

# The Impact of Microfinance Programs: A Review of Data and Methodological Dilemma

Peter W Muriu

School of Economics, University of Nairobi, Kenya

## Abstract

How robust is the evidence that microcredit works? Theory on microfinance initiatives has moved far ahead of evidence, and success stories remain unsubstantiated. Lack of convergence in findings together with the contexts in which impact studies have been conducted, render the interpretation and generalization of findings intricate. Consequently, thirty years into the growth of microfinance industry there is scant evidence that microfinance improves the lives of clients in measurable ways. This paper reviews methodological and data constraints in microfinance impact assessments. The study builds on earlier work and considers recent research on microfinance initiatives in order to isolate some emerging contestable issues.

*JEL classification:* G21;C52;C81

**Keywords:** Microfinance; Impact; Methodologies; Evaluation; Data

**DOI:** 10.7176/JESD/11-22-09

**Publication date:** November 30<sup>th</sup> 2020

## 1. Introduction

Microfinance has been regarded as an effective policy tool in the fight against poverty. Yet there is so far no consensus among researchers on the impact of microcredit. To date the emphasis has been on potential positive linkages. But why does reality look so different? Quite often we observe outcomes for households with microfinance, but not outcomes for the same households without microfinance at the same moment in time. In spite of the abundance of theoretical literature, there has been surprisingly little concrete and rigorous evidence on whether and how microfinance actually helps to reduce poverty. Theory has moved far ahead of evidence, and success stories remain unsubstantiated. The relatively few carefully conducted cross-sectional or longitudinal impact studies remain anecdotes (Armendáriz and Morduch 2010).

Evaluation of the impact of microfinance requires disentangling the simultaneous roles of income, location and entrepreneurial attributes. Most of the studies fail to control for endogeneity which leads to overestimating or underestimating the interventions. This reflects the difficulty of distinguishing the causal effect of microcredit from selection effects. To identify causal effects, researchers have to find ways to approximate the prediction of what would have happened without the microfinance intervention. The impact of microfinance initiatives therefore remain highly contestable (Muriu, Murinde and Mullineux, 2017). Previous studies have either overlooked or are unable to resolve the overarching issue of data and methodological challenges. Given such hurdles to overcome, the evidence generated by microfinance studies is contestable. This paper seeks to address these weaknesses by reviewing data and various methodologies employed in the evaluation of microfinance impact while at the same time underscoring the need for more rigorous evaluations.

## 2. Cross sectional studies

Traditional research in microfinance has relied on cross-sectional data to test theoretically driven hypotheses where the dependent, explanatory, and control variables each have a single measure, often at one point in time. Cross-sectional data requires the researcher to make strong assumptions about the comparability of different observations. OLS regression represents the most popular estimation technique. The main limitation of cross-sectional studies such as Al-shami, et al (2018), Samer, et al (2015) Ganle, et al (2015), Mazumder & Lu (2015), Ghaliba, et al (2014) is limited effectiveness in tackling the problem of selection & placement bias, attribution, and fungibility.

Selection and placement bias may occur due to; (i) individual participation is usually voluntary. Participants differ from non-participants in observable and unobservable ways. The treatment group systematically possessing an invisible attribute such as entrepreneurial ability which the control group lacks. Therefore receiving any form of intervention may result in a short-term positive response from the treatment group, (ii) difficulties in finding a location at which the control groups economic, physical and social environment matches that of the treatment group, (iii) the control group becoming contaminated by contact with the treatment group and (iv) the fungibility of the treatment (see Hume 2005).

Selection bias may also go in the opposite direction. Many MFIs target women and poor households. Poorer households are likely to be borrowers than their neighbours conditional on village of residence and other observable characteristics. In cross sectional studies, this outreach can lead to a downward bias on the estimated effect of credit on earnings (Morduch 1999). At the extreme, the effective targeting of poor households can yield

the impression that participation in the program makes clients poorer. Addressing selection bias reveals how participation increases earnings. For instance, Karlan (2007) and Ahlin and Townsend (2007a), show that not controlling for selection bias can lead to over estimation of the effect of participation on MFIs profit by as much as 100%.

Difference-in-differences (DID) approach controls for self-selection and non-random program placement bias due to variables that do not change over time. A major identifying assumption in difference in difference estimation is that there are no time-varying, unit-specific shocks to the outcome variable that are correlated with treatment status, and that selection into treatment is independent of temporary individual-specific effect. DID requires data collection both for control and treatment groups before and after intervention. DID estimates can control for fixed differences between clients and non-clients. But this only holds if the impact of the measured, unmeasured, and village attributes do not change over time, otherwise OLS estimates would be biased.

With DID approach, the time invariant unmeasured attributes drop out. But the unmeasured attribute may change over time which may amplify the selection/endogeneity problem. Although selection bias associated with fixed characteristics can be eliminated by DID approach, selection bias associated with growth prospects remain (Armendáriz and Morduch 2010). It might also be the case that people who choose to join MFIs would be on different trajectories. This invalidates comparisons over time between clients and non-clients (see e.g Tedeschi and Karlan 2013). The typical response to the problem is to estimate impacts while including region-level fixed effects or their equivalent (see e.g., Pitt and Khandker, 1998).

