinance is sometimes based on the voluntary formation of small groups to provide mutual, morally binding group guarantees in lieu of the collateral required by conventional banks.[3] This mechanism has allowed the Grameen Bank to experience repayment rates of up to 98 per cent. Grameen has been replicated worldwide and has inspired over 7,000 microfinance institutions in Latin America, Africa and Asia serving 25 million poor clients. Access to credit could increase expenditure of participants by encouraging project start-up and raising labour supply. It could also increase child enrolment in school if the opportunity cost of school decreases due to increased parental wealth.

The empirical evidence is much less clear. Using the Bangladesh data employed here, Pitt and Khandker (1998) use an identification strategy based on a discontinuity contained in the Grameen Bank's eligibility rules for programme participation (people with less than half an acre of land are eligible). The discontinuity rule evaluation compares individuals just above and below the eligibility rule. They concluded that programme participation raised household consumption by 18 Takas for every 100 Takas lent to an individual woman.

However, Morduch (1998) points out that the discontinuity rule is not respected in practice. Indeed, almost 25 per cent of participants are mistargeted in the sense that they own more than 0.5 acres of land. Morduch points to one participant possessing 13.4 acres of land. This analysis casts doubt on the Pitt and Khandker (1998) results. A situation where treatment does not depend in a deterministic way on a certain criterion such as acreage is called a fuzzy design as opposed to a sharp design in the regression discontinuity literature (Hahn *et al.*, 2001: 202). This does not preclude estimation. However, Hahn *et al.* (2001) show that results could be biased upwards if there is no discontinuity in the probability of receiving treatment at the threshold of 0.5 acres. Morduch prefers a simple difference-in-differences approach comparing eligible to ineligible in treated villages to the same difference in control villages. He finds no significant effect resulting from microfinance exposure. In his response to Morduch, Pitt (1999) noted that this difference-in-differences approach fails to deal with the key issue of programme placement. If the Grameen Bank focuses on areas where inequality between rich and poor is greatest, Morduch's estimate will be biased downwards.

In contrast to these two papers, I use the matching technique to compare outcomes of Grameen Bank members to the target group in control villages and to the target group non-members in Grameen Bank villages. This technique relies on two identification assumptions. First, the Conditional Independence Assumption (CIA) states that non-treated outcomes are what treated outcomes would have been had they not been treated conditional on a set of observables. It is not possible to test this hypothesis directly but I do use different sets of observables to provide a convincing set of observables. Second, the Common Support hypothesis states that there exists a set of individuals where all treated agents have a counterpart in the non-treated population and anyone constitutes a possible participant. This hypothesis is tested, finding that microfinance increased expenditure of participants by 3 per cent compared to expenditure of matched individuals in control villages. On the other hand, I did not find any significant differences between participants and non-participants in villages with microfinance. This indicates the presence of positive externalities of microfinance at the level of villages. I also find that microfinance does not act as a consumption smoothing mechanism, that males and females increased their labour supply, and that enrolment of boys and girls increased due to microfinance.

Given the difficulties met by previous authors on this database, the matching technique brings a useful contribution to this debate. Matching does not rely on a discontinuity that is not present in practice. Matching takes into account non-random programme placement by comparing treated individuals with the 'same' non-treated individuals in control villages. These 'same' non-treated individuals in control villages would have participated in microfinance had they had access to microfinance. Of course, matching can cover selection into the programme based on observables but not unobservables. However, Heckman *et al.* (1997a), in their seminal paper on matching, find the bias due to selection on unobservables empirically less important than other components that matching explicitly controls for.

Given the absence of randomised experiments on the impact of microfinance in the literature, this paper using the matching technique brings a contribution to the more general literature on the empirical impact of microfinance on poverty. Kaboski and Townsend (2005) evaluate the impact of participation in microfinance on economic outcomes using the presence of a microfinance institution in a specific village as an instrumental variable. However, the presence of a microfinance institution in a village might be endogenous and related to observable characteristics (average wealth in the village) or unobservable ones (entrepreneurial spirit in the village). This paper essentially replicates the same methodology by comparing individuals in villages with a microfinance institution to individuals without, but adds the refinement of matching by comparing an individual to its matched comparison. Karlan and Zinnman (2007) evaluated the impact of microfinance by using a unique credit expansion: a South African lender relaxed its risk assessment criteria by encouraging its loan officers to approve randomly selected marginal rejected applications. They found that consumer credit produced significant benefits. This paper, however, focuses on traditional 'productive' lending.

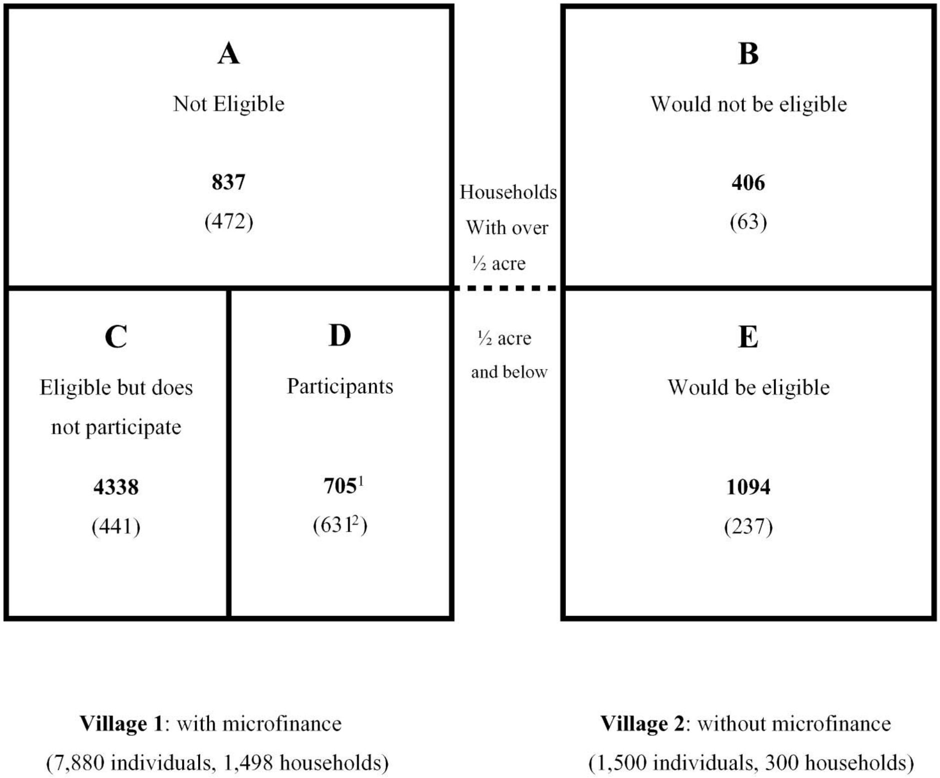
An additional contribution of the paper is a thorough analysis of the problem of late repayment as a source of cost to microfinance institutions. It is often claimed that repayment rates reach 98 per cent in some instances. This paper uses rigorous accounting standards to evaluate this claim. It is found that even after a 30-day grace period, only 71 per cent of the loans are actually repaid. This paper develops an indicator of cost at the individual level. This will permit the construction of a cost–benefit analysis at the individual level which may serve as a guide for improved customer selection. This could have important policy implications for microfinance institutions in order to select their customers.

The paper's structure is as follows. Section II will describe the data used. Section III will describe the matching technique. Section IV will present the preferred specification and results. Section V will calculate the costs due to late repayment at the individual level and compare costs and benefits at the individual level. Section VI concludes.

