ISSN: 2038-7296
POLIS Working Papers
[Online]

Dipartimento di Politiche Pubbliche e Scelte Collettive – POLIS
Department of Public Policy and Public Choice – POLIS

POLIS Working Papers n. 180

February 2011

Microcredit and poverty. An overview of the principal statistical methods used to measure the program net impacts

Cristina Elisa Orso

UNIVERSITA' DEL PIEMONTE ORIENTALE "*Amedeo Avogadro*" ALESSANDRIA

Periodico mensile on-line "POLIS Working Papers" - Iscrizione n.591 del 12/05/2006 - Tribunale di Alessandria

Microcredit and Poverty. An overview of the principal statistical methods used to measure the program net impacts.

Cristina Elisa Orso*

Abstract

The purpose of this paper is to examine different econometric approaches aiming to evaluate the impact of microcredit on poverty. Starting with a brief description of microcredit and the most common kinds of statistical biases connected to these studies, I describe the principal characteristics of Non-Randomized and Randomized approaches, in order to highlight strengths and weaknesses concerning the application of such methodologies.

JEL-Classification: C54, C58.

Keywords: Microcredit, poverty, program impact, randomized approach, non-randomized approach.

*Ph.D. Student in "Economic Policy", Department of Economic and Social Sciences, Catholic University, Piacenza. E-mail: cristina.orso@unicatt.it

1. Introduction

Scientific testing of the net impact of microcredit is very difficult.

It is possible to identify a specific question which represents the core of every serious impact study: if we find out that people who have obtained loans are doing better than those who haven't, does it necessarily mean that receiving loans caused the positive change?

In other words, the problem is to understand how outcomes have changed with the program compared to what would have happened without the program implementation.

The most important challenge, in this particular context, has been to determine a control group for comparison; it is very hard to identify a group of people who are like the program participants in all relevant features apart from not having received funds. The critical issue to evaluate how microfinance works is the measurement of the net effects caused by the programs.

As outlined by Armendàriz, Morduch (2010), the first contributions of microcredit impact literature mainly concerned non-experimental methods. In this context, researchers use treatment and control groups, but do not randomly assign subjects to the groups. The critical point in such studies is to establish causality relationships. Moreover, biases from omitted variables, non-random program placement, client selection and self selection, and attrition lead to important estimation problems (Karlan, 2001).

Some kinds of biases can be reduced by using longitudinal data. The effects of non-random participation and non-random program placement, under specific assumptions, can be mitigated by the implementation of this strategy. But if there are unobservable variables that change over time, attributes hard to measure such as entrepreneurial organization and business skills are probably correlated with participation status. The most popular longitudinal studies have been sponsored by USAID in the second part of 1990s. Researchers investigated net impacts on members of three different institutions: a microlender organization operating in the informal sector (SEWA) located in Ahmedabad (India), an ACCION International affiliate (Mibanco) situated in Peru and the Zambuko Trust in Zimbabwe. The sample households were observed a first time to collect baseline data, and then, after two years, they were resurveyed (Armendàriz, Morduch, 2010).

Some empirical work is based on household surveys from the World Bank and the Bangladesh Institute of Development Studies in

Bangladesh. The studies that exercise the most influence on the research community are Pitt and Khander (1998) and Khander (2005). As regards PK approach, they use cross-section data (from the three seasonal rounds in 1991/1992), while Khander takes into account the panel dimension of the data set (he also uses the 1999 round) in order to strengthen identification. To follow, many other studies are based on this data set. We report, for instance, Khander (1996, 2000), Pitt et al. (1999), McKernan (2002), Pitt and Khander (2002), Pitt et al. (2003), Menon (2005), and Pitt, Khander and Cartwright (2006).[1]

Concerning experimental approach, randomized controlled trials (RCTs) , when done correctly, can really provide credible and transparent estimates in difficult contexts. Such approach consists of giving loans to a subgroup randomly drawn from the population (for instance through a random algorithm to select people from a list), while the other subgroup, also randomly selected, will not obtain any funds.Using randomized approach, the difference in terms of average outcome between the two distinct groups represents a good estimate of the program's average impact. Under determinate assumptions, the result can be interpreted as the causal impact of microcredit introduction.[2]

However, we must notice that it is an average impact; therefore, if half of the participants gains by 100 percent while the other half loses by 100 percent, the average impact will be zero. In a similar context, zero is a clean estimate of the average impact, but it doesn't give us any relevant information about an individual's performance.

In order to obtain credible estimates, it is important that neither agreement to participate in the experiment nor the tendency to drop out are systematically related to outcomes of interest.

Randomization approaches also have limits: few published experimental studies of microfinance have been able to highlight short-term results only.

Banerjee et al.(2009) and Karlan and Zinman (2009) looked at microcredit participants over a short period of time (12-18 months). They didn't find any kind of improvement in household income or consumption, but both the studies showed some other benefits.

[1] For a more complete literature review, see Armendàriz, Morduch, 2010.

[2] Because of RCTs implementation, loans represent the only ex-ante difference between the treatment and the control groups.

The former concerns a randomized evaluation on the community-level impact of the introduction of new branches of a local microfinance bank. The baseline sample was randomly selected from urban Hyderabad (India) for the opening of new microfinance branches.

Before new institutions opened, 69 percent of the households had at least one outstanding loan from informal channels (moneylenders, family, friends). The authors found that the new branches settlement caused, in the interested areas, more new business openings, higher purchases of durable goods and relevant profits in existing business activities. However, it's important to note that the main effects concern target households starting new businesses, and the authors can't tell if the funds are actually invested in business activities or not.

Interestingly, Dupas and Robinson (2009) conducted a randomized field experiment in Kenya that found short-term welfare improvement regarding micro-savings, not microcredit.

They gave interest-free savings accounts in a local microfinance institution to a random sample of small entrepreneurs; these accounts did not pay interest and charged withdrawal fees, but they represented the only opportunity for formal savings in the area.

The results suggested wide variation in the intensity of formal account usage. Some microentrepreneurs did not accept the proposal, and many accepted without using the account effectively; roughly 50 percent of those with accounts used them only rarely, and only a small minority used them frequently.

Clients of this service increased their investment and daily expenditures for women, but there was no evidence in terms of men impact. However, this study presents a few shortcomings: first, the random sample is small (185 microentrepreneurs) and only few clients used the account frequently. Moreover, the target area of experiment was limited to a single site and to a single microfinance bank .

Karlan and Zinman (2008) conducted a randomized experiment in South Africa. The authors asked loan officers of a local lender to revalue and accept applicants for a microloan from a set who were initially refused but who got just below the cutt-off.

Many applicants who were reconsidered (but not all) showed income improvements and a more positive outlook on the future. Despite of a micro-entrepreneur set rejected and not reconsidered, they registered a higher level of stress and depression.

The results of this study can't generalize to other microfinance contexts, because the loans were consumer loans, not typically used for investments in business activities. In addition, applicants were not very poor, and interest rates were higher than those applied by typical microfinance institutions (Rosenberg, 2010).

The remainder of this paper is organized as follows. The next section provides a brief description of microcredit and its principal objectives, among which poverty alleviation. Section 3 gives a sketchy outline of the problems arising from reverse causation and selection bias in order to discuss about the econometric methodologies to use. Section 4 and section 5 respectively describe the two distinct empirical approaches, examining the results reached in both fields on the basis of recent literature. Section 6 compares randomized and non-randomized strategies. The last section concludes with a discussion of the principal impact issues.

2. Definition and aims of Microcredit

According to its basic definition, microcredit involves small loans to poor people for self employment projects that produce income, allowing them to care for themselves and their families.[3] These individuals, also called "unbankable", lack any kind of economic collateral and a verifiable (and valuable) credit history to offer a traditional bank. Hence, they cannot meet the basic qualifications required to access formal credit. Microcredit can be considered a field of microfinance, which embraces the provision of a wider range of financial services designed for the very poor. As highlighted by Armendariz, Morduch (2010), recently the terms *microcredit* and *microfinance* have often been used interchangeably even if they show some remarkable differences. The original focus on microcredit mainly concerned poverty reduction and social change, and the most important institutions involved were NGOs. Then, the shift to *microfinance* arose from the growing need of poor households for a broader range of financial services (like savings) and not exclusively credit for entrepreneurial activities. As a result, this lexical transition has produced an orientation change, toward less poor people and toward a different kinds of organizations, commercially oriented and with strong financial regulations.

The term microcredit is often used in distinct contexts (rural credit, agricultural credit, consumer credit, etc.) with different implications. In order to clarify the target and the specific objectives of

[3] http://w.w.w.microcreditsummit.org

the programmes, we introduce a broader classification to identify the various categories.

On the basis of Yunus taxonomy[4], it is possible to identify five different kinds of microcredit:

- Traditional informal microcredit (moneylender's credit, pawn shops, loans from friends and relatives, consumer credit in informal markets). Amongst those of that sector local moneylenders are an important source of credit to those borrowers usually refused by most financial institutions because of their particular economic conditions. Low income and lack of collateral and stable employment do not make borrowers creditworthy in the eyes of traditional banks. Moneylenders are better at serving clients neglected by the formal sector because they have a considerable market knowledge and lower transaction costs. At the same time, they can easily monitor borrower behaviour because of their proximity to the client and reliable information about his or her status. Poor people turn to moneylenders mainly for consumer loans or to cope with emergencies like health problems or to pay for high outlay connected with education, wedding, funerals, and so on. Hence this source of credit cannot be considered a valid engine of inclusion and local growth.

- Microcredit based on traditional informal groups (tontine, ROSCA). Tontine can be compared to our life insurance. The basic scheme is simple. Each participant contributes a sum of money to the tontine and, then, he receives an annual dividend on his investment. When each participant dies, his or her share is reallocated among the surviving subscribers, until only one investor survives. ROSCA (Rotating Savings and Credit Association) consists of a group of individuals who agree to meet for a specific period of time in order to save and borrow money from a common "pot", usually allocated to one member of the group (who changes each period). Both kinds of microcredit pursue partially different objectives compared to modern microcredit. The aim of tontine concerns insurance, while ROSCAs finance consumer credit (Armendàriz-Morduch, 2010).

- Microcredit through traditional banks, generally specialised in specific investment sectors (agricultural credit, livestock credit, fisheries credit, handloom credit, etc.).

- Cooperative microcredit and rural credit through specialised banks.

[4] http://www.grameen-info.org

- Modern microcredit (Grameen credit, bank-ONGs partnership based microcredit and consumer microcredit). Small loans typically designed for low-income clients who traditionally lack access to commercial banking for several reasons (absence of collateral, informal employment, unverifiable credit history, high transaction costs). This kind of credit allows poor people to start new entrepreneurial activities, expand existing businesses, to cope with shocks due to adverse climatic conditions and illness, and smooth out consumption.

Concerning the last category of microcredit, three models for lending have become globally popular: *village banking*, *solidarity groups*, and i*ndividual lending*. The choice of lending technology relies on the specific characteristics and needs of the target areas.

The *village banking* approach is most similar to the old cooperative credit movement. This method was innovated by FINCA International founder, John Hatch. A village bank is an informal self-help support group of 20-30 participants (predominantly women). The first loan comes from an institution like MFI or NGO, then following deposits come from individual savings amounts in group funds. For example, if the first loan is $ 50 and the participant saves in the same period $ 10, the second loan will be equal to $ 60. The length of the loan cycle is 4-6 months[5]. *SHGs* (self-help groups) have dominated the microfinance landscape especially in India. These involve small groups of 10-20 members that collect savings from their participants and, at the same time, provide them with loans. The members are jointly and severally liable for the funds obtained within the group itself (Nair, 2005).