Another source of bias is non-random program placement since most social interventions are targeted (Karlan 2007; Armendáriz and Morduch 2010). Upward biases arise when programs choose regions that are already doing well, and downward biases arise when programs favour disadvantaged areas. Many programs are set up specifically to serve the underserved. Thus, they are located in areas previously with weak financial services. This may lead to apparent negative impacts relative to control areas. Alternatively the programs may set up where there is good complementary infrastructure biasing estimates upwards. The pioneering theoretical work by Copestake (2007), shows that wealthier clients cost less. The signs and size of the biases are likely to change as programs expand over time into new areas. Including region-level fixed effects can exacerbate bias when program placement is predicated on unobserved qualities particular to target populations. Data suggest that this is often the case. But, with the exception of the results on reduced variability, the main qualitative results are robust both with and without controls for village-level unobservables (Armendáriz and Morduch, 2010).

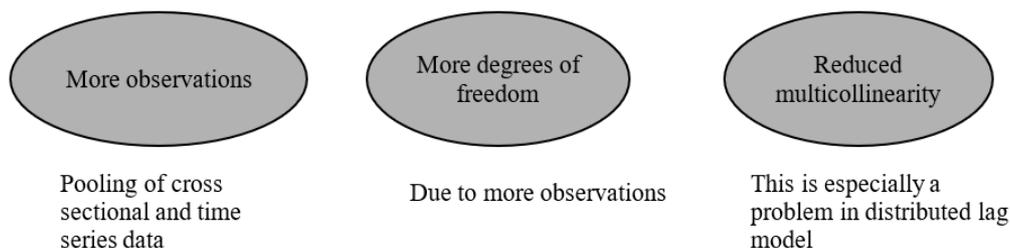
One of the least understood phenomenon when examining the impacts of microfinance using cross-sectional data is the high dropout rates also known as survivorship bias. Borrowers leave either because they are doing so well or they are in trouble (Tedeschi and Karlan (2013). It is likely that those who remain behind have the positive qualities of survivors, while the new borrowers are yet to be tested. If better off clients are more likely to leave, the pool of borrowers becomes poorer on average and when the poor leaves, impacts are overstated. If the failures are more likely to drop out, comparing old to new borrowers will overestimate impacts. Since some borrowers drop out of the program, the treatment group is incomplete. Another problem is due to non random exclusion (Pearlman, 2014).

The issue of fungibility of credit is a critical problem in precisely determining the impact of credit. Credit is often not used as originally indicated to the lender. Failure to control for fungibility may over-estimate the impacts of micro credit intervention. This is particularly true when households have access to multiple sources of credit (Khalily, 2004). The other limitation with pure cross-section data is historical context.

### 3. Longitudinal studies

Some microfinance studies have also used panel data to test static propositions. The novelty of panel data models is their power to overcome the problems of endogeneity bias, in addition to controlling for unobserved heterogeneity across households. Baltagi (2020) highlights several advantages of panel data compared with cross-sectional data.

#### Advantages of panel analysis



Although panel data offers a number of advantages, the analysis of the same requires the adoption of a

number of assumptions and trade-offs. The use of panel data often creates potential statistical problems for OLS regression. Specifically, panel data may create analytic problems in the form of error terms containing heteroskedasticity, autocorrelation, or contemporaneous correlation; the presence of such conditions creates non spherical error terms. Analyzing panel data without panel-level error terms assumes that MFIs are homogeneous. Beck and Katz (1995), highlight the benefits of panel-corrected standard errors to control for such analytic problems. Contemporaneous correlations arise when the errors of unit  $i$  at time  $t$  are correlated with errors of unit  $j$  at time  $t$ . This correlation violates the assumption of spherical error terms and could influence the results of OLS.

Contemporaneous correlation could be particularly harmful in microfinance research. Events occurring in the external environment may increase uncertainty and outcomes for a number of MFIs in a sample. To account for the influence of contemporaneous correlation, researchers have employed techniques such as Generalized Least Square (GLS) estimation. For GLS to model contemporaneous correlation, the number of time periods,  $T$ , in a panel data must exceed the number of units,  $N$ ; this requirement is a necessary condition. To the extent that these problems are present and not corrected, the analysis of panel data may actually produce incorrect estimation results (Baltagi, 2020).

Certo and Semadeni (2006) conducted two studies to examine the potential influence of contemporaneous correlation. They used Monte Carlo simulations to analyze the extent to which contemporaneous correlation influences the results of other estimators such as fixed and random effects regression. By extending Beck and Katz's (1995), they examined the behaviour and performance of different estimators in data sets where the number of units exceeds the number of time periods (i.e.,  $N > T$ ). They were quick to point that when contemporaneous correlation interferes with estimator accuracy, estimators may yield results that support (or reject) a theory when the results are, in fact, driven by contemporaneous correlation.

