Abstract: In this article, I review evidences from several Microfinance studies. The literature shows contradictory results and sometimes fails to provide enough evidence for some effects, especially on consumption, income and empowerment of women. Nevertheless, the results provided evidence that Microfinance can help the poor to become entrepreneurs, grow their business and save more. I also discuss about different design and statistical methodologies such as cross-sectional analysis, panel data analysis, RCTs, dynamic panel data and lab experiments. I also discuss about the importance not only of a clear research design but also of statistical transparency.

Keywords: microfinance, causal inference, literature review

1. Introduction

Microfinance can be a powerful tool in order to empower the poor. Poor people usually do not have access to financial services, making them “pariahs” in the financial system. When they have access to the financial system they may be able to finance their enterprises, or to overcome financial shocks. However, since they cannot offer collateral, how can banks be assured that the poor have positive incentives and repay their loans? Moreover, how to reach to those in need?

Cull et al. (2009) emphasizes that the biggest triumph of Microfinance is to show that poor households can be good and reliable clients. Actually, some microlenders report repayment rates of 98%. However how that is possible?

The Microfinance field of study developed fast in the past forty years. Empirical papers in Microfinance are from several different perspectives. From cross-section to RCTs, from surveys to experiments, the objective here is to inspect results, the flow of ideas and methodologies applied over time. With this article I hope to provide fundamental information about the pros and cons of each method, show some previous results and to help researchers how to better address their research problems.

2. Outcomes in Microfinance Studies

First of all, a definition about impact studies must be provided. As its own name suggests, are studies that investigates if some Microfinance program was successful or not. However, in order to estimate success one must define: success measured based in what? Straightforward measures comes to mind like increase in revenues and profits, access to bank services, etc. But how to define “successful” remains cloudy.
Differently from other branches of finance, the main goal of a Microfinance institution should not be maximizing their profits. They must be able to continue their growth, aiming to help more people, however profiting from the poor can be ethically questionable (Schmidt, 2010). Thus, many researches are using “measures of success” that take into account the well being of the clients. For instance, Angelucci et al. (2015) tested if the insertion of a Microfinance program had effect in several outcomes at the borrowers' level. They found a decrease of the amount of capital borrowed from informal entity (eg.: moneylenders), a decreasing amount of income coming from governmental subsidies (such as cash transfer programs), decreases in a depression index, and a increase in loans from either that institution or a different one.

Karlan & Goldberg (2011) proposes an agenda to evaluate Microfinance impacts in three different evaluation fronts: program, process, and policy.

The first relates to a single particular program. The goal is to assess if that Microfinance institution (MFI) is effective to improve their clients welfare. Therefore the comparison is between a program versus the absence of that program. Particularly, using new programs, one may be able to see if there is significant changes in households and subjects outcomes.

The second aims to compare processes, therefore the comparison is made between two different programs. A good example of this approach is Jamison et al. (2014) which identified no difference between giving savings account plus financial training and financial training only. They tested different ways to apply the program offered. Two possible interpretations for their findings is that education and account are substitutes or that only financial education leads to a n increase in savings.

Finally, the third identify cost-benefit reasons to apply given methodology in lending and policy implications. Armendáriz & Morduch (2010) argues that is important for policymakers to know if one dollar spent in a Microfinance program diminishes poverty more than the same dollar in other initiatives. Nevertheless, policy evaluation is quite hard to establish causality, since there are many possible endogenous variables (Karlan & Goldberg, 2011). For instance, Neri & Medrado (2005) raises as a problem in their study the fact that is quite difficult to evaluate the cost-benefit of a Microfinance program in terms of public spending given that the program was financed by the government because of endogenous reasons.

Another concern is about the duration of the effects. Studies like Karlan & Zinman (2013) observed not only short-run but also long-run evidence of price sensitivity by borrowers in Mexico. However, this may not be always the case and effects may "die during time", what makes the design of a study particularly important. When designing an RCT in Microfinance the researcher should be able to measure the short-term effects on the outcome variable as well as long-term effects.

After knowing what should be evaluated, we can focus now on the problems of designs in order to be able to argue about causality and not mere correlation. In the next section, I stress how problematic can be to find good counterfactuals in Microfinance research.
3. Causality in Microfinance Studies

As noted in the previous section, it is quite an effort to claim causality in Microfinance studies. In Microfinance studies the treatment and control group are seldom the same in terms observables variables such as loan amount, landholding and in unobservable characteristics like investment opportunities and intrinsic risk. These issues raise questions about the exogeneity of the treatment and claims of causality. Thus, for non-randomized studies the researcher must use the "statistical toolbox" in order to reduce the effect of possible alternative explanations.

As Karlan (2001) notes, the usage of cross-sectional data to compare early borrowers with late borrowers on the outcomes variables may be weak due to dropouts in later periods. This leads to different sources of bias related to dropout, such as incomplete sample bias or attrition bias.