Solidarity groups is a lending mechanism which allows a group of people to provide collateral through a group repayment pledge[6].Usually borrower groups are made up of three to seven members, most commonly five[7], and the patterns of disbursements and repayments is

[5]For detailed information, see:
http://en.wikipedia.org/wiki/Village_Banking

[6]http://www.accion.org/ Group_Lending

[7]Five is the right size of a Grameen group. Grameen Bank (or Rural Bank) was started by Muhammad Yunus, a Professor of the University of Chittagong (Bangladesh) in 1976. The bank mainly targeted rural women for its credit programmes. It introduced group lending strategy to make credit available to the poor, usually denied by commercial loans because of the lack of physical collateral. The income-generation also aims to empower the women and increases their participation in household decisions. For more details about its "mission" and history, see Armendàriz, Morduch, 2010, chapter 4.

regimented. According to the Grameen model, payments start immediately after disbursement every week. In practice, the first two participants take their loans and begin with repayment, then two more, and finally the fifth. The eligibility to further and larger amounts of funds is subject to the repayment of all group loans. Each borrower is responsible for the repayment of all the other loans within the group (joint liability). The advantages of this methodology concern lenders and customers. First, it is convenient for villagers because the bank comes to them (as in the cases of ROSCAs and local moneylenders), avoiding logistic and administrative problems. At the same time, for the lender, transaction costs connected to loan disbursement are considerably reduced. Although this methods boasts some evident strengths, the higher likelihood of group failure in the event of a single default, the huge training costs and the greater financial responsibility for others in a group have to be taken into consideration when looking into its implementation (Armendàriz, Morduch, 2010).

The case of *individual lending* substantially differs from previous approaches. Typically, microcredit is associated with group structures (solidarity groups or village banks). Whilst the first two lending models serve the poorest, individual lending serves less poor clients. This targeting choice arises from the high costs of underwriting new loans, monitoring behaviour and repayment of disbursed funds, and enforcing repayment process. Group lending strategies, as mentioned above, shift much of the responsibility onto borrowers, while lending to individuals involves (for the MFI) managing these costs directly. Interestingly, however, some microcredit institutions do not offer group but individual loans. BDB (Bank Dagang Bali) and BRI (Bank Rakyat Indonesia), two Indonesian private banks, are revealing examples of individual microfinance[8]. As in conventional lending, loan officers take collateral and collect information from credit bureaus, require pledges of title to land or to other property. Sometimes, the individual lending is a part of a larger "credit plus" program, where other particular kinds of services (such as skill development, education, health, etc.) are provided.

To sum up, it is possible to highlight some general key objectives of the microcredit programmes:

(a) to provide small loans to poor people at lower cost than informal sources;

[8] For more detailed information, see Armendàriz, Morduch, 2010.

(b) to avoid exploitation of the poor due to the growth of informal credit channels;

(c) to reach the "unbankable" that cannot be financed by traditional banks (because of the lack of collateral);

(d) to empower women both within the household (as decision makers) and in society (through active economic and political participation);

(e) to improve employment opportunities;

(f) to reduce poverty, increase growth and improve the living conditions on sustainable perspectives.

As regards the last point, microcredit is gaining importance as an effective tool of poverty alleviation. Impact evaluations aim to investigate the role of credit access in terms of poverty reduction. Although some interesting research[9] analyzes the effects of microcredit programmes through a multidimensional poverty perspective[10], financial outcomes remain the core matter in many relevant impact studies. The lack of reliable data and the selection problems connected with empirical evaluations represent a serious complication to estimate microcredit consequences.

3. Econometric impact evaluations: selection bias and causality

Microcredit and the other microfinance services can affect individuals and households in many different ways. The first question

[9] Karlan, Zinman (2010) conducted an interesting study using an experimental approach. They measure the impact of microcredit programmes (in Manila) also in terms of subjective well-being using indicators such as life satisfaction, job satisfaction, decision making power, optimism, etc. In addition, the authors investigated the treatment effects on different kinds of human capital.

As mentioned in the introduction, Karlan and Zinman (2008) created another interesting experiment aiming to estimate impact effects of microcredit in South Africa. In addition to financial impact measures (such as income and consumption), they consider particular types of indices concerning health aspects (physical and mental health index) and decision- making process, optimism, and position in the community socio-economic ladder (this information comes from subjective perceptions of sample households).

[10] Multidimensional poverty involves a group of deprivations that cannot be adequately expressed as income insufficiency. It refers to specific composite measures, such as the Human Poverty Index, that accounts for well-being indicators (like a decent standard of living, a long and healthy life, knowledge). For more details, see Tsui, 2002.

that researchers have to ask is the following: what are we trying to measure?

Concerning the impact of microcredit, we distinguish between two different effects:

(a) an income effect, that makes households wealthier and pushes up consumption levels (it can also increase the demand for children, health, children's education, spare time);

(b) a substitution effect, which may counterbalance the income effect. In fact, with increased female employment rates, hours spent raising offspring can be costlier in terms of foregone income, driving birth levels down.

But increasing income and consumption are not the only metric of judgement of microcredit impact.

As argued in a interesting book, *Portfolios of the poor: How the World's Poor live on \$2 a day* (Collins, Morduch, Rutherford, and Ruthven, 2009), the poor tend to use credit and savings not only in order to smooth consumption, but also to cope with emergencies like health problems and pay for expensive services such as education, weddings, or funerals.

A lot of studies have looked at the experience of people who have obtained microloans.

In order to have a clean estimate of evaluating impacts, the statistical problem is to separate out the causal role of microcredit program. In other words, it is necessary to gauge microcredit effects eliminating the various *reverse causation* and *selection* biases.

As regards *causality*, researchers have to answer fundamental questions. For example, if they note that wealthier households have larger loans, they have to ask if the loans enrich the households or do richer households merely have less difficulty in accessing credit, without a substantial increase in their productivity (Armendàriz, Morduch, 2010).

The challenge has been to identify an appropriate control group for comparison; we have to ask if changements as new business, new saving accounts, further education for children, etc. are due to the program implementation or would have happened also without microcredit introduction.

We can summarize the effect produced by a specific treatment (T) on a characteristic Y (in order to understand whether and in what size the treatment may be considered a determinant factor of Y) as follows[11]:

$$\alpha = E(Y_i^1/T_i = 1) - E(Y_i^0/T_i = 0)^{12} \tag{1}$$

In this context, the causal effect is measured as the difference between the outcome that would be observed if unit i received the treatment ($E(Y_i^1/T_i = 1)$), and the outcome of receiving no treatment of any kind ($E(Y_i^0/T_i = 0)$).

Econometric methods attempting to estimate how much of the distinct outcomes between treatment and control groups attributable to the program are a crucial tool in microcredit interventions, especially in non-experimental settings. The question that every evaluation seeks to solve is how would partecipants have done in absence of intervention. As well argued in Holland (1986), the fundamental problem of causal inference concerns the impossibility of evaluating, at the same time, treatment and no treatment, that is $T_i = 1$ and $T_i = 0$; obviously, it is not possible observing the results obtained on the same unit from the two different situations ($E(Y_i^1/T_i = 1)$ and $E(Y_i^0/T_i = 0)$) in order to gauge the causal effect of T on Y.

Different counterfactual statistical estimation methodologies have to be implemented on the basis of the specific evaluation context.

In particular, we imagine to have to measure the causal impact of microcredit on borrower income. The income level can be attributed to different kinds of sources: measurable attributes, for example, like job, business, pension, age, education and experience, that are generally available.

But another category of personal characteristics is hard to measure, for instance organizational ability, entrepreneurial skills, access to valuable social networks. In this latter category, we include economic shocks, and other types of casual events that could affect household outcomes (Armendàriz, Morduch, 2010).

In addition, a further set of attributes may be useful in order to estimate microcredit impact, such as village size, or the existence of scale economies related to specific production (actually, this kind of information is measurable but not gathered in surveys).

[11] See Duflo, Glennetster, and Kremer (2007)

[12] where "α" represents the measure of the impact, $E(Y_i^1 /T_i = 1)$ is the expected value of the outcome variable observed after treatment on target units, while $E(Y_i^0 /T_i = 0)$ is the expected value after program implementation on control group.

Estimating microcredit impact implies separating out its role from the roles of all these different attributes.

Bank officers work hard to screen potential ranges of customers, and calculate the optimal locations for new branches; loans are thought to attract the most gifted individuals, who choose to partecipate or not in a microcredit program on the basis of personal and strategic reasons (in particular, on the basis of perceived returns).

If target clients are wealthier and more productive than the non-treated group, it could be attributed to the strategic placement of the intervention, not to the active role of the microfinance institution in making these conditions. Hence, a high correlation between microcredit participation and, for example, the variables age and entrepreneurial ability is very probable.

In this context, if investigators manage measurable attributes (like age), there are simple strategies in order to control for age-related issues, but when there are typical unmeasurable attributes, such as entrepreneurial ability researchers have to be cautious in making comparisons between ex-ante and ex-post situations. The effect of being a good microentrepreneur could incorrectly be interpreted as an impact of program access (Armendàriz, Morduch, 2010).

The counterfactual estimation (and, at the same time, the impact estimation) can be affected by relevant problems, indicating as threats to validity (Bartik and Bingham, 1995). Concerning impact evaluations, the essential problem is the risk that the comparison between the target and the control group might be contaminated by factors which inhibit the untreated units from simulating the without-intervention situation (*selection bias*).

The problem with comparing microcredit participants to non-participants is that participants are self-selected and therefore not comparable to the non-participants. Many microcredit clients already have initial advantages respect to their neighbors.

In other words, program target units may have systematic differences compared to the control group units, and these differences could cause a biased evaluation of the intervention results.

Formally, if we consider equation (1) as a definition of program impact, and we subtract and add the term $E(Y_i^0/T_i = 1)$, we obtain

$$\alpha = E(Y_i^1/T_i = 1) - E(Y_i^0/T_i = 1) - E(Y_i^0/T_i = 0) + E(Y_i^0/T_i = 1)$$

$$\alpha = E(Y_i^1 - Y_i^0/T_i = 1) + E(Y_i^0/T_i = 1) - E(Y_i^0/T_i = 0) \qquad (2)$$

The term $E(Y_i^1 - Y_i^0/T_i = 1)$ is the treatment effect; the other term $E(Y_i^0/T_i = 1) - E(Y_i^0/T_i = 0)$ represents the measure of selection bias. It captures the difference in potential untreated outcomes between the target and the control groups. Treated units may have had different results on average even if they had not been treated. The bias can move in two distinct directions: if, for instance, program participants are more motivated in seeking goals, have high-level entrepreneurial skills, or live in a richer geographic area, they are more likely to achieve good results in terms of outcomes. In this case, $E(Y_i^0/T_i = 1)$ could be larger than $E(Y_i^0/T_i = 0)$. On the other hand, if the treatment implementation arises in particularly disadvantged communities, with an higher rate of poverty, the term $E(Y_i^0/T_i = 1)$ would be smaller than $E(Y_i^0/T_i = 0)$. The crucial point is that in addition to any kind of treatment effect there may be systematic differences between participants and non participants (Duflo, Glennerster, Kremer, 2007).

Given that the term $E(Y_i^0/T_i = 1)$ represents the expected outcome for a borrower who received a loan, if he had not received the loan, such term is not directly observable and we don't assess the size (and the sign) of the selection bias. Many empirical papers aim to identify in what cases selection bias does not exist or find strategies to correct for it (Armendàriz, Morduch, 2010).

4. Non-randomized approach

4.1 The estimation strategy

The main contributions of recent literature use innovative research designs to overcome selection bias problems.

Following Coleman's study (1999), we initially analyse a standard empirical specification concerning the evaluation of program impacts in the microfinance framework.