From a theoretical standpoint, researchers must carefully choose between fixed-effects and random-effects regression when analyzing panel data and use relevant tests to examine whether modelling heterogeneity improves model fit. For example, researchers examining panel data could estimate a model with OLS (Model 1) and then estimate another model with fixed effects (Model 2). A Chow test would enable the researcher to determine whether the inclusion of fixed effects in Model 2 improves the fit compared to Model 1. Choosing between fixed and random effects models remains difficult as each technique has its own advantages and disadvantages. It would be unrealistic to expect that the panel-level error term exerts a constant influence over time with fixed-effects models. The issue of covariates that do not vary over time may also inform the choice between fixed and random effects models.

It is important to note, that just like in DID models, fixed effects models implicitly allow researchers to control for all variables that do not vary over time. The key difference between the two techniques is that random effects models are able to report coefficients for these variables, whereas fixed-effects models do not (see Baltagi 2020). The choice between fixed and random effect models also depends on statistical assumptions underlying the models. Random effects models assume that the estimated panel error term is uncorrelated with the independent variables. The Hausman test helps in evaluating this condition. The Hausman test will reject models if this assumption does not hold. When the Hausman test is rejected, researchers should rely on fixed-effects models, which are not restricted by this assumption (Wooldridge 2002). It is also important for researchers to understand that GLS estimation does not account for unit heterogeneity.

In microfinance studies, fixed effects estimation does not control for lending mechanism features such as peer networks or other relevant characteristics that are specific to households in program villages. The village level fixed effects will only control for those unobservables that affect all households in a village-identically and linearly (Hsiao, 2003). While it is common to use fixed effects estimators to control for unobservable variables correlated with the placement of programs, using fixed effects estimators can exacerbate biases when dealing with targeted populations. It is for this reason that Morduch (1999) contested Pitt and Khandker (1998), findings since poorer households are more likely to be Grameen borrowers than their neighbours, conditional on village of residence and other observable characteristics which can lead to a downward bias on estimated effect of microcredit.

A variant of panel data estimation includes pooled regression model which assumes that any two country-years can be compared, whether across time or across space. Although Cull, Demirgüç-Kunt and Morduch (2007), pioneered the use of cross-country, cross-MFI data in statistical tests and provided a new dimension to the existing literature on MFIs performance, pooled regression model omits fixed effects, and omitting fixed effects risks omitted variables bias. Pooled regression estimates just a single intercept rather than different intercepts for each unit and/or time point. The omitted country-specific intercepts may be correlated with the explanatory variables. Additionally the disturbances may be correlated within groups.

Lensink et al (2018) employ panel data modelling to examine the potential effects of microfinance 'plus' on the financial and social performance of MFIs. Khandker et al (2016) used long panel household survey data to investigate various dimensions of microcredit effects on a set of behaviours. Muriu (2012), Ahlin et al (2011)

also estimates a baseline specification on a sample of MFIs. To the extent that there is likely to be persistence in the MFI outcome variables (see e.g Muriu, 2016), endogeneity remains an issue with these and other similar empirical studies in microfinance.

Fixed/Random Effects estimator (within) model in the presence of a lagged dependent variable among the regressors is both biased and inconsistent. Dynamic panel-data models include  $p$  lags of the dependent variable as covariates and contain unobserved panel-level effects, fixed or random. Therefore and by construction, the unobserved fixed or random effects are correlated with the lagged dependent variables, making standard estimators inconsistent (Baltagi, 2020). Ideally, these panel methods do not control, where possible, for endogeneity, measurement error and omitted variables. One prominent way to address these shortcomings has been through first-differenced generalized method of moments estimators (GMM) applied to dynamic panel data models. The relevant estimator was originally developed by Holtz-Eakin, et al (1988) and refined by Arellano and Bond (1991).

The basic idea of GMM estimator is to specify the regression equation as a dynamic panel data model then take first-differences to remove unobserved time-invariant country-specific effects. Further, instrument the right-hand-side variables in the first-differenced equations using levels of the series lagged two periods or more, under the assumption that the time-varying disturbances in the original levels equations are not serially correlated (Baltagi 2020). An advantage of the dynamic GMM estimation is that all variables from the regression that are not correlated with the error term including lagged and differenced variables can be used as valid instruments (Greene, 2008). This is an unexplored dimension in microfinance literature.

Microfinance endogenous placement problem can be dealt with by using group-level fixed effects. If for instance, placement decisions are made on the basis of observable and unobservable characteristics of the region or the group, the direct effect based on these characteristics will pass into the fixed effects. Because of the inconsistency of fixed-effect probit regressions, linear probability regression-probit, and logit specifications with robust standard errors can be utilized since robust standard errors in the linear probability model are substantially larger. Panel data modelling however, allows for construction and tests of more complicated behavioural models compared to other estimation techniques. The methodology also allows the dynamics of adjustment, identification and measurement of effects which would have otherwise not been detectable in the analysis of cross section or time series data (Baltagi, 2020).