II. Data and Existing Studies

This paper uses a large survey collected in Bangladesh in 1991–1992 by the Bangladesh Institute for Development Studies in collaboration with the World Bank, originally used in Pitt and Khandker (1998) and subsequently in Morduch (1999). Using the same dataset allows us to compare results across methodologies. Figure 1 describes the dataset used.[4] Three microfinance programmes (Grameen Bank, BRAC, BRDB) were operating in 87 villages, but 15 of them were not covered by microfinance. In Figure 1, group A corresponds to ineligible individuals in villages having access to microfinance whereas group B corresponds to ineligible individuals in villages without microfinance. Group C corresponds to eligible individuals who choose not to participate and group D to eligible individuals who choose to participate in villages where microfinance is available. Group E includes eligible individuals in villages without microfinance. The number of individuals is in bold and the number of households in brackets. The number of observations is estimated with de jure definitions (Morduch, 1999), pooling the participants over the three rounds (to over-sample participants). Column 1 of Table 1 (below) provides descriptive statistics of the main variables of interest.

The question of whether microfinance is beneficial to customers is a typical evaluation problem. The objective here is to identify the average effect of



- Collected in 1991–1992 by the Bangladesh Institute for Development Studies in collaboration with the World Bank, three microfinance programmes (Grameen Bank, BRAC, BRDB), 87 villages, 15 not covered by microfinance.
- The number of individuals is in bold and number of households in brackets.
- The number of observations is estimated with de jure definitions.
- De facto participants are actual participants. De jure participants are participants who have less than 0.5 acres of land. Morduch (1999) makes this distinction as some individuals participated in microfinance even if they had more than 0.5 acres of land and should have been excluded.

Notes: [1]Total number of participants using de facto definitions: 936; [2]total number of households participating using de facto definitions: 827

Figure 1. Description of the dataset with number of observations associated

participation in microfinance on, for example, expenditure, in cases where people have access to the former. A participant's expenditure should be compared to a counterfactual expenditure – that is, that of the same individual in the same situation at the same time without access to microfinance. Since the counterfactual is never observed, even with individual panel data observations, it must be estimated. Ideally, an experiment would randomly assign loans to people and compare average outcomes of groups with and without loans. Lacking a controlled randomised experiment, we must turn to non-experimental methods that mimic it under reasonable conditions.

One should not estimate microfinance's impact by differentiating expenditure between group D participants and group C non-participants. Such a difference would be misleading due to problems of self-selection. People self-select into microfinance programmes. This difference would measure entrepreneurial skills as well as microfinance impact. Equipped with such skills, participants might have done better with or without microfinance.

Pitt and Khandker (1998) use a regression discontinuity design based on the eligibility rule. Arguing that this rule is not respected, Morduch (1999) uses a simple difference-in-difference approach by comparing C+D to A, to E to B, and finds no significant effect resulting from microfinance exposure. Pitt (1999) noted however that this difference-in-differences approach fails to deal with the key issue of programme placement. The difference-in-differences strategy compares participating and non-participating individuals, before and after treatment occurs. However, pre-programme observations are unavailable in the dataset so the difference-in-differences technique cannot be applied. Other evaluation techniques cannot be used in this context. Zaman (2000) uses the number of eligible households in each village in 1992 as an instrumental variable. The rationale behind this is that while a larger number of potential members in a village will reduce the likelihood of eligible households from participating in microfinance, it is difficult to see why this variable should affect the poverty status of households. I have tested if the number of eligible households in each village in 1992 is correlated with participation in microfinance but not with the outcomes presently under consideration. It is not a significant variable in the participation equation and is not correlated with female non-land assets and female labour. It is, however, correlated with the logarithm of expenditure per capita, male labour, and male and female school enrolment. This instrumental variable is therefore not appropriate in this context.

Given the failure of traditional empirical strategies to appropriately measure microfinance's impact in this context, the technique of matching is preferable in this context since no clear criteria are available to explain the participation of individuals in microfinance. Indeed, the 0.5 acre eligibility cut-off is not respected in practice and participation in microfinance is voluntary. Comparing individuals across villages is also erroneous given the existence of non-random programme placement. In light of these difficulties, matching is preferable since it 'builds' an appropriate counterfactual for each participant.

III. Methodology

Statistical matching adjusts for differences in pre-treatment characteristics between treatment and non-treatment groups by pairing each treated individual to a non-treated unit with the 'same' observable characteristics. Given the assumption that relevant differences between two groups are captured by their observable characteristics (see online Appendix 1 for a mathematical expression of this first 'Conditional Independence' assumption of matching), the average outcome experienced by the matched pool of non-treated units/individuals identifies the counterfactual outcome the treated units would have experienced had they not been treated.

Put in a simple way, matching 'builds' out of the control group a synthetic individual that resembles best (based on some observable characteristics) each treated individual. It is then easy to compare the outcome of the treated individual to the outcome of the synthetic individual. This technique compares only what is comparable. Presented this way, matching does not seem to differ from a simple regression controlling for many factors. However, matching presents two advantages over a simple regression. First, matching compares a treated individual to its synthetic individual only if the synthetic individual is comparable enough to the treated individual (see online Appendix 1 for a mathematical expression of this second assumption of matching). If treated and untreated individuals are very different, a regression will give a result (by interpolating a linear relationship between the two groups). This is an undesirable result as individuals are not comparable. On the other hand, matching compares individuals only if they are 'close enough' to each other. This is called the Common Support assumption. I will later describe simple techniques making sure only individuals in the Common Support (comparable) are compared. Second, the matching technique is non-parametric as opposed to a linear regression imposing a linear structure on the data. After a synthetic individual is built, outcomes between the treated individual and the synthetic individual are differenced. No particular structure is imposed on the data. Heterogeneous treatment effects are allowed.

A final point has to be said on selection. The advantage of matching is that selection on observables is controlled for. This means that if participation in microfinance can be understood with observable variables, then matching would be consistent. However, the disadvantage of matching is that selection on unobservables is not controlled for. Heckman *et al.* (1997a), in their seminal paper on matching, argue that three causes of bias occur when estimating the average treatment on the 'treated'. First, there may be a difference in the support of observable characteristics of the two groups: individuals in both groups could be systematically different. Matching only on Common Support eliminates this bias. Second, there may be a difference between the two groups in the distribution of their characteristics over its Common Support. This bias is eliminated since matching compares one treated individual to one synthetic individual, effectively re-weighting the untreated dataset. Third, there may be a bias due to unobservables. The magnitude of that bias will depend on the adequacy of the Conditional Independence assumption in the specific problem.

Heckman *et al.* (1997a) are able to measure explicitly these three biases in the evaluation of a job training programme by comparing the results from matching to experimental impact estimates. They find that bias due to selection on unobservables is empirically less important than the other two components. They conclude that matching methods eliminate much of the bias as conventionally measured.

However, choosing the observable variables according to which matching is performed remains a critical step in the analysis. As a general rule, any variable that is thought to influence both participation and outcome is to be included.[5] The results section presents three different specifications in the choice of observable characteristics and justifies the choice of the preferred specification. Specification 1 is that used in Pitt and Khandker (1998) to explain participation in microfinance. Specification 2 includes variables that could be relevant to the participation decision based on economic theory in order to maximise the pseudo-R^2. In specification 3, the most relevant variables are retained in order to maximise the number of retained observations. The difference between specifications 2 and 3 illustrates the trade-off between explanatory power and multicollinearity. Clearly specification 2 should be preferred to specification 1: minimising the chance of having unobervables driving the participation decision means maximising the explanatory power of the specification by adding variables. As the data are not perfect, this comes at the cost of observations. A compromise might be reached by choosing the most explanatory variables that are also collected for the most observations in specification 3.

There are then three practical steps to follow when using the technique of matching. First, one has to match a treated individual with an untreated individual based on the chosen observable characteristics. It can be very time and effort consuming to match on a number of observable characteristics. This is called the curse of dimensionality. Rosenbaum and Rubin (1983) demonstrated that, instead of matching on a number of observable characteristics, one can calculate a score taking into account all these observable characteristics and match individuals on this single score. This score is called a propensity score and represents in this paper the propensity to participate in microfinance. In practice, one uses a logit regression of a participation dummy on the observable characteristics and the predicted score for each individual as the propensity score.