We start from the following specification:

$$\mathrm{B}_{ij} = H_{ij}\alpha_B + L_j\beta_B + \epsilon_{ij}, \qquad (3)$$

$$\mathrm{Y}_{ij} = H_{ij}\alpha_Y + L_j\beta_Y + B_{ij}\delta_Y + \mu_{ij} \qquad (4)$$

where B_{ij} represents the amount borrowed from the village bank by household i in village j, H_{ij} is a vector of household attributes, L_j is a vector of village characteristics and Y_{ij} is an outcome on which measuring the impact. The parameters $\alpha_B, \beta_B, \alpha_Y, \beta_Y, \delta_Y$ have to be estimated during the analysis. The error terms ϵ_{ij} and

μ_{ij} represent unmeasured household and village attributes that determine microcredit participation and outcomes, respectively. The parameter δ_Y measures the causal impact of a microcredit program on the outcome Y_{ij}.

A crucial assumption in order to obtain unbiased econometric estimation of the parameters is that the error terms ϵ_{ij} and μ_{ij} are uncorrelated. If this assumption is violated, the estimate of the parameter of interest (δ_Y) will be biased. This kind of correlation can arise from (a) self-selection into the village bank and (b) nonrandom program placement.

Concerning the first source of correlation, we consider a sample of households selected only from communities with a local bank. Some households will receive loans, while others will not participate in a microcredit program (on the basis of specific eligibility criteria). In this context, a correlation between ϵ_{ij} and μ_{ij} is almost certain; for example, if the more promising households are selected to be borrowers , the unmeasured "entrepreneurial ability"might affect both the choice to become a program participant and the impact estimation of the outcomes.

As regards the second source of correlation, we imagine to have another common kind of sample made up of households from a village with a local bank and randomly drawn households from communities without any village bank. If the placement of the local bank is not random, it is more likely that ϵ_{ij} and μ_{ij} are correlated across different villages. To illustrate this situation, consider a simple example: we suppose that an NGOs has to decide where placing a village bank, on the basis of its own criteria. Presumably, the institution will choose more entrepreneurial or better organized communities, with a higher level of income, then the terms ϵ_{ij} and μ_{ij} will be correlated.

Moffitt (1991) proposes three standard procedures used in the case of correlation between ϵ_{ij} and μ_{ij} . The first concerns the instrumental variables approach. The identifying instruments might be determinants of participating in the microcredit program, but not determinants of the impact measure[13].

The second strategy regards the introduction of panel data, in order to take into account the pretreatment systematic differences in outcome variables. But collecting a panel dataset is often difficult and costly.

[13] This will be covered in more detail later.

Finally, the third method suggested by Moffitt concerns assuming an error distribution (usually a normal distribution) of the outcome variable in absence of any kind of treatment. Then, the impact of the microfinance intervention can be defined by measuring the deviations from normality of the variable of interest within the treated units. This procedure involves three relevant problems:

(i) In many contexts, researchers haven't sufficient information on which to base assumptions about the error distribution;

(ii) The impact estimation is highly sensitive to the initial assumptions about the error distribution;

(iii) If analysts use, for instance, a censored regression (Tobit model), the identification of the impact effects is sometimes impossible[14].

[14] The Tobit model is appropriate when the dependent variable is censored at some upper or lower bound because of the way the data are collected (Tobin, 1958, Maddala, 1983). If we decide to censor at a lower bound, the empirical specification will be:

$Y^*_{ij} = X_{ij}\beta + \epsilon_{ij}$ (1)
$Y_{ij} = Y^*_{ij}$ if $Y^*_{ij} > 0$ (2)
$Y_{ij} = 0$ if $Y^*_{ij} \leqslant 0$ (3)

where Y^*_{ij} is an unobserved continuous latent variable, Y_{ij} is the observed variable, X_{ij} represents the vector of the independent variables, ϵ_{ij} is the error term and β is a vector of coefficients. In addition, we assume that the error term is uncorrelated with the vector X_{ij} and is independently and identically distributed. We can generalize the model by introducing a known nonzero constant in equation (2) and (3), in substitution of the threshold zero. Variations of the censoring point across observations may happen (Winship, D.Mare, 1992).

OLS estimation of the first equation involves a selection bias problem. In fact, for the set of information $Y_{ij} > 0$, the above model implies:

$$\begin{aligned} Y_{ij} &= X_{ij}\beta + E\left[\epsilon_{ij} \mid Y^*_{ij} > 0\right] + \eta_{ij} \\ &= X_{ij}\beta + E[\epsilon_{ij} \mid \epsilon_{ij} > -X_{ij}\beta] + \eta_{ij}. \qquad (4) \end{aligned}$$

η_{ij} represents the difference between ϵ_{ij} and $E\left[\epsilon_{ij} \mid Y^*_{ij} > 0\right]$ and is uncorrelated with both terms. We note that $E[\epsilon_{ij} \mid \epsilon_{ij} > -X_{ij}\beta]$ in equation (4) is a function of $-X_{ij}\beta$. The less the rate of censoring (that is $-X_{ij}\beta$), the greater is the conditional expected value of the error term ϵ_{ij}. In this context, the OLS regression estimates are biased and inconsistent because of the negative correlation between $-X_{ij}\beta$ and ϵ_{ij}. It is possible to construct a similar equation to the (4) for observations for which $Y_{ij} = 0$, generating a parallel analysis, but the inclusion of observation for which $Y_{ij} = 0$ leads to analogous inconveniences. Starting from the analysis of equation (4), Heckman (1979) shows how selection bias may be thought of as an omitted variable bias. Specifically, the term $E\left[\epsilon_{ij} \mid Y^*_{ij} > 0\right]$ can be interpreted as an omitted variable correlated with X_{ij} and that affects the outcome. Hence, biased and inconsistent OLS estimates of the vector coefficients (β) hinge on its omission (Winship, D.Mare, 1992).

4.2 The Coleman alternative specification

Coleman (1999) introduces a new approach consisting of using information on future clients before the microcredit program is started. The author exploits a particular way a microcredit intervention was implemented in Northeast Thailand and suggests an interesting way to address selection bias. The author gathered data on 445 households in fourteen communities. Of these, only eight had local banks beginning their activity at the start of 1995 .The other six did not, but local village banks will be set up a year later. The "control" village bank households would have, presumably, the same unobservable characteristics as the "treatment" group of village bank members who had already received the loans. Moreover, members and non-members of control and treatment villages were surveyed.

Taking into account this survey design, Coleman(1999) estimates the following regression equation[15]:

$$\mathrm{Y}_{ij} = H_{ij}\alpha_Y + L_j\beta_Y + M_{ij}\gamma + T_{ij}\delta + \nu_{ij} \qquad (5)$$

This kind of approach allows a refinement of the difference in difference method[16]. In particular, the dummy variables are introduced to control for membership status and location of the intervention. Specifically, M_{ij} represents a dummy variable equal to 1 if household i in village j self-selects into the microcredit program, and 0 otherwise. The term T_{ij} is another dummy variable equal to one if a self-selected member has already benefited from the credit interventions, and 0 otherwise. In this specification, the variables

---

[15] The author replaces equations (3) and (4) by an unique impact equation.

[16] Difference-in-difference designs use pre-intervention differences in outcomes between treatment and comparison group for control for unobserved heterogeneity between the groups, when data are available both before and after the intervention. Consider Y_1^T the potential outcome in the case of treatment (Y_1^C corresponds to the case of no treatment) in period 1, after the program implementation. Then, we denote by Y_0^T the potential outcome if the subject is treated (Y_0^C if the subject is not treated) in period 0, before the program occurs.

Subjects belong to group T or group C, and the T group is treated in period 1 and untreated in period 0. The control group (C) is never treated. Formally, the difference-in-difference estimator is:

$$DD=\left[\hat{E}\left[Y_1^T \mid T\right]-\hat{E}\left[Y_0^C \mid T\right]\right]-\left[\hat{E}\left[Y_1^C \mid C\right]-\hat{E}\left[Y_0^C \mid C\right]\right]$$

Under the assumption that $\left[\hat{E}\left[Y_1^C \mid T\right]-\hat{E}\left[Y_0^C \mid T\right]\right] = \left[\hat{E}\left[Y_1^C \mid C\right]-\hat{E}\left[Y_0^C \mid T\right]\right]$, this estimator provides an unbiased estimation of the program impact (Duflo et al. 2007).

of most interest are M_{ij} and T_{ij}. Coleman suggests that M_{ij} can be interpreted as a proxy for the unobservable attributes, which leads subjects to self-select into the local bank. In other words, M_{ij} captures the unobserved variables that caused the correlation between ϵ_{ij} and μ_{ij} across households. The variable T_{ij} represents the number of months that the loans of the village bank were available to members who have self-selected, which is exogenous to the household. Following Coleman (1999), we argue that M_{ij} controls for selection bias in order to obtain a consistent estimate of the causal treatment impact described by δ , the coefficient of T_{ij}.

Coleman's findings suggest that average treatment effects were not significantly different from zero after checking for endogenous member selection and microcredit program placement. In addition, the author expands the estimation frame to distinguish between impacts on "rank-and-file members" and members of the local bank committee (who are, usually, wealthier); the results show that most program impacts were not statistically significant for rank-and-file members, while there were some relevant impacts for the committee members in terms of wealth accumulation.

However, the study needs to be put into the larger financial outlook. Thai villagers are relatively wealthier than, for example, Bangladeshi villagers, and have an easier access to credit, from different sources. In this analysis, the village banks' loans may be too small to produce relevant average differences in the welfare of households. Coleman recognizes that one reason that wealthier borrowers may have performed consistent impacts was because they could manage larger loans.

4.3 Quasi experimental designs: Bangladesh studies

As mentioned above, a different approach to overcome statistical problems may be searching for an instrumental variable for microcredit program participation. This strategy allows analysts to address some kinds of omitted variable bias, reverse causality and problems arising from measurement error[17]. In practice, the instrumental variables method consists in finding an additional variable or set of variables that gives an explanation for levels of credit received, but that has no correlation with the error term in the regression framework. Then, the proxy variable formed on the basis of the instrumental variable approach can be use to extract

[17] For a more detailed discussion about instrumental variables strategy see Greene, 2008, Part III.

the causal impact of credit access. To find appropriate instrumental variables for microcredit is complex. But when we have within-village variation in program access the basis of the evaluation methodology can be the eligibility rules (this is the approach using in some important studies of microfinance impact in Bangladesh).

The most-famous studies about microcredit impacts on households are based on a survey fielded in Bangladesh in the 1990s.

Pitt and Khander (1998) develop a framework for estimating impact effects using the first round of data (1991-1992 cross-section). They analyze surveys of 1,798 households in 89 villages randomly drawn within 29 *upazillas*[18] of Bangladesh. The starting point is that the observations concerning the three programs evaluated (Grameen Bank, BRAC, and the state-run Rural Development Boards (RD-12)) all answer the same eligibility rule[19]. To focus the attention on the poorest, all the three program formally defined eligibility rule in terms of land ownership: households having over half an acre of land are not allowed to borrow.

Following PK estimation set-up, the crucial feature of the estimation problem is that the credit variables are potentially endogenous and censored. Moreover, outcomes of interest as labour supply and girl's school enrolment are censored or binary. To estimate impact parameters, PK use a limited-information maximum likelihood (LIML[20]) framework, that takes into account instrumental variables and handles censoring. According to the kind of outcome, the model will be as continuos and unbounded (household consumption), Tobit (female and male labour supply per month, female non-land assets), or Probit (school enrolment of boys or girls).