The incomplete sample bias, which is related to the fact that a researcher can only capture in the sample those that remained active in the Microfinance institution. A reason for this difference could be that those two groups were impacted differently, and that may overestimate the results. At the end, one will only be able to observe those that remained in the program but the true impact depends of entire sample: those that remained and those who dropped.

The second problem is related to attrition bias, which is independent of the outcome, but correlated to a different variable. For instance, wealthier participants may choose to stay in the program while the poorer decide to drop out. Thus, this attrition bias undermines a causal claim between a variable and an outcome.

The above-mentioned problems are common issues related to selection, since the participation is voluntary. As noticed by Coleman (2006), wealthier villagers are significantly more likely to attend in a Microfinance program than the poor, even after controlling for possible endogeneity and self-selection. Consequently, both groups are not comparable. He takes advantage that the program started one year after for six of the villages. This was critical to the evaluation because now he was able to compare who joined and who did not before the program started in the "late villages", through a quasi-experimental setup.

To address this problem with selection bias Karlan (2001) presents two solutions for attrition and sample bias. The first is to add in the veteran group those who drop out, ensuring that they are comparable to the early group. Although it is hard to do in practice, because this introduces noise to the estimations, this increases the validity of a causal claim. The second solution consists in adjusting the selection of groups in order to invite only those that are expected to remain.

An interesting approach was taken by Karlan & Zinman (2008) to provide robustness for their findings. In a study about short-term elasticity, they considered three groups as the most likely to read the bank's solicitation sent by mail. Namely, they compared the full sample with those with higher formal education level, those that borrowed in last 9 months
and those that already made two previous loans. They found that the loan amount changed in a similar amount for all these sub-samples, thus enhancing their causal claim.

Another example of the difficulty to find comparable groups is a study by Neri & Medrado (2005). They compared people from the Northwest region in Brazil, with people from other regions not attended by the CrediAMIGO program launched by Banco do Nordeste. The study uses a differences-in-differences strategy to estimate the effect of the introduction of the program. This approach is not fully exogenous and the affected/treated group is different in many observables from the unaffected/control, thus undermining claims of causality.

Spillovers are another problem, which Khandker (2005) defines as the extent in which a program benefits households beyond those that actually participated. It is especially worthy of concern when the treatment spills to the control groups. This usually happens when both treatment and control groups are in the same village or region. Using a survey conducted by World Bank in Bangladesh the author concluded that spillover effects are frequent and of significant magnitude. He used longitudinal/panel data, which is consisted of observations from the same subject over a period of time. The usage of village and household fixed effects to control for diverse invariant characteristics reduced possible sources of biases in his findings.

In other study Khandker & Samad (2014) uses data from 20 years (three points in time). The authors explored the dynamic panel data, i.e. allowing for dynamic variation in the variables to be explored with trends, time-varying effects of credit and in long-term perspective. They concluded that the effect of Microfinance changed over time. Past credit had a greater impact on income and expenditure than current loans. Thus the marginal effect of Microfinance programs tend to decrease over along period of time. The drawback of using panel methodologies is that the model is regressed using first differences, which eliminates time-invariant variables (e.g.: gender). Nonetheless, all time-invariant characteristics are controlled. Therefore, if the research question is based on a time-invariant variable other research methodologies may be used.

Leite et al. (2016) uses a hierarchical multilevel linear model to address the question if there is a difference between the financial sustainability ratios of for-profit and not-for-profit Microfinance institutions (MFIs), since in the database the profitability of an MFI was time-invariant, a normal fixed-effects panel analyses could not be estimated. The conclusion was that not-for-profit MFIs can be as sustainable as for-profit MFIs, albeit smaller for-profit MFIs charged higher interest rates than the not-for-profit ones. Nevertheless, this difference disappeared for the larger MFIs.

More recently, researchers entered in the field of randomized studies. Papers using Randomized Control Trials (RCTs) became popular and considered one of the best solutions to deal with the lack of exogeneity (Karlan et al., 2009). An RCT consists of choosing a random group of people to have access to a Microfinance program while another is the control (later access to the program). The average difference of these two groups can be considered causal evidence for the program assessment. The researcher must guarantee that neither the
acceptance to participate from the subjects nor the tendency to drop out are systematically related to the dependent variables (Armendáriz & Morduch, 2010).

However, the assignment must be random. And sometimes is really hard to guarantee it. For example a bank may choose to implement the program in certain villages based on some previous information (credit score, average income, number of people in a village, etc.), thus this decision makes the treatment and control groups not comparable. In a big review, Banerjee et al. (2015c) evaluated six RCTs. In the RCT performed in Bosnia the randomization was in individual level and is similar to the studies of Karlan & Zinman (2010) and Karlan & Zinman (2011). They used a credit scoring metric to define a range where they could manipulate randomly who would get the loan and who would not. For a clear and deep debate about RCT and its advantages see Karlan et al. (2009).