Second, each treated individual must be matched with an untreated individual or a set of untreated individuals. Four main methods exist. The most intuitive one is the 'Nearest Neighbour' method: a treated individual is matched with the untreated individual with the closest propensity score.[6] Although intuitive, the Nearest Neighbour technique must be considered very carefully since it does not impose the Common Support assumption: this technique will find a nearest neighbour even if there is no comparable 'enough' unit (with a very different propensity score). A refinement of this method is the 'Caliper' one. The 'Caliper' technique selects the nearest neighbour only if it is close enough. This technique imposes the Common Support assumption as opposed to the 'Nearest Neighbour' and should therefore be preferred. These two techniques match one treated individual with only one untreated individual. The 'Stratification' technique improves on these methods by using as a match an arithmetic average of all untreated individuals close enough (inside a 'stratum') to the treated individual in question.[7] Stratification imposes the 'Common Support' assumption and does not rely on a single untreated observation. For these reasons, it should be preferred to 'Nearest Neighbour' and 'Caliper'. However, equal weight is given to an individual at the limit of the stratum and to an individual close to the treated individual. The 'Kernel' estimation improves on this technique by giving each untreated individual a weight decreasing in distance to the treated individual in question. The Kernel estimation does not impose per se the Common Support assumption, since all untreated individuals in the comparison group are used, though it is very precise as more than one untreated individual is used to build a match. The Stratification and Kernel techniques have both advantages and drawbacks but should be preferred to Nearest Neighbour or Caliper. I will only present results using these two estimation techniques. Finding similar results with these two methodologies would strengthen the confidence in the results. It is also possible to test the power of each technique by calculating the balancing of the observable characteristics used before and after matching.[8] It is possible to *t*-test the equality of means in the treated and non-treated groups, both before and after matching. This test is equivalent to the ones performed for randomised experiments.

Third, one can simply difference outcomes experienced by a treated individual to its matched outcome to obtain the average treatment on the treated.[9] Additional sources of variability are introduced by estimating the propensity score and by the

Table 1. Propensity score estimates: determinants of the probability of participation

Independent variables	Means	Spec. 1	Spec. 2	Spec. 3
Highest grade completed	2.255	0.041	0.024	
	(3.173)	(0.03)	(0.036)	
Sex	0.513	−0.886***	−1.515***	−1.136***
	(0.499)	(0.123)	(0.182)	(0.128)
Age	22.327	0.051***	1.224***	1.065***
	(17.422)	(0.004)	(0.269)	(0.159)
Age of HH head	42.313	−0.046***	−0.035***	−0.014**
	(12.383)	(0.006)	(0.009)	(0.006)
No adult male in HH	0.024	1.951	2.854*	0.832***
	(0.153)	(1.268)	(1.562)	(0.308)
Parents of HH head own land	0.246	0.137	0.094	
	(0.56)	(0.14)	(0.147)	
Brothers of HH head own land	0.714	0.019	−0.023	
	(1.224)	(0.065)	(0.068)	
Education	0.551			0.336***
	(0.497)			(0.113)
Savings	1128.9		0.0002***	0.0002***
	(4201.37)		(0.0004)	(0.00003)
Own a non-farming enterprise	0.468		0.763***	0.630***
	(0.499)		(0.173)	(0.111)
Livestock value	3273.15		0.0000397	0.00005***
	(5533.9)		(0.00003)	(0.00002)
HH size	6.232		−0.117***	−0.147***
	(2.632)		(0.041)	(0.028)
Non-agricultural wage	4.023		−0.002	−0.006*
	(16.303)		(0.004)	(0.003)
Agricultural wage	2.987		0.013**	0.010**
	(9.755)		(0.007)	(0.005)
Age squared	802		−0.033***	−0.028***
	(1109.7)		(0.01)	(0.006)
Age to the power four	1874542		−1.73E-6*	−1.16E-6***
	(5029988)		(0.000000944)	(0.000000501)
Village dummies	yes	yes	yes	yes
Number of observations		4215	4205	5037
Pseudo R-squared		0.1502	0.3561	0.3313

Robust t statistics in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%. Specification 1 replicates the analysis of Pitt (1998). Specification 2 includes other control variables: landed assets, equipment assets, transport assets, injuries, change residence in the last 2 years, assets, expenses of the non-farming enterprises, agricultural costs, irrigated land, fathers still alive, marital status, agricultural income, mother's education, irrigated household land, mothers still alive, household land, highest grade completed by household head, sex of household head, no adult female in household, sisters of household head owning land, father's education, revenue of non-farming enterprises, dairy products sales; all insignificant.

matching process. We need therefore to obtain bootstrapped confidence intervals for the matching estimates, as there is no asymptotic distribution theory for these estimates.

What follows summarises the chosen methodology. The average outcome experienced by the matched pool of non-treated units identifies the counterfactual outcome the treated units would have experienced had they not been treated under the following assumptions: (1) relevant differences between the two groups are captured by their observable characteristics (in other words, the right set of *X*s described in the next section and included in the propensity score); and (2) treated individuals are matched with non-treated individuals on a Common Support (which is why we will only consider Stratification and Kernel techniques).

IV. Results

The first practical step of a matching analysis is the computation of its propensity score using the right set of observable characteristics. Once the preferred specification has been chosen, it is possible to obtain matching results of the benefits of microfinance.

Propensity Score

The left hand side variable is a variable that takes the value 1 when individuals participate in microfinance, 0 otherwise. Using a logit specification as the outcome is a binary variable, the predicted probability of each individual will then be its propensity score. The intuition for this technique is that two individuals with the same propensity scores, one participating in microfinance, the other not, will have no systematic differences beyond participation in microfinance. The difference in outcomes is due to microfinance and nothing else. The propensity score is crucial in this analysis since the key drawback of matching is that only selection on the observables X may be controlled for. Any variable influencing both participation and outcomes should be included. To deal with this problem, the three different specifications outlined above are employed.

Table 1 presents the results of this propensity score estimation. Column 1 presents descriptive statistics of each variable. The sample is restricted to individuals with less than 0.5 acres. All variables have 5,043 observations except 'Parents of HH head own land' and 'Brothers of HH head own land' which have only 4,372 and 4,368 observations. Column 2 of Table 1 presents results from the first specification. The explanatory power of the specification used by Pitt and Khandker (1998) is limited: the pseudo-R^2 of the participation regression is only 0.15. The propensity score coming from this specification will lack accuracy. Column 3 presents the results from specification 2, which includes every variable that could be thought of as having an impact on either participation or an outcome of interest: the pseudo-R^2 is 0.356. A surprising result in this specification is that the Highest Grade Completed is not correlated with participation. However, this variable was very seldom measured in the dataset and varies from one round to another for the same individual. Replacing this in specification 3, with a dummy variable indicating whether the individual went to school or not, is less precise but also less prone to measurement error.

Specification 3 excludes variables limiting the sample size, such as parents (brothers) of HH head own land. For this reason, 5,037 observations are now included. The coefficient preceding this variable is now positively significant in specification 3, as shown in column 4. The coefficient preceding gender is significantly negative since women are given preferential treatment in regards to loans by microfinance institutions. Individual's age is positively correlated with participation. This surprising result can be refined by introducing non-linear age variables. Individual age squared and to the power four are significantly negative in specifications 2 and 3. This indicates that older individuals are more likely to participate but with diminishing career concerns. Specifications 2 and 3 include variables that might influence both participation and outcomes. Individuals with savings, livestock or non-farming enterprises are more likely to participate in microfinance. This suggests that even though the Grameen Bank targets the poor, it is not the poorest who participate. It also highlights the fact that it is important to control for such variables when evaluating propensity scores in order to match comparable individuals. That the coefficient of agricultural wages is positive and that of non-agricultural wages is negative may suggest that the Grameen Bank targets agricultural professions. To conclude, specification 3 should be preferred since more observations are kept and the pseudo-R^2 is high.