#### **4. Search for convincing instrumental variables**

An alternative to the difference-in-difference model is the use of instrumental variables (IV), which is close to structural econometric method since it relies on exclusion restrictions (Angrist and Krueger, 2001). The IV method assumes that some components of the non-experimental data are random. It is perhaps the most widely used approach in measuring treatment effects (Islam, 2015). IV seeks to find a variable that is excluded from the outcome equation, but which is related to treatment status and has no direct association with the outcome. Several empirical studies such as Martinez, (2015), Alam, (2013), Kaboski and Townsend (2005) have utilized instrumental variables.

The search for convincing instrumental variables for credit has yielded little. The weakness of the IV approach is that such variables do not often exist, or that unrealistic assumptions must be maintained in order for them to be used to identify the treatment effect of interest. The difficulty in identifying instrumental variables often arise because many of the characteristics that define MFIs are likely to be endogenous in performance and outreach regressions. The problem is further compounded by the fact that variables which may be unrelated may be integrally linked due to non-separabilities driven by imperfect and incomplete markets. It then becomes less likely that a production side variable that explains credit use does not also help explain expenditure related outcomes independently (Morduch 1999, Armendáriz and Morduch, 2010).

Interest rate is a potential identifying variable but since achieving uniformity across branches is a common goal of microfinance programs, interest rates are very unlikely to vary within a given area and estimation is impossible without some variation. Even if interest rates vary, it is likely that the variation will at least partly reflect unobserved attributes of the borrower undermining their use as instruments. Pitt and Khandker's (1998), exploited the landholding criterion by using a weighted exogenous sampling maximum likelihood and regression discontinuity design to estimate marginal impacts rather than the average impacts. Their general approach is recast as an instrumental variables problem that helps clarify necessary identifying restrictions and illustrates broader methodological differences.

Using the data in Khandker (2003) and similar methodology, Khandker (2005) conducted a specification test and concluded that fixed-effects technique was more appropriate than fixed-effects with instrumental variables approach. Instrumental variables therefore lead to larger standard errors and less efficient results. Roodman and Morduch (2014) replicates Pitt and Khandker (1998), Morduch (1999) and Khandker (2005) applied the same methodology to the same data and performed closely related Two-Stage Least-Squares regressions which generated results opposite in sign. Their specification tests suggest that the instrumentation

strategy fails; that reverse or omitted-variable causation was driving the results, and that the sign and magnitude of the endogenous credit-consumption relationship vary by sub-sample, as well as the gender of the borrower. This explains the seeming gender differential in impact. Further analysis on these data questions the basis for the quasi-experimental identification in Pitt and Khandker (1998), and by extension in Morduch (1999) and show how, in Khandker (1998), exploiting the panel dimension does not compensate for the lack of clearly exogenous variation in the treatment variable.

## 5. Natural experiment

The main goal of microfinance impact assessment is to estimate the effect of an intervention in relationship to the counterfactual—what would have happened in the absence of the intervention. A common technique to achieve this is to compare program participants to a group of similar non-participants. The difficulty is to find groups that are sufficiently comparable. With most impact assessments in microfinance, the major hurdle is that the treatment group was self-selected.

Natural experiments occur when a researcher observes naturally occurring, controlled comparisons of one or more treatments with a baseline (Levitt and List 2009). Field experiment has only one source of possible selection bias: the decision to be in the naturally occurring market. The main disadvantage of natural experiments is that the researcher does not get to pick and choose the specifics of the treatments, and does not get to pick where and when the treatments will be imposed (Glenn et al, 2004). The difference between the parameter estimates of treatment and control groups is the effect of intervention. However, precision depends on the ability to control for observable and unobservable characteristics.

Quasi-experiments on the other hand seek to compare the outcomes of an intervention with a simulation of what the outcomes would have been, had there been no intervention. Estimation techniques include multiple regressions, but this has enormous demands for data on other possible causal factors and its assumptions. Innovations using structural evaluation of large-scale quasi-experimental initiatives include Bhuiya, et al (2016), You & Annim (2014), Nghiem, et al (2012), Kaboski and Townsend (2011). The quasi-experimental design, as argued by Habib and Jubb (2015) could be less effective in the absence of finding non-programme villages because of the depth of microfinance programs. The situation was partly experienced by Zohir et al (2001) while using disaggregated target participants by length of membership as a proxy for non-participant control group.

## 6. Randomised controlled trials

In microfinance, randomised controlled trials (RCTs) isolate the effect of a chosen innovation by assigning a random selection of individuals or villages to the innovation and another equivalent selection of villages or individuals to maintain the status quo and comparing results between the groups. RCTs enable researchers to estimate the impact of the intervention, free of the identified bias.

Methodologies for addressing selection bias can be ordered by their effectiveness at dealing with the problem and RCTs have proved better than others. RCTs require fewer assumptions in order to establish causality (See Imbens 2010). In randomized experiments, individuals are “assigned” to the treatment and control groups, they do not form the groups themselves. The characteristics of the participants are not related to the outcome, and the difference in outcomes between borrowers and non-borrowers is therefore only due to the loan.