Banerjee et al. (2015a) present solutions for several of the problems here mentioned. They randomly selected 52 neighborhoods in India and observed a group lending with women that were not necessarily entrepreneurs. An interesting aspect is that they followed the borrowers for about 3 to 3.5 years after the introduction of the program. With this design they could compare both short and long term effects. Neither in short nor in long-term they found significant effects in education, health and women's empowerment. An interesting result from this study is that households do increase their consumption in durable and temptation goods like parties and festivals.

Another RCT was conducted by Soman & Cheema (2011) in which they showed that poor people can increase their informal savings by saving at two different envelopes instead of one. Moreover, when they put a photo of their family in the envelope they also saved more. Thus, the photo worked as a reminder and enhanced the motive behind saving: the goal of improving their families' lives.

There are two other experiments worth of a mention. The first is an experimental survey. Bauer et al. (2012) introduced behavioral insights in Microfinance with a field experiment in India. They found a correlation of women with present-biased preferences to be more likely to become microcredit borrowers. By bringing psychological factors to the discussion they opened a new line for debate that includes behavioral characteristics.

Another study is a lab experiment conducted in an urban market in Peru by Giné et al. (2010). They invited owners and employees to receive loans and emulate joint-liability to manage between risky and safer investments. They conducted the study for 7 months, collecting data for the ten main variables at least 29 times, thus providing a very accurate measurement. The researchers, by using the same individuals, could control for important characteristics such as innate risk (eg.: one individual may be risk-averse while another may be risk-lover). A counter-intuitive result emerged: joint-liability increases the number of risky investments choices and default when participants are free to talk each other. Although that was a lab experiment, the external validity was assured by the use of real potential borrowers.

Nevertheless, RCTs are not perfect. The researcher may gamble on luck by including several dependent variables, or increasing the sample size after doing preliminary
estimations. The p-values are reliable in a single estimation: when a researcher starts using several dependent variables the likelihood of finding a false positive increases sharply. Thus, to claim causality, the researcher not only must focus his or her attention on design, but also on the statistical methods and assumptions behind the estimated models.

3.1 RCTs in different cultures

Although RCTs do make claims of causality stronger, this claim of causality may be restricted to a certain country, region or culture. In order to make results more generalizable the researcher must choose a more representative sample. In this subsection, I discuss three recent multicultural RCTs and their results.

The first paper is the paper written by Banerjee et al. (2015c), in which they evaluated six different experiments published at AEJ: Applied Economics 7(1). These studies found that the demand for microcredit is modest (around 30% of acceptance), nevertheless the studies consistently found positive effect on both profits and revenues from the business that the borrowers ran at the time. As the authors explained: "each study finds at least some evidence, on some margin, that expanded access to credit increases business activity".

Nevertheless, not all results were promising. All the studies failed to find any significant association between Microfinance programs and increase of household income, although business income increased in the four papers that studied that outcome. Again, of the four papers that studied family consumption three found no effect, while one found a negative relationship.

The second paper was written by Banerjee et al. (2015b). They performed RCTs in six different countries with a total of 10,495 subjects: Ethiopia, Ghana, Honduras, India, Pakistan and Peru. They combined some Microfinance programs (offer of savings accounts and financial education) and cash/assets transfers.

They found an increase in assets, savings, consumption and income. Nevertheless, they failed to find any support for improvement of women's decision making power. Regarding other outcomes they found that subjects had higher food security, spent more time working, increased their political involvement and reported an increase in health measures.

Karlan et al. (2016) assessed the impact of reminders on savings behavior of poor in three countries: Philippines, Peru and Bolivia. The researchers found that receiving reminders increased the likelihood of meeting their commitment goals. However they did not found conclusive evidence that the reminders increased the amount saved. They also shown that gain or loss-framed reminders had similar impacts.
4. Conclusion

Microfinance research developed fast in recent years. Mapping several methods I summarized the pros and cons of each one briefly. I explored several methods in this paper: cross-sectional analysis, panel data analysis, RCTs, dynamic panel data and lab experiments. Also, I showed some important studies and their main results. I hope that with the information presented in this paper the reader may be able to have a more solid perspective about the problems that a researcher will face conducting empirical Microfinance research. In addition this paper provides an important overview of findings in the Microfinance literature.

Discussing the results presented in this paper one can see that they are mixed. The literature shows contradictory results and sometimes fails to provide enough evidence for some effects, especially on consumption, income and empowerment of women. Nevertheless the results provided evidence that Microfinance can help the poor to become entrepreneurs, grow their business and save more efficiently.

Of course some research questions make a randomized approach impossible, thus the researcher must use other statistical methods aiming to diminish the effect of possible omitted variable bias: such methods could be propensity score matching, genetic matching, fixed effects and multilevel models. Nevertheless, exogenous shocks may provide quasi-experimental evidence without the need for the researcher to randomize the samples.

Further research in this area is focusing on establishing causality and making the causal assessment more generalizable. Current research is focusing more on cross-country RCTs and experiments. Nevertheless, as discussed, RCTs alone are not enough to provide a reliable estimation: transparency on sample selection and estimation are key elements to strengthen the causality claim of Microfinance studies.
Bibliography