The second step is to predict the propensity score of each individual. The logit specification was estimated only on the sample of individuals in villages with microfinance. This is logical because there are no observations about participation in the control villages. This specification thus does not provide coefficients on village dummies for individuals in control villages. I therefore predict a 'corrected' propensity score by equating all village variables to zero for individuals with access to microfinance in order to make it comparable to individuals without such access.[10]

Figures 2–4 present the distribution of the corrected propensity score for participants, non-participants in villages with microfinance and people without such access. In Figure 2, the propensity score for participants is well spread out over the interval [0, 1]. On the contrary, the distribution of the predicted propensity scores of non-participants in villages with microfinance in Figure 3 is skewed to the left. This is expected as these individuals did not participate. Their scores should therefore be low and Figure 3 is evidence that specification 3 has good explanatory power. Matching can very easily be understood on this graph: for each treated individual with a certain propensity score, a non-participant with the same propensity score (or according to one of the four techniques of matching described earlier) will be selected. Their outcomes will then be compared. These figures give a visual interpretation of the Common Support: treated individuals will be compared to a non-participant if and only if a non-participant is close enough. Treated individuals with a high propensity score will therefore be excluded from the analysis due to the lack of a similar untreated individual. Figure 4 presents the distribution of the predicted propensity scores of non-participants in villages without microfinance. Fewer individuals have a low propensity score. This is understandable as these villages contain individuals who would have sought a loan if only a microfinance institution was operating in the village. Common Support between participants and non-participants in villages without microfinance is greater than between participants and non-participants in villages with microfinance. Results comparing

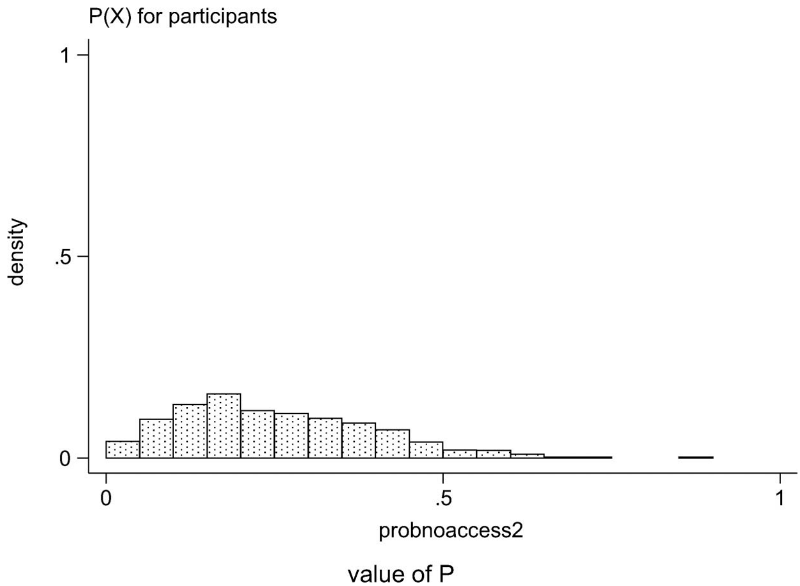


Figure 2. Distribution of the corrected propensity score for participants according to specification 3

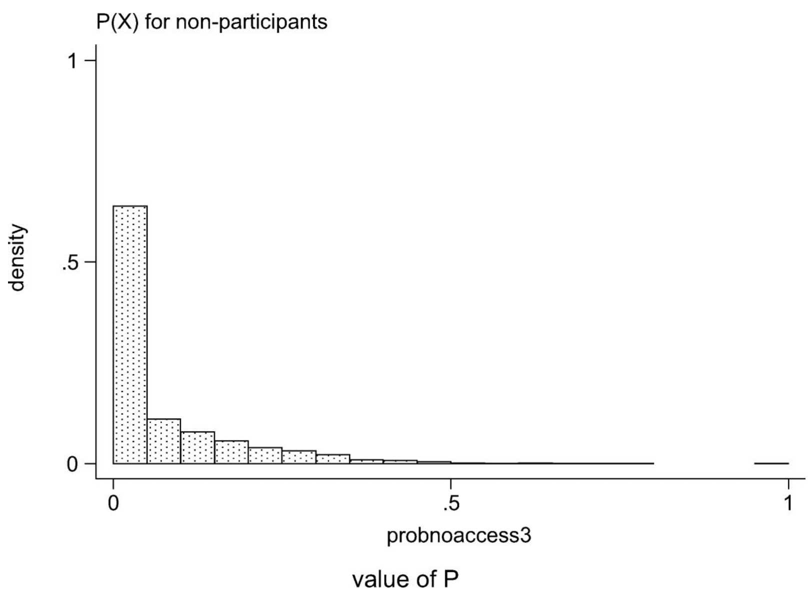


Figure 3. Distribution of the corrected propensity score for non-participants in villages with access to microfinance according to specification 3

participants and non-participants in villages without microfinance will therefore be more precise than results comparing participants and non-participants in villages with microfinance.

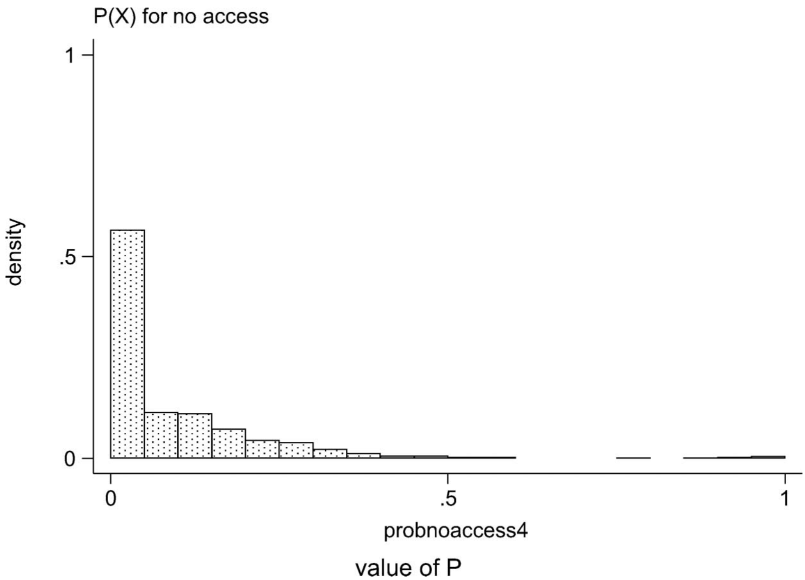


Figure 4. Distribution of the corrected propensity score for individuals in villages without access to microfinance according to specification 3

Matching Results

Matching results with a propensity score estimated with the preferred specification 3 and with the two preferred matching techniques (Stratification and Kernel) are presented in Table 2.[11] There is no clear cut answer in the matching literature about the parameters to be used when matching, so three strata are used for the Stratification technique: 20, 10 and 5 (the propensity score being between 0 and 100). A smaller stratum imposes the assumption of the Common Support but includes fewer observations. Three bandwidths are used for the Kernel technique (0.05, 0.02 and 0.01): a smaller bandwidth imposes the assumption of Common Support while Kernel matching converges to the nearest neighbour with decreasing bandwidths. The similarity of the results across parameters adds strength to the results.