RCTs have become increasingly popular in development economics as a way of overcoming many of the identification concerns that can arise with non-experimental approaches. Many of the earliest applications were in education and health (see, for example, Kremer 2003). Later on microfinance studies gained traction. These include Karlan and Zinman (2009, 2010, 2011), Ashraf, Karlan, and Yin (2010), Kaboski and Townsend (2011), Dupas and Robinson (2013). These studies were able to circumvent the endogeneity problem by making use of RCTs, and estimating a difference-in-difference (using pre-post and treatment-control data) model which enabled them to rule out the possibility that social connections correlates with other group characteristics in influencing repayment. This process of group formation exogenously creates groups with different levels of initial social ties, which enables the actual measurement of the impact of these social ties on monitoring and enforcement efforts within the group.

Subsequent rigorous analysis to evaluate the impact of access to microcredit on borrowers and their households, that controlled for endogeneity problem were set up in Bosnia and Herzegovina (Augsburg, et al 2015), Ethiopia (Tarozzi, et al 2015), India (Banerjee, et al 2015), Mexico (Angelucci, et al 2015), Mongolia (Attanasio, et al 2015), Morocco (Crépon, et al 2015) and cross country (Banerjee, Karlan and Zinman 2015). Studies that have augmented RCTs with qualitative analysis include Guerin, et al (2014).

RCTs have been embraced as the panacea for evaluations. But randomization is not always possible, nor desirable and designing and implementing randomized experiments is difficult (Banerjee and Duflo 2009; Deaton 2009; Imbens 2010). First, the methodology provides an estimate of the average impact of an intervention. It does not tell us anything about the median impact, and offers little about the distribution of impacts. Second, while RCTs excel at providing a clean estimate of impact, they are by necessity implemented

in a particular setting, and may not be generalized to other settings (external validity). For instance, Prina (2015) shows that the design of the field experiment with randomization at the household level, rather than at the village level, could not allow her to study the general equilibrium effects of giving access to bank accounts to the entire sample of households. Similarly, a randomized evaluation of flip charts as teacher's aide in schools in Kenya (Glewwe, et al 2000) only tells us whether the flip charts helped raise test score for the students in Western Kenya. It is plausible that students or schools in other parts of Kenya, India, or Latin America have different educational needs, and would benefit differently from flip charts. Additionally, can one conclude from a one-year follow-up that microfinance does not work?

Another concern with the current trend towards credible causal inference in general, and towards RCTs in particular, is that it may lead researchers to avoid questions where randomization is difficult, or even conceptually impossible. There are many such questions, and many of them are of great importance. Questions such as will digital banking reach the very poor or concerning the causal effects of macro-economic policies on microfinance cannot be settled by randomized experiments. In some cases randomized experiments raise ethical concerns, and are ultimately not feasible. However, the importance of questions for which randomization is difficult or infeasible, should not pre-empt RCTs. Well-implemented randomized experiments provide particularly powerful ways to measure impact or improved product design because they guarantee the best unbiased estimates.

Future research on randomized control trials should emphasize on the need for replications in other settings before drawing conclusions. Because randomized experiments are carefully planned and implemented, expansion to a large scale may yield different results. The best way to overcome external validity is through theory and repetition. By testing ideas in sufficient locations, with a variety of contexts, researchers can confidently state that a theory holds and can be used to inform policies. It is time microfinance researchers started collecting rigorous evidence about what works and why, so that policymakers and practitioners can maximize the impact of microfinance programs around the world. Other econometric approaches will continue to be valuable in assessing directions of impact and the ensuing relationships between economic and financial variables.

Randomized and non-randomized techniques have different strengths and weaknesses. For example, non-randomized techniques can opportunistically exploit natural experiments and therefore from a policy perspective the optimal research portfolio should blend the two techniques. It is however not clear when non-experimental studies are worth undertaking. Therefore when analyzing causality with strong endogeneity, non-experimental identification should be rigorous. The evidence adduced so far also casts doubt on the power of sophisticated parametric techniques to compensate for the lack of such (Roodman and Morduch 2014).

## **7. Data challenges in microfinance research**

The main difficulty in microfinance impact studies arising from methodological techniques is partly due to the challenges of obtaining reliable data on the working of microfinance programs and for testing concrete hypothesis. Data is critical for articulating policy initiatives, but as a relatively youthful industry, microfinance data have been slow to accumulate. Researchers use different sources of data to test various hypotheses. Different methodological approaches suggest that rigorous impact studies can be increased by basing them on reliable data (Chemin 2008).

The Microfinance Information Exchange, Inc. (MIX), through its MixMarket platform, and the Microcredit Summit Campaign are the main source of secondary data and have been critical in collecting large data sets on MFIs but both have their own targets and priorities. The MixMarket collects a wide array of financial and institutional data, supplemented by a limited amount of social data. The MIX also prepares the Microbanking Bulletin which involves data collection on a subset of institutions. These statistics are then adjusted to improve comparability and reveal implicit subsidies. Microcredit Summit Campaign on the other hand collects a limited number of indicators, most of them related to social outreach, though the number of MFIs reporting is higher relative to the MIX. The MIX is skewed towards financial data, while Microcredit Summit Campaign lays more emphasis on social and economic change. The strength of the Microcredit Summit Campaign dataset is the number of MFIs reporting to it and the diversity of these MFIs. The Summit collects data on many small institutions and a few very large ones, whereas the MIX database includes a set of medium to large MFIs.