The first model specification concerns the introduction of the credit choice variables denoting if females and males in a household can

[18] The districts of Bangladesh are divided into subdistricts called *upazillas.* (http://en.wikipedia.org/wiki/Upazilas_of_Bangladesh).

[19] In this study, the lending model was solidarity group. Some of these groups were all-male and more were all-female (none was mixed).

[20] The LIML estimator is based on a single equation under the assumption of normally distributed disturbances. It has the same asymptotic distribution as the 2SLS estimator. The advantage of the LIML estimator is its invariance to the normalization of the equation (see Greene, 2008, pp. 375-376 for the analytical demonstration). In particular, Davidson and MacKinnon (2004) show that LIML can be used with success when the sample size is reduced and the number of overidentifying restrictions is large.

borrow. Therefore, we have

$c_f = p_f e$

$c_m = p_m e.$

Let p_m and p_m dummy variables denote if microcredit group of men or women are effectively operating in a specific village, while e is a dummy that explains whether a household meets the eligibility criteria of the microfinance programs. Let y_0 denote the outcome. Then, the model choice (Tobit or Probit) relies on the kind of outcome y_0. But since we restrict our attention on household consumption, we will assume outcome is continuous and unbounded.

According to PK approach and subsequent contribution of the Roodman-Morduch study, we consider the following problem:

$$
\begin{aligned}
&y_0 = y_{fm}\prime\gamma + x\prime\beta_0 + \epsilon_0 \\
&y_f^* = x\prime\beta_f + \varepsilon_f \qquad if \qquad c_f = 1 \\
&y_m^* = x\prime\beta_m + \epsilon_m \qquad if \qquad c_m = 1 \qquad (6) \\
&y_f = 1\left\{y_f^* \geq C\right\} \cdot y_f^* \\
&y_m = 1\left\{y_m^* \geq C\right\} \cdot y_m^* \\
&(\epsilon_0, \epsilon_f, \epsilon_m)\prime \sim N(0, \Sigma).
\end{aligned}
$$

where y_f and y_m are the total borrowings of all women and all men household members. Let $\mathbf{y}_{fm} = (y_{f1}, y_{f2}, y_{f3}, y_{m1}, y_{m2}, y_{m3})\prime$denotes the six credit variables disaggregated by program and gender; $\mathbf{x}$ is a vector of exogenous controls, C represents the credit censoring level, Σ is a $3{\times}3$ positive definite symmetric matrix, and $1\{\}$ indicates a dummy variable.

The above econometric model presents some unusual features. First, as suggested by Roodman and Morduch (2009), "*Superficially, there appear to be no excluded instruments. Meanwhile, the credit equations' samples are restricted, which means that the number of equations in the model varies by observation.*"[21] Second, the outcome equation contains six different endogenous credit variables but in the model there are only two instrumental equations (for y_f^* and y_m^*). Apart from these unusual features, the key assumption behind the model is similar to a traditional two-stage instrumental variable set-up. Specifically, they estimate the following two equations by using LIML set-up:

$y_f^* = c_f \mathbf{x}\prime\beta_f + C + \epsilon_f$

[21] See also Wilde (2000) for a more advanced discussion.

$$y_m^* = c_m \mathbf{x}\prime \beta_m + C + \epsilon_m \qquad (7)$$

In this specific setting, PK effectively instrument for the borrowing variable creating interactions between the credit choice dummies and the exogenous variables included in the model. The PK exogenous variables are: age, sex, education of the household head, other household characteristics, a set of village characteristics and individual characteristics (in the case of regression on individual-level data). In addition, the authors included also the constant terms, c_f and c_m , which are instruments themselves.

The PK models for distinct outcomes have several characteristics. First, they are *conditional* (i.e., it means that their specifics, such as number of equations, vary by observation, being conditional on the data). Second, they are *recursive* in the sense that the specifications contain plain stages and do not model any kind of simultaneous causation. Moreover, as argued in Roodman (2009), the observed y_f and y_m appear in the equation (y_0); this implies that they are also *fully observed*. Finally, the models combine equations that show different types of censoring (*mixed process*)[22].

An important issue in this model concerns spherical errors. Since heteroskedasticity can make Tobit-type models inconsistent, the critical point is how much the previous assumption can be relaxed. In PK's carefully study homoskedasticity is implicitly assumed. Specifically, the authors assume identically but not independently distributed errors. Starting from assumptions on c_f and c_m, the identification framework is based on the exogeneity (after conditioning on controls) of this constant terms. In other words, the factors driving credit choice (i.e., the formation of credit groups by village and gender, and the eventual eligibility of individual households) must be exogenous. PK's approach does not suggest valid arguments in support of the exogeneity of the first factor. As regards the second (landholdings), they argue:

"Market turnover of land is well known to be low in South Asia. The absence of an active land market is the rationale given for the treatment of landownership as an exogenous regressor in almost all the empirical work on household behaviour in South Asia"[23]

But this seems to describe a case in which landholdings is external to the model and not exogenous (Heckman, 2000). The *exogeneity* notion is different: it requires that landholdings are related to

[22] For more details, see Roodman (2009).

[23] From Pitt and Khander (1998), p.970.

outcomes only through microcredit after linearly conditioning on controls[24].

Concerning the two PK identifying assumptions , Morduch (1998) remarks relevant questions. First, PK acknowledge that unobserved factors might influence both group formation and outcomes, generating endogeneity. Their strategy consists of including village dummies to control for any factors at the community-level. The Morduch's criticism is about sub-village effects (i.e., the village effects are not fixed within villages). For instance, we imagine a local community in which eligible households are not very poor. Then, in a similar context, group formation might be more likely and outcomes systematically better. In his following study, Pitt (1999) introduces interaction terms between landholdings and all the X variables to PK's instrument framework to strengthen their findings. Second, in the PK data land markets are active and there exists substantial endogenous mistargeting. In other words, 203 of the 905 borrowing households in the 1991-92 sample owned more than 0.5 acres before the microcredit intervention. Microlenders were not following the eligibility criteria strictly so that some of the over-half-acre households that received the loans may have been met with an alternative eligibility rule (as discussed in PK (1998), footnote 16: "*The quasi-experimental identification strategy used here is an example of the regression discontinuity design*"). Pitt (1999) replies to Morduch's criticisms pointing out that identification with LIML does not require the eligibility rule be perfectly respected but it merely drives an exogenous component of variation in borrowing.

With regard to impact results, PK found that "*annual household consumption expenditure increases 18 taka for every 100 additional taka borrowed by women...compared to 11 taka for men.*" In addition, they found that lending to female reduces the use of contraception and has a positive impact on schooling of boys (as regards Grameen Bank and RD-12, while women participation only in Grameen Bank has a positive effect on schooling of girls).

[24] According to Merriam-Webster's dictionary, *exogenous* means *caused by factors or an agent from outside the organism or system.* But the consistency of IV estimator implies that the instrument be orthogonal to the error term, which is not involved by the Merriam-Webster definition (Leamer, 1985). Heckmann (2000) suggests the term external for the Merriam-Webster definition. Its use concerns variables whose values are not set or caused by the variables in the model. Therefore, it is more correct keeping exogenous for the orthogonality condition that is required to obtain consistent estimation in IV context.

These findings reinforce two fundamental concepts about microcredit: first, its reliability as instrument to alleviate poverty, second the important role of women credit.

We now focus our attention on Morduch's study (1998). He uses an estimation strategy analogous to PK's, but simpler and less efficient. As a first step, he performs simple difference-in-difference estimates, then adds controls. Despite PK's study, Morduch finds no sharp evidence for strong impacts of microcredit on household consumption. However, he finds some evidence that microcredit helps households to diversify income flows so that consumption volatility is less pronounced across seasons. Moreover, the results hinge on the assumption that the village dummies totally capture all critical aspects about the communities that might affect the microlender's decisions concerning the program placement. In this analysis the village dummy variables only control for unobservable attributes that influence all households in a village identically and linearly[25]. In addition, Morduch finds weaker evidence that households with credit access tend to manage actively not only their spending, but also their labour income. He finally concludes (without any direct evidence) that the ability to smooth income over the year drives smoother within-year consumption levels.

The above empirical contributions arise from the analysis of the first round of data (1991-1992)[26]. But exhaustive impact studies using a single cross-section requires some important assumptions. These studies are based on intensive use of statistical methods to overcome the limitations of the data set. There are relevant questions about the validity of the critical assumptions that hold up the statistical framework (Roodman and Morduch, 2009).

Khander (2005) points out that the availability of panel data helps to eliminate one potential source of bias in the PK and Morduch cross-section studies. This source concerns unobserved but *fixed* attributes simultaneously influencing microcredit borrowing and other outcomes of interest. In his study, Khander analyzes three different outcomes: household food consumption, non-food con-

[25] This aspect will be strongly criticized by Roodman and Morduch (2009).

[26] Concerning the process of data collection, during the years 1991 and 1992 the World Bank and the Bangladesh Institute of Development Studies surveyed nearly 1,800 households in 87 villages in Bangladesh. Of these, 15 were not served by microlenders. Then, in 1998 and 1999, researchers returned to analyse the same sample. The only change is that all villages were served by microfinance services. Because of the attrition, 1,638 were the households interviewed in both rounds (Armendàriz, Morduch, 2010).

sumption and total consumption. Despite PK, he includes households owing more than 5 acres.

Following Khander's carefully study, we now examine in detail the poverty effects of microcredit intervention. In his analysis, the author distinguishes between moderately poor and extremely poor (as argued in Khander, 1998, he defines moderate poverty as household consumption level below 5,270 taka/person/year and extreme poverty as 80% of that, 3,330 taka/person/year)[27].

Table 1 contains data from Bangladesh context in Khander(2005).

We can note, from this survey, that in program villages microcredit participants had relevant declines in poverty rates (in terms of moderate poverty). Concerning the voice "All program participants", we can see a decrease from a rate of about 90 percent in the first period to about 70 percent in the second round of data, approximately a 20 percentage point decline. But for eligible households that did not participate in credit programmes (in the period 1991/92) the falling in terms of poverty rate was 19.2 percentage points, as nontarget nonparticipants (19 percent). At this point, the question is: did microcredit intervention play a role in this process? Pessimists may answer that the decline of average poverty incidence for program participants might have happened even without intervention. On the other hand, optimists think that the net impact of programmes have been substantial, involving also nonparticipants. In other words, they argue that the spillover effect could explain the general improvement in areas with programs. Comparing the results for program areas to results for villages without intervention in 1991/1992, we can note a similar trend: poverty rates decreased by around 19 to 20 percentage points (only target nonparticipants had a poverty reduction of 4.5 percentage points).

In his conclusions, Khander argues that microcredit contributed to approximately one third to one half of these poverty contractions. Specifically, Khander asserts that lending 100 taka to a female leads to a growth in household consumption by around 8 taka annually. Moreover, he suggests that the impact of microfinance intervention is stronger for extreme poverty than moderate poverty, and microcredit effects are more relevant for women borrowing than for men borrowing.

[27] This classification is based on the poverty line criterion.

4.4 The Roodman-Morduch revisitation.

The Roodman-Morduch approach to PK's analysis affirms, as Morduch (1998), that the baseline assumptions used in the study do not hold. Morduch (1999) does not find evidence of consumption impacts in the 1991/1992 data and criticizes the identifying assumptions of the PK's framework. On the other hand, he suggests that microcredit decreases consumption volatility. In their replication, Roodman and Morduch (2009) use the same methods to the same data as in PK. Applying two-stages least-squares (2SLS) regression, they contest the positive results obtained in the previous study. Concerning the PK findings, Roodman-Morduch achieve opposite (in sign) results. On the basis of specific tests , they argue that instrumentation strategy is failing. Reverse or omitted-variable causation drives the final outcomes, and the endogenous links between credit and consumption varies by subsample (i.e., borrower sex). It can explain the differences in terms of gender impact. But they do not conclude that microcredit does not affect the lives of poor borrowers; rather, they suggest that the statistical setting is not due to the task.