The first row contains results comparing expenditure of participants to expenditure of matched non-participants in treated villages. To make sure that results do not come from systematic differences across villages, the logarithm of per capita expenditure is removed from village effects by regressing this quantity on village dummies from both the programme and control villages only, and then estimating the residual arising from this regression. This quantity is termed the 'pure' logarithm of per capita expenditure since it is now freed from any village level effects. The difference in expenditure is negative, contrary to prior expectations, but not very significant. This may indicate the presence of positive externalities resulting from microfinance. Non-participants in treated villages do not fare worse than participants. The second row compares participants and non-participants in control villages. The propensity score is corrected and per capita expenditure are purged from village level effects to take into account non-random programme placement

Table 2. Impact of microfinance on log of per capita expenditure from matching with the non participants in treatment villages and individuals in control villages using specification 3

	Stratification			Kernel		
Control group	20	10	5	0.05	0.02	0.01
Non participants in treatment villages	−0.035*	−0.044*	−0.044*	−0.039*	−0.044*	−0.046*
Individuals in control villages	0.028	0.028***	0.028*	0.028***	0.028***	0.028***

Estimates are bootstrapped. *significant at 10%; **significant at 5%; ***significant at 1%. Specification 1 is the specification used by Pitt (1998), Specification 2 includes more relevant variables to maximise the explanatory power of the propensity score, Specification 3 retains the most significant variables in order to maximise observation numbers and explanatory power. I use three strata for the stratification technique of 20, 10 and 5 (the propensity score being between 0 and 100). A smaller stratum imposes the assumption of the Common Support but includes less observations. I use three bandwidths for the kernel technique of 0.05, 0.02 and 0.01. A smaller bandwidth imposes the assumption of Common Support while kernel matching converges to the nearest neighbour with decreasing bandwidths. Propensity scores are corrected for non-random programme placement by equating all village dummies to 0 for treated individuals. The outcome is purged from village level effects by regressing it on village dummies and accounting for residuals.

and according to the procedure described earlier. The results of the Stratification technique are significantly positive. The Kernel estimate results are significantly positive and very close to those of stratification. The former are very robust in regards to bandwidth changes. The Kernel estimate is about 0.028 and significant at the 1 per cent level. Standard errors have been bootstrapped with 100 replications where bootstrap samples are of the original sample size with replacement. Results are the same according to the percentile or bias corrected method. Logarithms of per capita expenditure are the outcomes. This means that a participant would be able to spend 3 per cent more than a comparable individual in a control village. Individually, this means that on average people spend 250 Takas more than non-participants. Interpreted another way, it means that out of a 100 Taka loan, people may spend 2.8 Takas more.

This result is lower than Pitt and Khandker (1998), whose analysis was potentially biased upward. Using a regression discontinuity design, they basically compare individuals with landholdings just below 0.5 acres to individuals whose landholdings are just above 0.5 acres. However, as Morduch (1999) has shown, a number of individuals with landholdings greater than 0.5 acres actually received microfinance loans. There is no participation discontinuity in the 0.5 acre programme. Morduch points to one participant possessing 13.4 acres of land. A situation where treatment does not depend in a deterministic way on a certain criteria such as acreage is called a fuzzy design as opposed to a sharp design in the regression discontinuity literature (Hahn *et al.*, 2001: 202). This does not preclude estimation. However, Hahn *et al.* (2001) show that results could be biased upwards if there is no discontinuity in the probability of receiving treatment at the threshold of 0.5 acres.[12] My result is greater than the one of Morduch (1999) since the latter found the impact of microfinance insignificant. Indeed, Morduch only estimates the effect of exposure to, and not participation in, microfinance. Moreover, as Pitt (1999) showed, Morduch (1999) did not take into account non-random programme placement. Villages without microfinance may be more equal, while the poor in microfinance villages may be systematically worse off than the poor in non-microfinance villages. This will underestimate the impact of the programme.

Microfinance may impact the lives of the poor in many other ways. We can replicate our analysis with six different outcomes used in the previous literature. Variation in expenditure is used in Morduch (1999). Women's non-land assets, female and male labour supply, and female and male school enrolment are studied in Morduch (1999) as well as in Pitt and Khandker (1998). To systematically compare results to the previous literature, Table 3 presents results comparing treated individuals with non-treated individuals in villages without microfinance, correcting for non-random programme placement in the propensity score and in the outcome, using the Kernel technique.

The first row of Table 3 concerns variation in the log of per capita expenditure, calculated as the variance of the log of per capita expenditure over the three rounds of data collection, following Morduch (1999). As Morduch (1999) has shown, we find microfinance having a negative impact on expenditure variation. Microfinance has an income smoothing effect. However, the results are not very significant, as opposed to Morduch (1999). This discrepancy could be explained by the fact that Morduch does not take into account non-random programme placement as Pitt (1999) noted. If microfinance institutions targeted villages with higher inequality, it may be that ineligible households are richer on average in treatment than in control villages. Expenditure variability could be higher for richer individuals. Therefore, expenditure variability of microfinance participants could appear lower because of non-random programme placement. The matching technique is immune to this problem because it compares a treated individual with the 'same' untreated

Table 3. Impact of microfinance on 6 different outcomes using a kernel technique

	Bandwidth of kernel		
Outcomes	0.05	0.02	0.01
Variation of log of per capita expenditure	−0.008	−0.008	−0.008
Log of women non-land assets	0.037	0.037	0.038
Female labour supply	9.503	9.507	9.521
Male labour supply	17.001***	16.996***	16.974***
Girl school enrolment	0.051***	0.051***	0.052***
Boy school enrolment	0.035*	0.035	0.036

Estimates are bootstrapped. *significant, at 10%; **significant at 5%; ***significant at 1%. Specification 3 is used. I use three bandwidths for the kernel technique of 0.05, 0.02 and 0.01. A smaller bandwidth respects the assumption of Common Support while kernel matching converges to nearest neighbour with decreasing bandwidths. Propensity scores are correctcd for non-random programme placement by equating all village dummies to 0 for treated individuals. The outcome is freed from village level effects by regressing it on village dummies and accounting for residuals.

individual in treatment villages or a treated individual to the 'same' untreated individual in a control village.[13]

Women also appear to benefit from microfinance as far as their non-land assets are concerned as visible in the second row of Table 3, although the increase is not significantly positive. Pitt and Khandker (1998) find a positively significant effect of microfinance on women's non-land assets. But their results could again be biased upward due to the fuzzy design of their regression discontinuity. Moreover, the aforementioned risk reduction would appear to result from income rather than consumption smoothing. People tend to work more when participating in microfinance. Women typically work around nine hours more per month as visible in the third row of Table 3. This is a smaller estimate than for men, which suggests that the former's access to microfinance increases household consumption, presumably by increasing the productivity of their market time rather than by increasing the supply of that time. One possible explanation for the fact that men increase their supply of labour more than women may be found in the average size of loans available to both sexes. Men tend to borrow bigger amounts. The average Grameen Bank loan to male customers was 13,642 Takas, whereas the average loan to female customers was 11,542 Takas. Men's projects tend to be bigger although the difference in loan amounts granted to female and male customers is not that large. These results are at odds with Pitt and Khandker (1998), who find a positive impact for women and a negative impact for men. The last point is associated with the impact of microfinance on education. Microfinance seems to increase both male and female school enrolment. Girls' school enrolment in particular is positively affected by the participation in microfinance. Weaker results for boys may simply be a reflection of their greater initial enrolment. In fact, 60 per cent of boys were enrolled, compared to only 56 per cent of girls. Results on enrolment are lower than those of Pitt and Khandker (1998). Their results could be biased upward due to the fuzzy design of their regression discontinuity. It is interesting to note that consistent with this theory, all results (except for male labour supply) are lower than the ones of Pitt and Khandker (1998).

V. Benefits and Costs of Microfinance

Aims of a microfinance institution may vary. Some might favour outreach and benefits felt by borrowers. Some others might strive for financial sustainability. Finally, some might want to achieve both of these aims. This section presents a lending guide for the Grameen Bank which could help this institution to better select its customers depending on which goal it wants to achieve. More than its direct policy implications for the Grameen Bank, this section presents a methodology that could be used by other microfinance institutions.