Although MixMarket and Microcredit Summit Campaign are geared towards global coverage and encourage broad submission of data, MFIs appearing in the two data sets differ systematically by geographic location and by their focus on poverty and financial performance. Different statistics are therefore obtained when asking the same questions on the same variables. The extent of this data inconsistency was documented by Bauchet and Morduch, (2010) who found positive relationship between financial performance and reaching more women customers in MIX data set, but a negative relationship using Microcredit Summit Campaign data set.

MixMarket data has limitations. For instance, it could not allow Cull, Demirgüç-Kunt and Morduch (2007) to answer the controversial question on whether subsidies in microfinance are justified. Answering the question

would have required reliable data on social impacts but which was not forthcoming. Data is heterogeneous across MFIs and countries. Survey is undertaken using different methodologies, in different geographical zones, prevailing economic circumstances, different legal and regulatory framework e.g on interest rates cap and incomplete entries. There is therefore a justification on the need to standardize the available data prior to making meaningful analysis.

Since reporting to any microfinance database is voluntary, analyses based on these data are vulnerable to self-selection bias. Using a quasi-experimental technique on a panel data, Tedeschi, (2008) shows that self-selection into the lending program is a significant problem in the data. MFIs reporting to any source are likely to be different from those not reporting at all. This bias is likely to be large in magnitude, though it is difficult to measure and overcome. Additionally, MFIs self-select into reporting to either the MixMarket or Microcredit Summit Campaign or both. Consequently, MFIs reporting to one outlet might differ from those reporting to another, or from MFIs reporting to more than one outlet. MFIs may also choose to report some indicators and conceal others, and for some years discretionary.

While participation in firm and household level surveys is in most cases voluntary, some choose to skip questions entirely. This pattern of voluntary participation and of discretionary response to and skipping questions is not random. Additionally, high-income households are very difficult for researchers to engage. Developing countries tend to have less-developed statistical systems and irregular surveys and therefore samples, whether at the individual or macro level are seldom full representation of the underlying populations (Bauchet and Morduch, 2010).

Several studies such as Ahlin and Townsend (2007b), Gugerty (2007), have utilized survey data. This includes collecting information on borrowers before and after program participation. No statistical analyses of the differences are conducted and their data did not allow them to control for the possibility of endogenous program placement. Changes in the household survey response rates over time as incomes rise and frequent changes in the design of household survey within the same country make inference based on the within a specific country less reliable compared to inference that relies on within-MFIs. Hence, while the survey data could be used for measuring the impacts of microfinance on poverty among other variables, it has not been adequate for testing concrete theoretical propositions.

Recall data is often used in the absence of a baseline survey since it can control for both non-random participation and non-random program placement. However, it is also subject to potential biases due to time varying unobservables. Differencing noisy data such as that used by Mosley (2001) can exacerbate measurement error. Noisy recall data may bias downward coefficients which captures program impacts (Armendáriz and Morduch 2010).

MFIs that are rated have a common interest in accessing funding and increasing their sustainability. The leading rating agencies in the microfinance industry include Microrate, Microfinanza, Planet, Crisil, M-CRIL, Moody's and Standard & Poor. The rating reports are narratives consisting of contextual and MFI-specific information including accounting details, organizational features and benchmarks. Annual rating reports have featured in microfinance research (see e.g Mathews et al 2007) but due to poor book-keeping by MFIs, the data is often incomplete, unreliable and not repeated across samples. The reports are not fully standardized and therefore differ in their emphasis and in the amount of information available. For example, in Grameen Bank all accounting definitions are not standard e.g the reported overdue rates are calculated as the value of the loans overdue, greater than one year, divided by the current portfolio. The problem is that the current portfolio tends to be much higher than the portfolio that existed when the overdue loans were first made. Reported profits do not conform to international accounting standards. Grants from donors are considered part of the income in the profit calculation.

Experimental data have recently become popular. Every empirical researcher who assumes that an exogenous variable varies independently of an error term views their data as being generated from an experiment. In some cases this belief is a matter of *a priori* judgment; in other cases it is based on auxiliary evidence and inference; while in others it is built into the design of the data collection process (Glenn et al, 2004). That notwithstanding, the question still abounds-does this solve data problems? Does the experimental data match simulation data well or rather do human agents behave identically? Do the behavioural differences appear to be large enough to affect the outcome of the experiment? Qualitative Impact Protocol (QUIP) is one such effort that develops guidelines on data collection and analysis but does not resolve quantitative data issues.

## 8. Conclusion

This paper reviews methodological and data hurdles in microfinance impact evaluations and the reasons why more rigorous studies are necessary. As MFIs innovate in products and programs at a rapid pace, it's important for policymakers and practitioners to understand the relative impact of different mechanisms, particularly on the borrowers. The study has observed that cross sectional methodology does not measure the impact of a microcredit program satisfactorily. This reflects the difficulty of distinguishing the causal effect of microcredit

from selection effects.