Roodman and Morduch findings about Khander's analysis reduce the confidence that the key identifying assumptions for causal inference effectively hold in a similar context. In particular, they doubt that the Khander's assertion concerning the relevant impact of microcredit services in reducing extreme poverty could arise from a direct estimation. The critical point is that the introduction of the panel framework does not seem to overcome the problem of the lack of clearly exogenous variations in the use of microcredit. The distinction between moderately poor and extremely poor conducted by the author is based on estimated baseline poverty levels. Then he compares changes in consumption (about the two kinds of poor households) using regression coefficients that would have sense only if all the households have similar net impacts.

All the three studies (Pitt-Khander 1998, Morduch 1998, Khander 2005) tend to reinforce the general positive idea about microcredit. According to PK (1998) and Khander (2005), microloans can reduce poverty, especially for women. In addition, Morduch (1998) suggests that small-size credits help households to attenuate consumption variability across the seasons. Roodman and Morduch (2009) do not contradict these considerations, but highlight the absence of decisive statistical evidence in support of these studies.

5. Randomized approach

5.1 Randomization as potential solution of selection bias: analytical foundations.

As discussed in section 3, selection bias represents a relevant problem in order to obtain a clean estimation of the microcredit program impact.

The aim, in this particular context, is to gauge the net effect of credit access on the revenue of a borrower. The causal impact, that is the term $E(Y_i^1 - Y_i^0/T_i = 1)$, is not observable but as argued in Duflo et. al. (2007) is "logically well defined".

Randomized experiments allow to manage selection bias problems through a random assignment of the program to the treatment and comparison groups.

How does this methodology function in practice?

First, a sample of N individuals, or households, is selected from a population of interest[28]; second, the initial sample is randomly divided into two distinct subgroups: treatment and comparison or control groups.

The target units are exposed to the "treatment" (i.e. they receive the loan), while the non-target units aren't. Then researchers observe the outcome Y, and compared the results for both different subgroups.

As argued in section 3, we can estimate the average program impact as follows:

$$\alpha = E(Y_i^1/T_i = 1) - E(Y_i^0/T_i = 0) \qquad (8)$$

The advantage in adopting this approach is that the two groups are expected to be identical before the microcredit program, because of their random selection. The only difference is due to the exposure of the treatment. This implies that the selection bias term $E(Y_i^0/T_i = 1) - E(Y_i^0/T_i = 0)$ is equal to zero.The term in question describes how both participant group and control group would have performed if nobody had had access to credit. Moreover, if the outcomes of a subject are unrelated to the treatment of the other individuals[29], we obtain

[28] It's important to highlight that the population of interest can't be randomly drawn from the whole population, but could be selected on the basis of observable attributes (Duflo et al. 2007).

[29] See Angrist, Imbens, Rubin (1996)

$E(Y_i^1/T_i = 1) - E(Y_i^0/T_i = 0) = E(Y_i^1 - Y_i^0/T_i = 1) = E(Y_i^1 - Y_i^o)$ (9)

If randomization has been completed with success, experimental approaches can provide a valuable instrument in order to overcome selection bias problems.

In the opposite case, we can meet with problems. For instance, the individuals that apply for loans successfully may have more business ability, organizational skills and entrepreneurial experience than subjects that don't apply for a microcredit program. Besides, the choice of the location from a microfinance institution may be addressed to the villages with good life and economic conditions respect to other more disadvantaged sites.

Many estimation problems concern the cases of "non-random" attrition (the less promising clients are the first to drop out of the program) and contamination (for instance, a new competitor starts his financial activity during the study period).

Generally, the bias of impact estimation is upward. But there are also particular forms of selection bias, such as contamination, which may lead to downward biases; performing a correct randomization implying that $E(Y_i^0/T_i = 1) = E(Y_i^0/T_i = 0)$.[30]

5.*2 Randomization in microcredit impact evaluations: sample-size and the power of experiments.*

In general, two particular factors affect the success of an experiment: statistical power and the role of spillovers (Armendàriz, Morduch, 2010).

Duflo et al. (2007) suggest the basic principles of power calculation[31], starting from a simple regression frame. The estimate of the average impact is the OLS coefficient of β in the following regression:

$$\mathrm{Y}_i = \alpha + \beta T + \epsilon_i \qquad (10)$$

Following Bloom's (1995) approach, we assume that only a possible treatment exists, and that a specific proportion P of the sample receives the treatment in question. Since we suppose that each

[30]As much as discussed above depends on the properties of expectations of linear operators (average impact).

As argued in Armendàriz, Morduch (2010, p. 296): "*But the basic set-up does not permit us to say anything about the medians and very little about the distributional features of impacts. And we need to be careful in analyzing data on the impacts for particular subgroups in a population.*"

[31]See also Bloom (1995).

unit was randomly sampled from an identical population, the observations can be considered to be *identical independent distributed* (*i.i.d.*), with variance σ^2.

Clearly, the variance of the OLS estimator of β, that is $\hat{\beta}$, is determined by:

$$\frac{1}{P(1-P)}\frac{\sigma^2}{N} \qquad (11)$$

Now, we focus our attention on testing the hypothesis H_0, that is the impact of the treatment is equal to zero against the alternative hypothesis, H_1, that it is not.

The significance level related to a specific test describes the likelihood of rejecting the null hypothesis when it's true.

Figure 1: Statistical Power

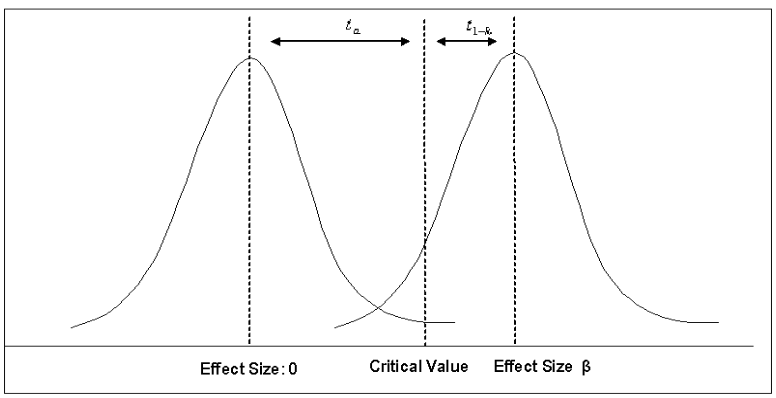

Figure 1 draws two distinct bell shaped curves: on the left there is the distribution of $\hat{\beta}$ under the null hypothesis of absence of impact $\mathrm{H}_0 = 0$, on the right the distribution of $\hat{\beta}$ if the true effect is effectively β.

In the first specification, If $\hat{\beta}$ falls to the right of the critical value, for a determinate level of significance, the hypothesis $\mathrm{H}_0 = 0$ will be rejected; formally, it is true if

$$\mid \hat{\beta} \mid > t_a * SE_{\hat{\beta}} \qquad (12)$$

where t_a hinges on the significance level and it derives from a standard t-distribution.

In the second case, in order to evaluate the *power* of the test for a true impact size β we take into account the part of the area under this curve which is to the right of the critical level t_a. In other words, it corresponds to the probability of rejecting the null hypothesis when it's effectively false.

The achievement of a power k implies the following condition:

$$\beta > (t_{1-k} + t_a)\, SE\left(\hat{\beta}\right) \qquad (13)^{32}$$

Duflo et al. (2007) consider the issue of power as the "minimum detectable effect size" for a given statistical power (k), significance level (a), sample size (N), and set of individuals that belong to the treatment group (P). It can be given by

$$MDE{=}(t_{1-k} + t_a) * \sqrt{\frac{1}{P(1-P)}}\sqrt{\frac{\sigma^2}{N}} \qquad (14)^{33}$$

where the term t_{1-k} represents the level of statistical power, t_a captures the confidence level, P includes the portion of sample that received the program, σ^2 is the variance of the impact and, finally, N is the total size of the sample.

From equation (14), we note a trade-off between statistical power and sample size. In fact, when N increases, the minimum detectable effect size decreases, and vice versa. Power calculation describes the linkage between impact size and sample size, with typical statistical confidence levels (5 percent, 10 percent, 20 percent). But in general the effect size is not known before starting with the intervention. Many different studies have developed potential solutions to overcome such problems: the first practical approach consists in making predictions on the basis of previous researches, the second concerns the introduction of a small pilot study.[34]

Cohen (1988), for instance, express the impact size in terms of standard deviations from the mean of the outcome; he suggests, in his analysis, that an impact of 0.2 standard deviation is negligible, 0.5 is intermediate, and 0.8 is relevant. Clearly, these values are indicative and have to be considered in each specific context.

Equation (14) also provides useful advice concerning the division of the sample between treated and non-treated units. We assume, for example, that there is a single treatment, and the most relevant cost of the study is data collection. It follows that an equal distribution between treatment and control group is an optimal allocation (in this case, the equation above is minimized at P=0.5).

But if program implementation is costly and the data are easily available for both the groups (the data collection process isn't ex-

[32] The term t_{1-k} is simply given by a t-table.

[33] It refers to a single sided test. If we introduce a two-sided test, the term t_a will be substituted by $t_{a/2}$.

[34] For a more advanced discussion, see Duflo et al. (2007, pp 22-24).

pensive), the optimal division will require a larger control group. Specifically, in order to obtain the optimal proportion of treated units, equation (14) must be minimized under the budget constraint as follows:

$$\underset{P}{Min} \quad \left[(t_{1-k} + t_a) * \sqrt{\frac{1}{P(1-P)}}\sqrt{\frac{\sigma^2}{N}}\right] \qquad (15)$$

$$\text{sub} \qquad Nc_c + NPc_t \leq B$$

where N is the total sample size, c_c represents the unit cost per comparison subject, and c_t is the unit cost for treatment subject. From (10), we obtain the optimal allocation rule:

$$\frac{P}{1-P} = \sqrt{\frac{c_c}{c_t}} \qquad (16)$$

As argued in Duflo et al. (2007), "*the ratio of subjects in the treatment group to those in the comparison should be proportional to the inverse of the square root of their costs.*"

The above setting can be applied to sample size calculations when there are multiple treatment.[35]

The level of randomization represents a crucial matter for the sample size. Many experimental designs involve randomization over groups rather than single individuals. For example, PROGRESA[36] program used the village as the unit of randomization, even if single individual data were available for statistical evaluation.

In similar contexts, we have to consider that the error term may not be independent across single individuals. If the program participants in a group have some attributes in common, information about single units will cause weak variations in the final result compared to the case of individual-level randomization. Since borrowers of the same group can be subject to common shocks, a correlation between outcomes is very likely, and leads to a wrong interpretation of the program impact.[37]

The key issue is to compare the proportion of the outcome variance coming from the group impact and the proportion coming from individual impact: if the first value is higher, also the sample needs to be bigger, or, alternatively, the effective size necessary for detection.

[35] See Bloom (1995).

[36] *PROGRESA* (*Programa de Educaciòn, Salud y Alimentaciòn*) is an anti-poverty program implemented in Mexico in the late 1990s (for more information, see http://www.ifpri.org/dataset/mexico-evaluation-progresa)

[37] For an analytical treatment, see Bloom (2005)

One important difficulty encountered when we operate in the microfinance context is that some experimental designs only encourage subjects to participate in a specific credit program (treatment) so that "eligible" people can accept or refuse the invitation. At the same time, people in the control group may take up the treatment even if they are not directly encouraged (Armendàriz, Morduch, 2010).