A key advantage of matching is to provide individual measures of benefits brought by microfinance. It is thus possible to understand the factors determining success at the individual level. Results are presented in column 1 of Table 4 (the sample size is restricted to the Grameen Bank so there are only 158 observations). A microfinance institution wishing to maximise benefits felt by its customers should lend to borrowers presenting the characteristics found to be highly significant in column 1 of Table 4. One surprising outcome is that education does not seem to have a significant impact on earned benefits. In other words, education affects participation in microfinance (as was visible in Table 1) but not the amount of benefits earned once participating. Among the Grameen Bank customers, education does not seem to play a big role in determining success, although they are systematically more educated than the rest of the population. Gender does not affect benefits. The popular perception that females benefit more seems wrong. On the other hand, savings and non-agricultural wages positively affect benefits. The fact that savings

Table 4. Determinants of the benefits, costs and cost-benefit using specification 3

Independent variables	Benefit	Cost	Cost-benefit
Education	212.045	16.778	−90.768
	(326.552)	(16.691)	(352.802)
Sex	826.728	−23.454	1412.204**
	(560.222)	(20.131)	(605.256)
Age	334.917	54.930***	241.468
	(246.153)	(21.023)	(265.94)
Household land	6.338	−0.13	4.546
	(10.564)	(0.112)	(11.413)
Age of HH head	21.2867	2.386***	20.84
	(19.132)	(0.883)	(20.67)
Savings	0.331*	0.007*	0.542***
	(0.18)	(0.004)	(0.194)
Own a non-farming enterprise	−284.309	60.966***	−183.072
	(370.924)	(17.313)	(400.741)
Duration of ownership of a non-farming enterprise	3.671**	0.036	3.024*
	(1.705)	(0.075)	(1.842)
Livestock value	0.045	−0.002	0.061
	(0.048)	(0.002)	(0.052)
Agricultural income		−0.010***	0.003
		(0.003)	(0.15)
Household size	−175.677*	−0.333	−179.733*
	(98.804)	(3.86)	(106.747)
Agricultural wage	−4.614	1.581**	−21.262
	(17.514)	(0.638)	(18.922)
Father's education	130.211**	−4.001	113.608*
	(58.156)	(2.611)	(62.83)
Age squared	−13.678	−2.289***	−7.747
	(10.311)	(0.761)	(11.14)
Sales of dairy products	0.922	−0.049***	1.442*
	(0.849)	(0.016)	(0.917)
Non-agricultural wage of the household	16.349***	0.351	15.236***
	(5.037)	(0.244)	(5.442)
Agricultural wage of the household	10.98	0.103	15.068*
	(8.165)	(0.354)	(8.822)
Number of observations	158	1364	158
Pseudo R-squared	0.5219	0.3707	0.5687

*significant at 10%; **significant at 5%; ***significant at 1%. Specification 3 is used. The benefits are derived from individual matching results. Costs are profits made by microfinance institutions and corrected by late repayments. The Cost-Benefit is the difference between benefits and costs.

and outside wages positively affect benefits may suggest that the loans with the weekly repayment schedule imposed by the Grameen Bank can only be put to good use by those who can pay the instalments with other funds while a longer term project funded by the Grameen loan germinates.[14] Household size also impacts benefits negatively. Microfinance does not seem to benefit the poorest of the poor.

Another goal of microfinance institutions could be financial sustainability. Morduch (1999) provides a detailed analysis of costs borne by microfinance institutions. This paper focuses on a different source of hidden costs: late repayments. The Grameen Bank displays an impressive 98 per cent repayment rate. This claim is true in the long run, as visible in Figure 5 which shows actual repayment as a function of time to scheduled repayment date. However, the origin of the graph indicates that some borrowers repay late. After 30 days, the repayment rate is 71 per cent. This represents an opportunity cost for microfinance institutions which could have lent the money to another borrower at the same interest rate. People differ in their repayment delay. In this dataset, I am able to measure repayment delays for each loan. To understand what drives repayment, I have regressed in column 2 of Table 1 the repayment delay on the same individual characteristics as previously found in column 1. Microfinance institutions striving to maximise repayment rates would do well to lend to young people (preferably women), people with high livestock values and agricultural income, and people selling dairy products. Focusing on these significant factors could result in a quicker and cheaper selection of successfully repaying customers.[15]

This analysis is very close in spirit to Ahlin and Townsend (2007), Cassar *et al.* (2007) and Karlan (2007) who try to uncover the fundamental determinants of

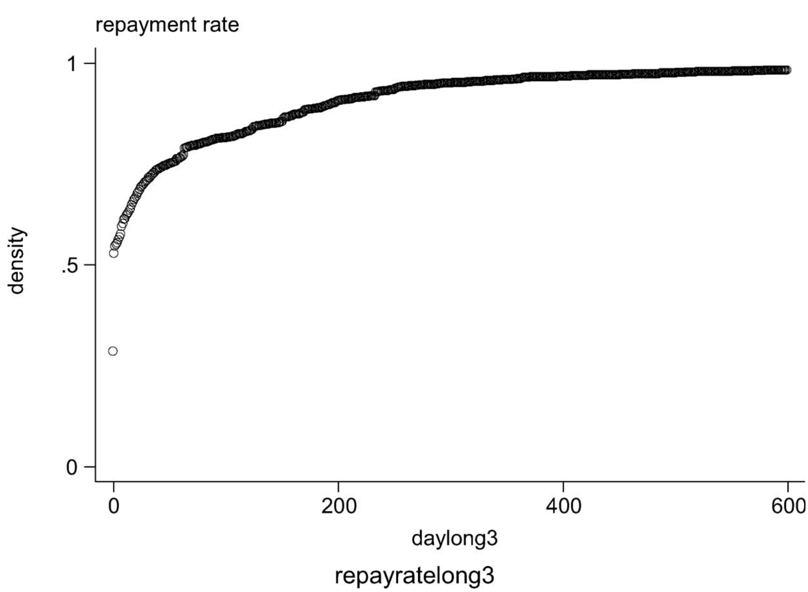


Figure 5. Percentage of people having repaid depending on time of delay (from 0 to 600 days, from 0 to 80 days, from –500 to 600 days)

repayment. For example, Cassar *et al*. (2007) find that personal trust between group members and social homogeneity are important to explain repayment. Karlan (2007) finds that more socially connected groups perform better. This paper considers a different dependent variable: late repayment. An additional contribution of this paper is to compare costs to benefits in the same analysis to show that social benefits and financial sustainability might be attainable at the same time.

It is possible to transform a delay measured in days into monetary value.[16] It is then possible to compare for each customer the benefits brought by microfinance to the costs imposed on the microfinance institution by his late repayment. Figure 6 provides a visual interpretation of the results. In this figure, the x-axis represents the profit made by a microfinance institution with a particular customer, taking into account the problem of late repayment. The y-axis represents the benefits brought by microfinance to the customer, measured from the matching technique. A microfinance institution looking for maximisation of social benefits should lend to individuals in the top part of this graph. A microfinance institution looking for financial sustainability should avoid late repayment individuals and lend to individuals on the right side of the graph. The best customers are those in the first quadrant, who benefit from microfinance and generate profits. One can then easily find the determinants of such individuals by calculating an index comparing these two measures. The results are shown in column 3 of Table 4. Customers who maximise cost–benefits are males, have high savings, high paternal education levels, high dairy product sales and high agricultural as well as non-agricultural household wages. Focusing on these factors, which are found to be significant, may be a quicker and cheaper way of selecting customers. This result contradicts popular perceptions surrounding microfinance. According to column 2, women have higher repayment rates while men obtain better cost–benefit ratios from microfinance.