RCTs literature has made great strides towards improving the credibility of empirical work. But they have limits: they are not always feasible, not always representative, and not always focused on the larger questions of interest. Moreover, there have been limited large-scale randomized trials with the potential to examine what happens when microfinance becomes available in a new market. It would be regrettable if randomized trials led researchers to avoid questions that cannot be answered through randomized or natural experiments. Quasi-experimental methods are not always available, panel data is time consuming and expensive for most MFIs, while experimental methods sometimes are not feasible and requires expertise which is not available to the MFIs. This leaves the problem of loan fungibility unresolved. The issue of fungibility is more important than the issues of selection bias and endogeneity in the accurate assessment of impact of microfinance.

As long as the issue of endogeneity is recognized, it can be resolved through application of appropriate econometric techniques. Disaggregated analysis by access to different sources of credit like households with micro-credit and formal bank credit only, households with micro-credit and informal credit only and households with access to all sources of credit may enable a researcher to derive policy implication of microfinance quite clearly. Both descriptive and case studies as well as econometric approaches can be applied in such a situation. Such studies can inform the debate on microfinance by sharpening the practitioners' understanding of the role of microfinance in reaching the poor, its impact in different environments, and its cost-effectiveness as a poverty intervention.

Considering credit as endogenous, predicted value of credit along with instrumental variables can be used in the function of welfare outcomes. Moreover, household economic portfolio approach may also be applied to control for fungibility. Estimates of parameters are sensitive to changes in the sample as well as to changes in the model specification. Much evidence has accumulated about the sensitivity of results to changes in model specification or data. Thus, for some research problems, the question remains, can useful estimates of structural parameters be obtained from observational data?

## References

- Ahlin, C. and Townsend, R (2007a). Selection into and across credit contracts: Theory and field research, *Journal of Econometrics*, **136**:665-698.
- Ahlin, C. and Townsend, R (2007b). Using repayment data to test across models of joint liability lending, *The Economic Journal*, **117**: F11-F51.
- Ahlin, C., Lin, J and Maio, M (2011). Where does microfinance flourish? Microfinance institution performance in macroeconomic context, *Journal of Development Economics*, **95**:105-120
- Alam, S (2013). The impact of credit and non-credit aspects on self-employment profit: a comparison of microcredit programs and commercial lenders in rural Bangladesh, *The Journal of Developing Areas* **47**: 23-45
- Al-shami, S., Razali, R. M., Rashid, N (2018). The effect of microcredit on women empowerment in welfare and decisions making in Malaysia, *Social Indicators Research* **137**:1073-1090.
- Angelucci, M., Karlan, D and Zinman, J (2015). Microcredit Impacts: Evidence from a randomized microcredit program placement experiment by Compartamos Banco, *American Economic Journal: Applied Economics* **7**:151-182.
- Angrist, J and Krueger, A (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments, *Journal Economic Perspectives*, **15**:69-85.
- Arellano, M and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *Review of Economic Studies*, **58**:277-297.
- Armendáriz de Aghion, B. and Morduch, J (2010). *The Economics of Microfinance, second edition* MIT Press, Cambridge, Massachusetts.
- Ashraf, N., Karlan, D and Yin, W (2010). Female empowerment: Impact of a commitment savings product in the Philippines, *World Development*, **38**:333-344
- Attanasio, O., Augsburg, B., De Haas, R., Fitzsimons, E and Harmgart, H (2015). The impacts of microfinance: Evidence from joint-liability in Mongolia, *American Economic Journal: Applied Economics* **7**:90-122.
- Augsburg, B., De Haas, R., Harmgart, H and Meghir, C (2015). The impacts of microcredit: Evidence from Bosnia and Herzegovina, *American Economic Journal: Applied Economics* **7**:183-203.
- Baltagi, H (2020). *Econometric Analysis of Panel Data*, 6th Ed. Springer International Publishing
- Banerjee, A and Duflo, E (2009). The experimental approach to Development Economics, *Annual Review of Economics*, **1**:151-178.
- Banerjee, A., Duflo, E., Glennerster, R and Kinnan, C (2015). The miracle of microfinance? Evidence from a randomized evaluation, *American Economic Journal: Applied Economics* **7**: 22-53.
- Banerjee, A., Karlan, D and Zinman, J (2015). Six randomized evaluations of microcredit: Introduction and further steps, *American Economic Journal: Applied Economics* **7**:1-21.