Ashraf, Karlan, and Yin (2006) encouraged a randomly selected group of people to sign up a new savings account. In this framework, the element of randomness was the invitation; this implies that the effects of the program must be evaluated comparing invited and non-invited subjects. Clearly, it follows that not all invited individuals accepted the proposal. Since the impact measured at the *invitation level* is reduced, a bigger sample size is required.

Finally, a stratification (or block) of the sample can be introduced in order to improve estimate precision. Stratifying involves dividing the sample into subgroups that have similar values of particular observable characteristics. Then, the randomization is conducted for each single block (subgroup) separately.

While the randomization procedure ensures that treatment and comparison groups will be similar in terms of expectation, stratification process is used in order to ensure that the assignment to a specific group (treatment or control group) is random in practice, along observable dimensions of the stratification.

A stratified design allows us to gauge the impact of the program for each subgroups separately using statistical methods. Since each block tends to be more homogeneous compared to the entire sample, a little change in the outcome levels can be found out with the same sample size. The relevant consequence is that a smaller total sample is needed[38].

5.3 Spillover effects

Experimental designs can make externalities such that non treated individuals are affected by the intervention. Moreover, there are spillovers also when an individual transfers from the treatment group to the comparison group, or vice-versa.

Following Duflo et al. (2007), we consider a simple case in which a microcredit program is randomly attributed across a population of

[38] For a more advanced discussion, see Imbens, King, and Ridder (2006).

individuals. The estimate of the intervention is $\alpha = E(Y_i^1/T_i = 1) - E(Y_i^0/T_i = 0)$; if we interpret this difference as the impact of the treatment, "*the potential outcomes for each individual are independent of his treatment status, as well as the treatment group status of any other individual.*[39]"

If the previous condition is violated, the term $\hat{E}(Y_i^0/T_i = 0)$[40] in the sample does not correspond to $E(Y_i^0/T_i = 0)$ in the population, because the sample contains at the same time treated and non treated units. This implies that the potential outcome for each subject and the intention-to-treat estimate (α) hinge on the entire set of allocations to participants and non participants on the program. If we assume, for instance, that the externalities on untarget individuals are positive, the estimate of the term α will be smaller than it would have been in absence of externalities.

If the spillover effects become relevant in the impact study, researchers can design experiments *ad hoc*, in order to identify their size and power[41].

In the microfinance framework, spillovers can happen when, for example, a target individual receiving funds divides part of the money with a friend of the comparison group, or when a microcredit client participating in business training program shares the new knowledge with another borrower who wasn't selected for the training activity.

Externalities affect the experimental design at distinct levels. Individual switching between groups is not a random process, and the identification of the program effects depends on the randomness of the treatment assignment. Therefore, if an individual switches from treatment to comparison group, a problem of selection bias in the impact estimation reappears. As regards spillovers due to sharing benefits concerning only the treated status, it may reduce (or enlarge) the observable effects of the program implementation.

[39] This is the standard unit treatment value assumption (SUTVA): we assume that the potential outcomes for each individual are unrelated to the participation status of other subjects (Angrist, Imbens, and Rubin, 1996).

[40] We estimate the average treatment impact as the difference in empirical means of Y between the two distinct groups as follows

$\hat{\alpha} = \hat{E}(Y_i^1/T_i = 1) - \hat{E}(Y_i^0/T_i = 0)$

where the term $\hat{E}$ describes the sample average. Hence, as the sample size grows, the previous difference tends to

$\alpha = E(Y_i^1/T_i = 1) - E(Y_i^0/T_i = 0)$

(Duflo et. al., 2007).

[41] For a more advanced discussion, see Duflo et. al. (2007), pp. 56-58.

In similar contexts, a bigger sample is needed. Another potential solution concerns randomizing at the group level rather than the individual level.

When a microfinance istitution lends funds to a group of individuals (group lending strategy), spillover effects can affect the impact estimation. Since this procedure involves a random assignment of the status of program participation within the group, it is likely that resentment, confusion and distrust emerge from the group members and produce biased impact evaluations.

5.4 Hawthorne and John Henry effects

Another interesting aspect involving impact evaluations is that the evaluation itself may lead to changes in treatment and control groups behaviour.

Specifically, behaviour changes among the target units are called Hawthorne Effects[42], while behaviour changes among the untarget units are called John Henry effects[43].

In microcredit programs, the treated units might appreciate receiving the loan and be aware of being observed, which may lead them to modify their behaviour during the experiment. On the other hand, the control units may alter their behaviour because of their resentment towards the program participants, and decide to compete with those receiving the loan (for instance, working harder in order to show their abilities) or on the contrary, slack off.

One way to disentangle these effects from the program impacts is suggested by Duflo and Hanna (2006)[44]. The authors continued to monitor the effects of the program implementation after the duration of the experiment. If the results obtained in absence of any kind of experiment are similar to the outcomes at the beginning of the evaluation period, the results are not affected by Hawthorne effects.

[42] The term was coined by H. A. Landsberger (1950), when examining older experiments from the period 1924-1932 at the Hawthorne Works (a Western Electric Company in Chicago).

During a study comissioned in order to evaluate the effect of work conditions on worker productivity, it was suggested that workers tend to increase their production due to the motivational effect of being observed.

(http://en.wikipedia.org/wiki/Hawthorne_effect)

[43] The term John Henry effects comes from the story of John Henry trying to lay railroad track faster than the machine (Duflo et. al., 2007).

[44] For more information concerning the program, see the article "Monitoring Works: Getting teachers to come to school" (Duflo, Hanna, 2006).

Experimental designs can also be created in order to disentangle Hawthorne or John Henry effects. For example, in Ashraf, Karlan, and Yin's study (2006), the authors were concerned about the role of the marketer visits (in the SEED program[45]) to the homes of clients, emphasizing the importance of savings. They suggested that these kind of relationships may persuade individuals to improve their performances.

To overcome this inconvenience, they added a further treatment group to which they offered regular savings accounts. In other words, in addition to the clients assigned to the SEED program, a part of the comparison group was selected for the pure control group, while the other part was assigned to a third group, called the "marketing treatment" group. The units of this group receive the same marketing campaign as the SEED clients, with an exclusive difference: the marketing activity was limited to traditional savings products of the microfinance institution.

Comparing savings performances of the SEED's participants with the results of the "marketing treatment " clients, Ashraf et al. isolated the direct effect produced by the SEED program from the effect of the marketing activity.

5.5 Randomization limits

A first limit of a randomized approach arises from the estimate of an average impact of the intervention. It offers little information about the distribution of the effects, and nothing concerning the median impact. One method to improve the learning process about the distribution of impacts can be stratification.

As discussed above, the introduction of impact estimations for subgroups (from the start of the experiment) may help avoid relevant risks, such as "data mining" and the finding of spurious results. The implementation of a stratification method, both in randomized and non-randomized approaches, allows us to specify in advance which subgroups (for instance, men and women, richer and poorer clients, young and aged people) and which hypotheses may be significant, thus restricting the analysis to these.

[45] The SEED is a commitment savings product (Save, Earn, Enjoy deposits) account designed for a small rural bank in the Philippines. This kind of account requires that individuals undergo restrictions in withdrawing their funds until they reach a goal date, or that a given amount of money was deposited. (Ashraf, Karlan, and Yin, 2006).

Another relevant limitation of randomized experiments concerns the difficulty in generalising the findings to other settings, that are different from the original one. In short, they can have a high internal level of validity, but not an external one. Hence, analysts that use this approach have to replicate a specific experiment before reaching general conclusions.

Non-randomized studies usually use data coming from large geographical areas, with diversified populations and their findings are applicable to a wider range of different contexts. But often these methods are less efficient in terms of internal validity.

Randomized experiments are designed carefully, and their implementation is planned closely, but extension to a large scale can yield different results. Pilot studies[46] can be used to evaluate if a policy produces relevant impacts on a small scale, and then eventually to procede with the implementation on a wide scale.

When a randomized experiment is implemented, the initial random assignment of the treatment must be maintained during the entire period of analysis. The limits here are due to attrition and contamination. The first problem (attrition bias) is not predictable and can overestimate or underestimate the impact of intervention. Contamination may happen if, for example, a new microfinance institution starts working with a comparison group, giving financial services to the people that do not receive any kind of treatment.

Finally, randomized approaches impose their logic on the selection of program participants. This implies that not the entire population will receive the treatment, and the choice of who obtains the loan cannot be made based on fairness assumptions (such as "those who need it most"). Several nongovernment organizations may be unwilling to accept this statistical method. (Armendàriz, Morduch, 2010).

6. Experimental versus Non-experimental: a comparison

Recently, a growing amount of literature has tried to use randomized experiments to validate non-experimental methods.

Lalonde (1986) suggests that many statistical procedures used in impact evaluations did not lead to precise estimates and often these differ substantially from experimental results. Glazerman, Levy,

[46] Pilot study represents the phase before the program is scaled up. This is an occasion for researchers to assess the effectiveness of the program and a chance to improve the experimental design. (Duflo et al., 2007)

and Myers (2003) compared experimental and non-experimental approaches in different fields: welfare, job training and employment service programs in the US. They found, by the examination of twelve distinct design replication studies, that retrospective estimators frequently lead to outcomes strongly different from randomized evaluation. In addition, the bias can be significant[47].

Interestingly, Cook, Shadish, and Wong (2006) compare randomized and non-randomized studies, especially in educational settings. They note that the results coming from the two different approaches are similar when analysts use a regression discontinuity or "interrupted time series" (as non-experimental strategy). The authors conclude that quasi-experimental designs (for instance regression discontinuity) might generate clean impact estimates as a well-done experimental evaluation.

Diaz and Hanna (2006) focus on PROGRESA, an anti-poverty program implemented in Mexico. Specifically, they compare estimates from an experimental design with those obtained by propensity score matching, and conclude that the second strategy can be valid, when a consistent number of control variables is available.

Duflo, Kremer, and Robinson (2006) conducted an interesting study in which they compared experimental and non-experimental estimates of peer effects over fertilizer adoption in Kenya. They found that there is a correlation between an individual's decision and the decisions of other contacts, probably due to the sharing of the same environment. In this context, randomization ensured exogenous variations in the likelihood that some members of a specific network adopted the new technology. The study shows substantially different results from the non-experimental framework; in particular, the authors highlight no evidence of learning effect.

Comparative studies can be useful in order to assess the existence and the size of biases in retrospective impact evaluations, providing a valuable guidance to address estimation problems.

Clearly, it is important that retrospective evaluations are conducted before the results of experimental approach are available, in order to avoid selection of plausible comparison groups and methods with the intention of matching experimental estimates (Duflo et al., 2007).

[47] For a more complete discussion, see Duflo et al. (2007).

7. Summary and Conclusions

Is microcredit really moving people out of poverty? This is the crucial question that researchers are attempting to solve. But, as discussed above, completing reliable impact evaluations is often difficult, as well as interpreting the results obtained.

Impact evaluations play a crucial role in allowing microfinance institutions to reliably assess the effectiveness of their operations on poverty reduction and for investors to learn about the more promising types of programmes to implement. But the lack of valuable studies explains the difficulty in evaluating interventions in which different clients use financial services with several degrees of intensity and participation is voluntary.

To sum up, it is possible to highlight some important issues concerning the impact evaluations framework.

(i) Microlenders tend to carefully select communities and villages in which to offer their services, and target citizens to whom to lend. Therefore microcredit programmes can be implemented in those areas which have more business opportunities, or have good-level infrastructures (highways, markets, large towns). Such criteria to select places for intervention lead to biased estimation of program impact.