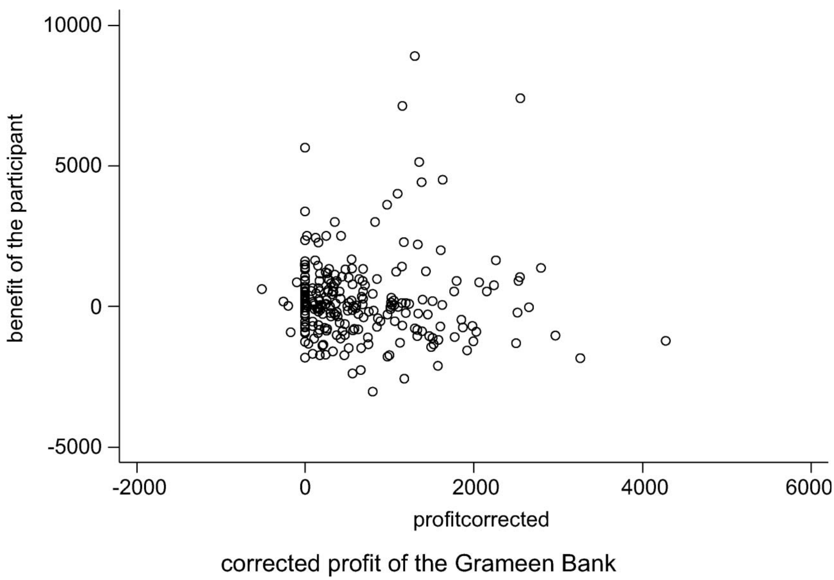


Figure 6. Distribution of the benefits according to the *corrected* profit generated

This study suffers from a number of limitations. The first concerns the range of the dataset; the final regression rests on a mere 158 observations. Even if the R^2 is high, it is insufficient to draw strong implications. Another arbitrary aspect of this work is to compare costs and benefits on an equal basis. Microfinance institutions' first concern might be to ensure their financial sustainability by considering cost regressions. Only once this is ensured could they put more weight on benefits experienced by customers and therefore maximise social benefits.

VI. Conclusion

By comparing participants to matched individuals in non-treated villages, I found that microfinance has a positive impact on participants' expenditure, supply of labour and male/female school enrolment. Participants spend on average 3 per cent more than similar non-participants. This result is lower than Pitt and Khandker (1998) because their analysis, based on a 0.5 acre eligibility rule that was not enforced, was potentially biased upward. The result is greater than Morduch (1999), who found microfinance to be of insignificant impact but limited estimation to the effect of exposure to, and not participation in, microfinance. Moreover, as Pitt (1999) showed, Morduch (1999) did not account for non-random programme placement. While villages without microfinance may be more equal, the poor in villages with microfinance may be systematically worse off than the poor in villages without it. This will tend to underestimate the impact of the programme.

This paper finds significantly smaller estimates than previous studies or the Grameen Bank estimates, and also suggests that microfinance is not good at targeting the poorest in participating villages. However, a word must be said regarding the limitations of this approach. Microfinance does have a number of other positive impacts not evaluated here. It empowers women formerly restricted by social custom from working outside the home, promotes self-sufficiency and enhances education by providing training. Moreover, traditional evaluation literature focuses on the programme's direct effect and is not a general equilibrium analysis. There is some evidence of positive externalities to microfinance inside villages, but more research is needed on this topic.

The paper also considered the rarely discussed problem of late repayment. The repayment rate of 95 per cent is in fact on average achieved 291 days after a specific repayment date; after 30 days, the repayment rate is only 71 per cent. This generates cost. A formal institution would provision against potential losses. Analysing the corrected profit of the Grameen Bank at the individual level identified two characteristics of a loan: the benefit it brings to customers and the cost it generates for microfinance institutions. A selection guide can therefore be provided to analyse the most suitable customers, improving repayment rates and/or benefits earned by individuals.

Acknowledgements

I am indebted to my advisors Robin Burgess and Tim Besley for their many comments and fine guidance. I would like to thank the participants of LSE/EOPP seminars for numerous useful comments and two anonymous referees. Financial support form ESRC, STICERD, Royal Economic Society is gratefully acknowledged.

Notes

1. See http://www.grameen-info.org/bank/bcycle.html. The target group is the group of people with less than 0.5 acres of land, which is the eligibility condition for microfinance.
2. For example, if the programme was first placed in villages where people would benefit more from the programme, such a comparison could yield a positively biased result.
3. See the work of Besley and Coate (1995) for a theoretical explanation as to why social ties within group lending could improve repayment performance. See Karlan (2007) and Cassar *et al*. (2007) for an empirical investigation of joint liability group lending.
4. See Pitt and Khandker (1998: 974) for a thorough description of the data.
5. It is interesting to note that we need variables good enough to provide a reasonable explanation of participation but not too good, since in cases where a certain X allowed us to discriminate perfectly between participant and non-participant, then a regression discontinuity design with such a variable would be more appropriate than a matching technique (Heckman *et al*., 1997a: 637).
6. See online Appendix 2 for a mathematical definition of these methods. This Appendix is available with the online version of the journal.
7. See online Appendix 2 for a mathematical definition of these methods.
8. Using the command pstest in Stata.
9. See online Appendix 2 for a mathematical definition of these methods.
10. This is equivalent to setting all coefficients on village dummies to zero. The reason for this is that there are no coefficients for village dummies in control villages.
11. The *t*-tests for equality of means of observable characteristics in the treated and non-treated groups, both before and after matching, are not presented but were always conclusive.
12. This is obvious when looking at the regression discontinuity coefficient expressed by Hahn *et al*. (2001: 202). If the denominator goes to zero (no difference in the probability of treatment because of the threshold), then the fraction could become infinitely large. In plain terms, if there is no discontinuity there cannot be a regression discontinuity design.
13. Additionally, expenditure variations are netted out of village level effects through the procedure calculating the 'pure' expenditure variation (residuals from a regression of expenditure variation on village level dummies).
14. I thank an anonymous referee for this insightful suggestion.
15. Ownership of non-farming enterprises is negatively related to repayment. This may suggest that people with such enterprises do not appear to repay on time. Microfinance seems more adapted to agricultural activities.
16. See online Appendix 3 for an explanation of the monetary transformation of a repayment delay.

References

Ahlin, C. and Townsend, R. (2007) Using repayment data to test across models of joint liability lending. *Economic Journal*, 117, pp. F11–F51.

Banerjee, A. and Newman, A. (1993) Occupational choice and the process of development. *The Journal of Political Economy*, 101, pp. 274–298.

Besley, T. and Coate, S. (1995) Group lending, repayment incentives and social collateral. *Journal of Development Economics*, 46, pp. 1–18.

Cassar, A., Crowley, L. and Wydick, B. (2007) The effect of social capital on group loan repayment: evidence from field experiments. *Economic Journal*. 117, pp. F85–F106.

Hahn, J.,Todd, P. and Van Der Klaaw, W. (2001) Identification and estimation of treatment effects with a regression discontinuity design. *Econometrica*, 69(1), January, pp. 201–209.

Heckman, J., Ichimura, H. and Todd, P. (1997a) Matching as an econometric evaluation estimator. *Review of Economic Studies*, 64(4), pp. 605–654.

Kaboski, J. and Townsend, R. (2005) Policies and impact: an analysis of village-level microfinance institutions. *Journal of the European Economic Association*, 3(1), pp. 1–50.

Karlan, D. (2007) Social connections and group banking. *Economic Journal*, 117, pp. F52–F84.

Karlan, D. and Zinnman, J. (2007) Expanding credit access: using randomised supply decisions to estimate the impacts. CEPR discussion paper DP6407.

Morduch, J. (1998) Does microfinance really help the poor? New evidence from flagship programs in Bangladesh. Department of Economics and HIID, Harvard University, June.

Morduch, J. (1999) The role of subsidies in microfinance: evidence from the Grameen Bank. *Journal of Development Economics*, 60, pp. 229–248.

Pitt, M. (1999) Reply to Jonathan Morduch's 'Does microfinance really help the poor? New evidence from flagship programs in Bangladesh'. 14 October.

Pitt, M. and Khandker, S. (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), pp. 958–996.

Rosenbaum, P. and Rubin, D.B. (1983) The impact of group-based credit programs on poor households in Bangladesh: does the gender of participants matter? *Biometrika*, 70, pp. 41–55.