(ii) Self-selection of program participant represents another important source of bias. Since it is probable that households with greater entrepreneurial spirit and organizational skills are more likely to join the microcredit program, this can produce biases in estimating the effects of the program (specifically, when these kinds of unobservable attributes are correlated with obtaining credit).

(iii) The standard methods applied in microcredit impact evaluations tend to consider the intervention as a homogeneous process, that produces a steady influence on outcome variables like consumption, income or poverty. In practice, the causal process is more complicated. For instance, there might be hidden relationships between the selected exogenous and endogenous variable, or externalities, which may obscure the causation linkage.

Recently, a set of interesting impact evaluations have incorporated experimental designs into the program implementation to achieve reliable estimates of the impact net effects. In my opinion it represents a possible strategy to follow, keeping in mind that both approaches have to be made with care.

If we focus our attention on the hundreds of millions of microfinance clients, they demonstrate how important microcredit is in their lives by a behavioural mechanism called "voting with their feet". According to Rosenberg (2010), the microfinance experience in three decades shows that customers arrive without a need to advertise and repay loans with high reliability rates. The key incentive to repay is the "*borrowers' desire to keep access to a highly valued service*" (Rosenberg, 2010, p. 4). It seems that target borrowers wish to preserve the microcredit access over time, in order to cope with possible future shocks. Another observable aspect concerns the willingness to pay high interest rates on the funds received, and accept trivial or no return on savings. In addition, microfinance customers usually return using micro-financial services, even in the presence of high desertion rates. The repeated use does not necessarily imply that these services provide benefits. It is clearly possible that some borrowers will over-indebt themselves, producing a negative final outcome. But the high repayment rates over the long term may justify the assertion that microfinance is not over-indebting great amounts of its users. Surely this presumption needs to be studied and tested by new research.

Finally, as suggested by Rosenberg (2010), an important advantage of microcredit introduction can be thought not as its net impact, but rather as its cost in term of subsidies. In other words, while social programs such as education and health care might require consistent amounts of continued subsidies, microcredit, if well-implemented, can require relatively small initial subsidies that allow such organizations to expand their business without any further kind of subsidy.

References

- ACCION Foundation web-site. "Group lending."
(http://www.accion.org/Page.aspx?pid=257#Group_Lending)

- J. D. Angrist, G. W. Imbens, and D. B. Rubin. 1996. "Identification of Causal Effects using instrumental variables". *Journal of the American Statistical Association*, 91, 444-455.

- B. Armendàriz, J. Morduch (2010). *The Economics of Microfinance.* 2nd edition, Cambridge, Mass.: Mit Press.

- N. Ashraf, D. S. Karlan, and W. Yin. 2006. " Tying Odysseus to the Mast: evidence from a Commitment savings product in the Philippines." *Quarterly Journal of Economics.*

- A. Banerjee , E. Duflo, R. Glennerster, and C. Kinnan. 2009. "The miracle of microfinance? Evidence from a randomized evaluation." Cambridge, Mass.: MIT Poverty Action Lab, May.

-T.J. Bartik , and Bingham R.D..1995. "Can economic development programs be evaluated?". W.E. Upjohn Institute for Employment Research, Kalamazoo MI.

- H. S. Bloom. 1995. "Minimum detectable effects: A simple way to report the statistical power of experimental designs." *Evaluation Review*, 19, 547-56.

- H. S. Bloom. 2005. "Randomizing groups to evaluate place-based programs". NY: Russell Sage Foundation, chap. Learning more from social experiments, pp. 115-172.

- J. Cohen. 1988. "Statistical power analysis for the behavioural science." Hillsdale, NJ: Lawrence Erlbaum, 2nd edition edn.

- B. Coleman. 1999. "The impact of group lending in northeast Thailand." *Journal of Development Economics*, 60: 105-142.

- D. Collins, J. Morduch, S. Rutherford, and O. Ruthven. 2009. *Portfolios of the Poor: How the World's Poor live on $2 a day.* Princeton, N.J.: Princeton University Press.

- T. D. Cook, W. R. Shadish, and V. C. Wong. 2006. "Within study comparisons of experiments and non-experiments: can they help decide on evaluation policy." Mimeo, Northwestern University.

- R. Davidson, J. MacKinnon. 2004. "Econometric Theory and Methods." New York: Oxford University Press.

- J.J. Diaz, S. Handa. 2006. "An assessment of Propensity Score Matching as a non-experimental impact stimator: Evidence from Mexico's PROGRESA program." *The Journal of Human Resources*, Vol. 41, N. 2, pp.319-345.

- E. Duflo, R. Glennerster, and M. Kremer. 2007. "Using randomization in development economics research: A toolkit." In T. Paul Schultz and John Strauss, eds., *Handbook of Development Economics*, Vol. 4. Amsterdam: North-Holland.

- E. Duflo, R. Hanna. 2006. "Monitoring works: getting teachers to come to school." NBER Working Paper No. 11880.

- E. Duflo, M. Kremer, and J. Robinson. 2006. "Understanding technology adoption: fertilizer in western Kenya, Preliminary results from field experiments." Mimeo.

- P. Dupas, J. Robinson. 2009. "Savings constraints and microenterprise development: evidence from a filed experiment." Working paper 14693. Cambridge, Mass.: National Bureau of economic research, January.

- S. Glazerman, D. Levy, and D. Myers. 2003. "Nonexperimental replications of social experiments: a systematic review". Princeton, NJ: Mathematica Policy Research, Inc.

- Grameen Foundation web-site. "Microcredit taxonomy."
(http://www.grameen-info.org)

- W. Greene. 2008. "Econometric analysis." 6th edition. Englewood Cliffs, NJ: Prentice Hall.

- J. J. Heckman. 2000. "Causal parameters and policy analysis in economics: a twentieh century retrospective ." *Quarterly Journal of Economics*, 115, 45-97.

- J. J. Heckman. 1979. "Sample selection bias as a specification error." Econometrica 47(1): 153-161.

- P. W. Holland. 1986. "Statistics and causal inference." *Journal of the American Statistical Association*, Vol.81, No. 396, 358-377.

- G. Imbens, G. King, and G. Ridder. 2006. "On the benefits of stratification in Randomized Experiments." Mimeo, Harward.

- D. Karlan. 2001. "Microfinance impact assessments: The perils of using new members as a control group." *Journal of Microfinance* 3(2): 76-85.

- D. Karlan, and J. Zinman. 2008. "Expanding Microenterprise Credit Access: Using Randomized Supply Decisions to estimate the impacts." New Haven, Conn.: Innovations for Poverty Action, January.

- D. Karlan, and J. Zinman. 2009. "Expanding Microenterprise Credit Access: Using Randomized Supply Decisions to estimate the impacts in Manila." New Haven, Conn.: Innovations for Poverty Action, July.

- S. R. Khander. 1996. "Role of targeted credit in rural non-farm growth." *Bangladesh Development Studies* 24 (3 & 4).

- S. R. Khandker. 1998. "Fighting Poverty with Micro-credit: Experience in Bangladesh." New York: Oxford University Press.

- S. R. Khander. 2000. "Savings, informal borrowing and microfinance." Bangladesh Development Studies 26 (2 & 3).

- S. R. Khandker. 2005. "Microfinance and Poverty: Evidence Using Panel Data from Bangladesh." *Word Bank Economic Review,* 19(2): 263-86.

- R. J. Lalonde. 1986. "Evaluating the econometric evaluations of training programs using experimental data." American Economic Review, 76(4): 602-620.

- E. E. Leamer. 1985. "Vector autoregressions for causal inference?" Carnegie-Rochester Conference Series on Public Policy, 22: 255-304.

- G. S. Maddala. 1983. "Limited-dependent and qualitative variables in econometrics." Cambridge, UK: Cambridge University Press.

- S. McKernan. 2002. "The impact of microcredit programs on self-employment profits: do non-credit program aspects matter?" *The Review of Economics and Statistics* 84(1): 93-115.

- N. Menon. 2005. "Non-linearities in returns to participation in Grameen Bank programs." *Journal of Development Studies* 42(8): 1379-1400.

- Microcredit Summit web-site. "Microcredit definition." (http://w.w.w.microcreditsummit.org)

- R. Moffit. 1991."Program evaluation with nonexperimental data." *Evaluation Review* 15: 291-314.

- J. Morduch. 1998. "Does microfinance really help the poor? New evidence on flagship programs in Bangladesh." Draft, MacArthur Foundation project on inequality working paper, Princeton University.

- A. Nair. 2005. "Sustainability of microfinance self-help groups in India: would federating help?". World Bank Policy Research Working Paper 3516.

- M. Pitt , 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 Economics*, 106(5): 958-996.

- M. Pitt. 1999. "Reply to Jonathan Morduch's 'Does microfinance really help the poor? New evidence from flagship programs in Bangladesh.'"Typescript, Department of Economics, Brown University.

- M. Pitt , and S. Khandker. 2002. "Credit programs for the poor and seasonality in rural Bangladesh." Journal of Development Studies 39(2): 1-24.

- M. Pitt, S. Khandker, O. H. Chowdhury, and D. L. Millimet. 2003. "Credit programs for the poor and the health status of children in rural Bangladesh." *International Economic Review*, Vol. 44, No. 1, pp.:87-118.

- M. Pitt , S. Khandker, and J. Cartwright. 2006. "Empowering women with microfinance: evidence from Bangladesh." Economic Development and Cultural Change 54(4): 791-831.

- D. Roodman, J. Morduch. 2009. "The impact of Microcredit on the Poor in Bangladesh: Revisiting the evidence." Working Paper n. 174, June 2009, Centre for Global Development.

- D. Roodman. 2009. "Estimating fully observed recursive mixed-process models with cmp." Working paper 168. Washington, DC: Center for Global Development.

- R. Rosenberg. 2010. "Does Microcredit really help poor people?" Focus note 59. Washington , D.C.: CGAP.

- J. Tobin. 1958. "Estimation of relationships for limited dependent variables." *Econometrica* 26: 24-36.

- K. Tsui. 2002. "Multidimensional Poverty indices." *Social Choice and Welfare*, 19: 69-93.

- Wikipedia. 2010. "Upazilas of Bangladesh."
(http://en.wikipedia.org/wiki/Upazilas_of_Bangladesh).

- Wikipedia. 2010. "Mexico evaluation Progresa."
http://www.ifpri.org/dataset/mexico-evaluation-progresa)

- Wikipedia. 2010. "Hawthorne effect."
(http://en.wikipedia.org/wiki/Hawthorne_effect)

- Wikipedia. 2011. "Village Banking."
(http://en.wikipedia.org/wiki/Village_Banking)

- J. Wilde. 2000. "Identification of multiple equation probit models with endogenous dummy regressors." Economics Letters 69: 309-12.

- C. Winship, and R. D. Mare. 1992. "Models for sample selection bias." *Annual Review of Sociology*, Vol.18, pp. 327-350.

TABLES

Table 1: Poverty incidence on the basis of program participant status and survey period

	Moderate Poverty			*Extreme Poverty*		
Program Participation Status	*1991/92*	*1998/99*	*Difference*	*1991/92*	*1998/99*	*Difference*
Program villages						
Program participants (targeted)	90,3	75,5	14,8	57,3	36,8	20,5
Program participants (mistargeted)	82,3	57,2	25,1	37,8	22,9	14,9
All program participants	90,3	70,1	20,2	52,5	32,7	19,8
Target nonparticipants	91,1	72,0	19,1	58,9	44,0	14,9
Nontarget, nonparticipants	69,8	50,8	19	23,6	19,3	4,3
Total	83,7	65,5	18,2	45,0	31,4	13,6
Nonprogram villages						
Program participants (targeted)*	89,3	79,0	10,3	60,2	51,6	8,6
Program participants (mistargeted)*	93,0	61,2	31,8	51,4	32,8	18,6
All program participants*	90,8	71,6	19,2	56,6	43,8	12,8
Target nonparticipants	87,4	82,9	4,5	57,0	51,2	5,8
Nontarget, nonparticipants	72,7	53,2	19,5	35,5	26,0	9,5
Total	80,3	67,7	12,6	46,6	38,3	8,3

**Program and nonprogram villages are based on 1991/1992 program placement.*
All villages receive program by 1998/1999.
Source: Khander (2005), table 7. Armendàriz, Morduch (2010), table 9.1.

Table 2: Characteristics of the principal microcredit impact studies

Study	Methodological approach	Data	Critical aspects	Results
Khander, S. (1998)	-Non randomized evaluation; - Double difference between eligible and ineligible households and between program and non program communities.	-Bangladesh (Grameen, BRAC); - World Bank and Bangladesh Institute of Development Studies survey (cross-section, years 1991/1992)	- Eligibility criteria are not clearly exogenous (Microcredit participation is restricted to "functionally landless" - households owning less than 0.5 acre of land).	- roughly 5% of households involved in the program lift out from poverty annually; - lending 100 taka to a female leads to additional consumption of 18 taka annually
Pitt, Khander (1998)	-Non randomized evaluation; - Double difference between eligible and non-eligible households and program with and without microcredit interventions. - Estimation process is conducted separately for male and female clients.	-Bangladesh (Grameen, BRDB*, BRAC); -World Bank and Bangladesh Institute of Development Studies survey (cross-section, years 1991/1992)	- Eligibility criteria are not clearly exogenous (Microcredit participation is restricted to "functionally landless" - households owning less than 0.5 acre of land). - No evidence of consumption impacts (Morduch, 1998). - Identifying assumptions do not hold (Morduch, 1998; Roodman, Morduch, 2009).	- Positive impact on microcredit participation on weekly expenditure per capita, women's non-land assets and women's labor supply. - Positive effect of female participation in Grameen Bank on schooling of girls. - Microcredit intervention can contribute to change attitudes and general characteristics of the villages.
Coleman (1999)	- Non randomized evaluation. - Double difference comparison between participants and Non-participants and between villages with program and villages in which program will be introduced later.	- Thailand (village bank). - 445 households in 14 villages: 8 with village banks that started their activity in 1995, the other 6 an year later.	- Control group made up of self-selecting participants in villages identified for later inclusion in the program. - Generally, Thai-villagers are wealthier than Bangladeshi-villagers and have easier access to credit. - The results cannot be extended to a large context.	- No evidence of program impact. Village bank membership does not impact on income variables.
Pitt et al. (2003)	- Non randomized evaluation. - Maximum likelihood estimation controlling for endogeneity of individual microcredit participation and of the placement of microfinance intervention. - Impact variables in the study are health status of male and female measured through the following indicators: arm circumference, body mass index and height-for-age.	- Bangladesh (BRAC, BRDB, Grameen Bank); Panel data from World Bank and Bangladesh Institute of Development Studies (years 1991/1992; 1998/1999).	- Doubt on reliability of identifying assumptions.	- Positive effect of female credit on height-for-age and arm circumference of both boys and girls. Male credit does not produce a significant impact on health children measures.
Khander (2005)	- Non randomized evaluation. - Fixed-effects estimator in order to drop out the fixed (and unobservable) individual-specific and village characteristics.	- Bangladesh (BRAC, BRDB, Grameen Bank); Panel data from World Bank and Bangladesh Institute of Development Studies (years 1991/1992; 1998/1999).	- Key identifying assumptions for causal inference do not hold (Roodman, Morduch, 2009). - Introduction of panel data does not solve the lack of clearly exogenous variations in using microcredit. - Microcredit impact in reducing extreme poverty does not arise from direct estimation (Roodman, Morduch, 2009).	- Lending 100 taka to a female leads to additional consumption of 8 taka annually. - Microcredit contributes to approximately one third to one half to reduce poverty. - Microcredit effects are stronger for extreme poverty.
Ashraf, Karlan, Yin (2006)	- Randomized evaluation. (Microsavings). - SEED* program is randomly assigned about 1.800 subjects to either receive a proposal to open a savings account (treatmetn units) or not (comparison units). A random subset of 707 individuals received the program offer, but only 202 accepted the proposal and opened the account (28,4 percent).	Baseline household survey conducted by the authors of 1.777 existing or former clients of Green Bank (rural bank of Philippines).	- Microsavings impact measured at invitation level (not all invited-subjects accepted the offer - 202 of 707): this requires a bigger sample size.	- In the initial program period, strong positive effect on savings. - Significant increase in women's decision-making power and positive impact on purchase of durable goods.
Karlan, Zinman (2008)	- Randomized evaluation. - A "randomizer" software encouraged loan officers to reconsider randomly selected marginal rejects. Specifically, the randomizer's treatment can be though as "encouragement to reconsider" rather than "randomized approval", because of loan officer assessment.	The Lender* operated in South Africa (Capetown, Port Elizabeth and Durban areas). The sample concerns marginal applicants, with three characteristics: new, rejected but potentially creditworthy (787 applicants).	- People are not very poor. (marginally rejected applicants) - Interest rates are higher than those applied by typical microfinance institutions. - Only consumer loans. - Short-run results (6-12 months)	- High level of stress and depression. - Positive effects on job retention, income, food consumption quality and quantity, household decision-making control and mental outlook.

Study	Methodological approach	Data	Critical aspects	Results
Dupas, Robinson (2009)	- Randomized evaluation (Microsavings, Kenya). - Sampled subjects were randomly divided into treatment and control groups, stratified by gender/occupation. Then the treated units received a proposal to open a savings account at no cost, while the control units did not benefit from any assistance in opening the same account.	- Four sources of data: 1) survey conducted by authors about baseline characteristics of participants (household composition, marital status, health, etc.). 2) Administrative data from the village bank (Bumala, Kenya). 3) Cognitive ability measures and time and risk preferences directly elicited from respondents. 4) Detailed data collected on respondents through daily self-reported logbooks (including information about income, expenditure, health, income shocks, etc.).	- Random sample is small (185 microentrepreneurs). - Target area is limited to a single site and to a single microfinance institution.	- Access to saving account had a positive impact on female productive investments. - Increase in income levels for female entrepreneurs. - High rate of return to capital for women, rougly 5.5% per month at the median.
Duflo et al. (2009)	- Randomized evaluation. - Random selection of 52 of 104 similar poor areas as location for new branches of Spandana (in the Indian city of Hyderabab).	- Spandana (India). Baseline survey conducted in 2005 (120 areas and 2.800 households). Sixteen areas were dropped from the study before randomization (because of large presence of migrant-worker households). Hence, only 104 areas were available for the analysis. - Spandana does not specified loan purpose.	- The potential borrowers in selected areas are poor, but not "the poorest of the poor". - Short-run results (15-18 months).	- No impact on measures of health, education and women's decision making. - Households with an existing business at program time increase investment in durable goods; households with high propensity to start new activity increase durable goods spending, while those with low propensity rise nondurable spending.
Karlan, Zinman (2010)	- Randomized evaluation. - Quantitative Credit scoring is used to randomize the approval decision for marginally creditworthy applicants.	- First Macro Bank (Philippines, Manila) - The sample is made up of 1.601 marginally creditworthy applicants to which 1.272 were assigned to the treatment group and 329 to the control group.	- People are not very poor. Using credit scoring to randomize allows to measure only treatment effects on the marginally creditworthy.	- Treated group tends to shrink their businesses compared to the control group. - No evidence that increasing access to credit improves subjective well-being. - Positive impact of expanding credit access on male profits. - Male microentrepreneurs use increased profits to invest in human capital of their kids.

Source: personal elaboration on the basis of literature referred to.
Note: I also introduce in this prospect microsavings programmes. A great deal of microcredit programmes include, in addition to a mere distribution of the loans, savings services.
* BRDB corresponds to Bangladesh Rural Development Board, a public sector organization working for rural development and poverty alleviation in Bangladesh.(http://www.brdb.gov.bd).
* SEED is a banking product designed to help clients save by binding their funds until they reached a specific savings purpose.
* The Lender in question was merged into a large bank holding company in 2005; it does not exist as a single entity. (Karlan, Zinman, 2010).

Recent working papers

The complete list of working papers is can be found at http://polis.unipmn.it/pubbl

***Economics Series**	****Political Theory Series**	**ε Al.Ex Series**
T Territories Series	**t Transitions Series**	**Q Quaderni CIVIS**

2011 n.180* Cristina Elisa Orso: *Microcredit and poverty. An overview of the principal statistical methods used to measure the program net impacts*

2011 n.179** Noemi Podestà e Alberto Chiari: *La qualità dei processi deliberativi*

2011 n.178** Stefano Procacci: *Dalla* Peace Resarch *alla Scuola di Copenhagen. Sviluppi e trasformazioni di un programma di ricerca*

2010 n.176** Fabio Longo and Jőrg Luther: *Costituzioni di microstati europei: I casi di Cipro, Liechtenstein e Città del Vaticano*

2010 n.175* Mikko Välimäki: *Introducing Class Actions in Finland: an Example of Lawmaking without Economic Analysis*

2010 n.174* Matteo Migheli: *Do the Vietnamese support* Doi Moi?

2010 n.173* Guido Ortona: *Punishment and cooperation: the "old" theory*

2010 n.172* Giovanni B. Ramello: *Property rights and externalities: The uneasy case of knowledge*

2010 n.171* Nadia Fiorino and Emma Galli: *An analysis of the determinants of corruption: Evidence from the Italian regions*

2010 n.170* Jacopo Costa and Roberto Ricciuti: *State capacity, manufacturing and civil conflict*

2010 n.169* Giovanni B. Ramello: *Copyright & endogenous market structure: A glimpse from the journal-publishing market*

2010 n.168* Mario Ferrero: *The cult of martyrs*

2010 n.167* Cinzia Di Novi: *The indirect effect of fine particulate matter on health through individuals' life-style*

2010 n.166* Donatella Porrini and Giovanni B. Ramello: *Class action and financial markets: Insights from law and economics*

2010 n.165** Corrado Malandrino: *Il pensiero di Roberto Michels sull'oligarchia, la classe politica e il capo carismatico. Dal* Corso di sociologia politica *(1927) ai* Nuovi studi sulla classe politica *(1936)*

2010 n.164[ε] Matteo Migheli: *Gender at work: Productivity and incentives*

2010 n.163[Q] Gian-Luigi Bulsei and Noemi Podestà (Eds): *Imprese differenti. Le organizzazioni cooperative tra crisi economica e nuovo welfare*

2010 n.162* Claudia Cusinello and Franco Amisano: *Analysis for the implementation of a sustainable transport model in the eastern Piedmont county of Alessandria, Italy*

2010 n.161* Roberto Ricciuti: *Accumulazione del capitale e crescita economica tra Italia liberale e regime fascista*

2010 n.160* Carla Marchese and Giovanni B. Ramello: *In the beginning was the Word. Now is the Copyright*

2010 n.159[ε] Peter Lewisch, Stefania Ottone and Ferruccio Ponzano: *Free-riding on altruistic punishment? An experimental comparison of third-party-punishment in a stand-alone and in an in-group environment*

2009 n.158* Rongili Biswas, Carla Marchese and Fabio Privileggi: *Tax evasion in a principal-agent model with self-protection*

2009 n.157* Alessandro Lanteri and Stefania Ottone: *Economia ed etica negli esperimenti*

2009 n.156* Cinzia Di Novi: *Sample selection correction in panel data models when selectivity is due to two sources*

2009 n.155* Michela Martinoia: *European integration, labour market dynamics and migration flows*