Zaman, H. (2000) Assessing the poverty and vulnerability impact of micro-credit in Bangladesh: a case study of BRAC. Policy research working paper 2145, The World Bank.

Appendix 1

Statistical matching adjusts for differences in pre-treatment characteristics between treatment and non-treatment groups by pairing each treated individual to a non-treated unit with the 'same' observable characteristics. Given the assumption that relevant differences between two groups are captured by their observable characteristics, the average outcome experienced by the matched pool of non-treated units/individuals identifies the counterfactual outcome the treated units would have experienced had they not been treated.

Matching's two assumptions are:

- the Conditional Independence Assumption (CIA): it is a condition on the set of observables X. Non-treated outcomes are independent of the participation status:

$$Y_0 \perp D|X \quad \text{for } X_i \in S$$

(S being defined in the next assumption). In other words, it means that the non-treated outcomes are what the treated outcomes would have been had they not been treated.

- The Common Support (set S): all treated agents have a counterpart on the non-treated population and anyone constitutes a possible participant:

$$P(D = 1|X) < 1 \quad \text{for } X_i \in S$$

Upon verification of these two assumptions, matching appears well-suited to deal with potential bias. This is made clear by decomposing the treatment effect in the following way:

$$\begin{aligned} E(Y_1 - Y_0|X, D = 1) = {} & [E(Y_1|X, D = 1) - E(Y_0|X, D = 0)] \\ & - [E(Y_0|X, D = 1) - E(Y_0|X, D = 0)] \end{aligned}$$

The first term is observed, while the second is called the bias conditional on X. Three causes of bias may occur when estimating the average treatment on the treated. Following Heckman *et al.*'s analysis (1997a), there may, first, be a difference in the support of X in the two groups: individuals in both would be systematically different. Matching only on Common Support eliminates this bias. Second, there may be a difference between the two groups in the distribution of X over its Common Support. This bias is eliminated since matching reweighs $D = 0$ data in order to equate the distribution of X in the $D = 1$ sample. Third, there may be a bias due to unobservables. The magnitude of that bias will depend on the adequacy of the CIA assumption in the specific problem.

Appendix 2. Four Matching Methods

Each treated individual must be then paired with a group of comparable non-treated individuals (depending on the matching technique). The outcome y_i of individual i is associated with a matched outcome $\hat{y}_i$ equal to a weighted outcome of each component of a comparison group:

$$\hat{y}_i = \sum_{j \in C^0(p_i)} w_{ij} y_j$$

where C^0 (p_i) is the set of neighbours of treated individual i in the $D = 0$ group and w_{ij} is the weight on control j in forming a comparison with treated i.

I used four matching techniques:

- Nearest Neighbour: treated unit i is matched to non-treated unit j such that:

$$\left|p_i - p_j\right| = \min_{k \in \{D=0\}} \{|p_i - p_k|\}$$

- Caliper Matching: for $\delta > 0$, i is matched to j such that:

$$\delta > \left|p_i - p_j\right| = \min_{k \in \{D=0\}} \{|p_i - p_k|\}$$

This technique imposes the Common Support hypothesis as opposed to Nearest Neighbour.

- Stratification: for a pre-specified length of a stratum, the matched outcome is the arithmetic average of the outcome of the individuals in the stratum.
- Kernel Matching: here we consider all observations with a decreasing weight with the distance between two observations:

$$\hat{y}_i = \frac{\sum_{j \in \{D=0\}} K\left(\frac{p_i - p_j}{h}\right) y_j}{\sum_{j \in \{D=0\}} K\left(\frac{p_i - p_j}{h}\right)}$$

I have chosen here a Gaussian K: $K(u) \propto \exp(-u^2/2)$ which takes into consideration all non-treated units.

The average treatment of the treated (ATT) is then:

$$ATT = \sum_{i \in \{D=1 \cap CS\}} (y_i - \hat{y}_i)\omega_i$$

where $\omega_i = 1/$(number treated within the common support).

These four techniques should be compared. The Nearest Neighbour technique must be considered very carefully since it violates the Common Support hypothesis. Even if there is no comparable 'enough' unit, this technique will provide an estimate. We should not be overly attentive to the results of this estimate. Imposing a Caliper is more appropriate but remains imprecise compared to Stratification and Kernel techniques since only one individual is matched with each participant. The Stratification technique uses an average of several individuals as a matched individual. However, equal weight is given to an individual at the limit of the stratum and to an individual close to the observed unit, since the average is only arithmetic. The Kernel estimation gives each individual a weight decreasing in distance compared to the unit being studied. The Kernel estimation does not impose per se the Common Support hypothesis, since all individuals in the comparison group are used, though it is very precise, since several individuals with the proper weights are used in the matched outcome. The Stratification and Kernel techniques should therefore be preferred for these reasons. I will present only those results obtaining from these two estimation techniques.

Additional sources of variability are introduced by estimating the propensity score and by the matching process. We need therefore to obtain bootstrapped confidence intervals for the matching estimates, as there is no asymptotic distribution theory for these estimates.

What follows summarises my chosen methodology. The average outcome experienced by the matched pool of non-treated units identifies the counterfactual outcome the treated units would have experienced had they not been treated under the following assumptions: (1) relevant differences between the two groups are captured by their observable characteristics (in other words, the right set of Xs described in the next section and included in the propensity score); and (2) treated individuals are matched with non-treated individuals on a Common Support (which is why we will only consider Stratification and Kernel techniques).

Appendix 3. Costs of Microfinance

To transform a delay measured in days into monetary value, one has to consider institutional details of the Grameen Bank repayment system. Interest rates on loans are 20 per cent. The Grameen Bank employs the declining balance method: it takes into account the fact that loans are regularly repaid throughout the year in weekly instalments. Let P be the principal of the loan, the amount borrowed. Let f be the weekly repayment rate and r the interest rate:

$$f = 1 + \frac{r}{52}$$

At the end of the week, the balance is thus fP. Repayment is a constant x every week. The remaining balance to be paid at the end of the week is:

$$fP - x$$

The following week, this sum is multiplied by f and x is repaid. The remaining balance is:

$$f(fP - x) - x$$

The balance at the end of n weeks is:

$$Pf^n - x(1 + f + f^2 + \cdots + f^{n-1}) = Pf^n - x\frac{1 - f^n}{1 - f}$$

Here, r is 20 per cent. Depending on the principal P and the repayment duration n (in weeks), we can compute x the weekly repayment amount (the balance must be zero at the end of repayment):

$$x = \frac{Pf^n}{\frac{1-f^n}{1-f}}$$

Interest income earned by the Grameen Bank is thus:

$$nx - P = n\frac{Pf^n}{\frac{1-f^n}{1-f}} - P$$

I considered for every loan a hypothetical profit earned from a loan of principal P paid on time after n weeks. However, when there was a delay in the repayment, the Grameen Bank could have lent the principal plus the interest (nx) at the interest rate of 20 per cent during the delay. The profit earned by using this operation would have been:

$$delay \times \frac{nxf^{delay}}{\frac{1-f^{delay}}{1-f}} - nx$$

In the profit formulae, I replaced principal P with nx or new principal, and the duration of this hypothetical loan with the delay. The total profit for the bank on a particular loan when delays are taken into account is now:

$$\underbrace{\left(n\frac{Pf^n}{\frac{1-f^n}{1-f}} - P\right)} - \underbrace{\left(delay \times \frac{nxf^{delay}}{\frac{1-f^{delay}}{1-f}} - nx\right)}$$

The first term is the theoretical profit the Grameen Bank makes after each loan dispensation, the second is the negative impact of a delay. The difference is therefore the actual profit made by the Grameen Bank on each loan with the repayment problem taken into account. It is easy to compute this variable from the data. It is now possible to compare costs and benefits at the individual level since I have transformed the delay into a monetary value.