- Bauchet, J and Morduch, J (2010). Selective knowledge: Reporting biases in microfinance data, *Perspectives on Global Development and Technology*, **9**:240-269
- Beck, N and Katz, J. (1995). What to do (and not to do) with time series cross sectional data, *America Political Science Review*, **89**:634-647.
- Berge, L. I. O., Bjorvatn, K and Tungodden, B (2014). Human and financial capital for microenterprise development: Evidence from a field and lab experiment, *Management Science* **61**:707-22.
- Bhuiya, M.M., Khanam, R., Rahman, M.M & Nghiem, H.S (2016). Impact of microfinance on household income and consumption in Bangladesh: Empirical evidence from a quasi-experimental survey, *The Journal of Developing Areas*, **50**:305-318.
- Certo, S and Semadeni, M. (2006). Strategy research and panel data: Evidence and implications, *Journal of Management*, **32**:449-471.
- Chemin, M (2008). The benefits and costs of microfinance: Evidence from Bangladesh, *Journal of Development Studies*, **44**:463-484
- Copstake, J. (2007). Mainstreaming microfinance: Social performance management or mission drift? *World Development*, **35**:1721-1738.
- Crépon, B., Devoto, F., Duflo, E and Parienté, W (2015). Estimating the impact of microcredit on those who take it up: Evidence from a randomized experiment in Morocco, *American Economic Journal: Applied Economics* **7**:123-50.
- Cull, R., Demirgüç-Kunt, A and Morduch, J. (2007). Financial performance and outreach: a global analysis of leading microbanks, *Economic Journal*, **117**:F107-F133.
- Deaton, A. (2009). Instruments of development: randomization in the tropics, and the search for the elusive keys to economic development, *Proceedings of the British Academy, 2008 Lectures* 162: 123-160 .
- Dupas, P and Robinson, J (2013). Savings constraints and microenterprise development: Evidence from a field experiment in Kenya, *American Economic Journal: Applied Economics* **5**:163-192.
- Ganle, J. K., Afriyie, K., Segbefia, A. (2015). Microcredit: Empowerment and disempowerment of rural women in Ghana, *World Development* **66**:335-345
- Ghaliba, A. K., Malki, I & Imai, K.S (2014). Microfinance and household poverty reduction: Empirical evidence from rural Pakistan, *Oxford Development Studies*, **43**:84-104.
- Glenn, H. W and List, J. (2004). Field experiments, *Journal of Economic Literature*, **42**:1009-1055.
- Glewwe, P., Kremer, M., Moulin, S and Zitzewitz, E (2000). Retrospective vs. prospective analyses of school inputs: The case of flip charts in Kenya, *Journal of Development Economics* **74**:251-268
- Greene, W. H. (2008). *Econometric Analysis*. 6th ed. Upper Saddle River: Prentice Hall
- Guerin, I., Morvant-Roux, S., Roesch, M & Moisseron, J.Y. (2014). Adding value to randomization with qualitative analysis: The case of microcredit in rural Morocco, *World Development*, **56**:302-312.
- Gugerty, M. (2007). You can't save alone: Testing theories of rotating savings and credit associations, *Economic Development and Cultural Change*, **55**:251-282.
- Habib M., Jubb C. (2015). Poverty alleviation with microfinance: bangladesh evidence. In: Heshmati A., Maasoumi E., Wan G. (eds) *Poverty reduction policies and practices in developing Asia*. Economic studies in inequality, social exclusion and well-being, Springer, International Publishing
- Holtz-Eakin, D., Newey, W and Rosen, H. (1988). Estimating vector autoregressions with panel data. *Econometrica*, **56**:1371-1395
- Hsiao, C. (2003). *Analysis of Panel Data*, second edition, Cambridge University Press, Cambridge.
- Hulme, D. (2005). Impact assessment methodologies for microfinance: theory, experience and better practice, in *Finance and Development: Surveys of theory, evidence and policy*, Green, C.J., Kirkpatrick, C.H. and V. Murinde (eds). Edward Elgar: Cheltenham.
- Imbens, G. (2010). Better late than nothing: some comments on Deaton (2009) and Heckman and Urzua (2009), *Journal of Economic Literature*, **48**:399-423
- Islam, A. (2015). Heterogeneous effects of microcredit: Evidence from large-scale programs in Bangladesh, *Journal of Asian Economics* **37**:48-58
- Kaboski, J and Townsend, R (2011). A structural evaluation of a large - scale quasi - experimental microfinance initiative, *Econometrica* **79**:1357-1406
- Kaboski, J., and Townsend, R. (2005). Policies and impact: An analysis of village-level microfinance institutions, *Journal of the European Economic Association*, **3**:1-50.
- Karlan, D and Zinman, J (2010). Expanding credit access: using randomized supply decisions to estimate the impacts, *Review of Financial Studies*, **23**:433-464
- Karlan, D and Zinman, J (2011). Microcredit in theory and practice: Using randomized credit scoring for impact evaluation, *Science* **332**:1278-1284.
- Karlan, D. (2007). Social connections and group banking. *Economic Journal*, **117**:F52-F84.
- Karlan, D., Zinman, J. (2009). Observing unobservables: Identifying information asymmetries with a consumer

- credit field experiment, *Econometrica*, **77**:1993-2008
- Khalily, M. (2004). Quantitative approach to impact analysis of microfinance programs in Bangladesh-what have we learned? *Journal of International Development*, **16**:331-353.
- Khandker, S (2003). Micro-Finance and poverty. Evidence using panel data from Bangladesh, World Bank Policy Research Working Paper, 2945, Washington, D.C.
- Khandker, S (2005). Microfinance and poverty: Evidence using panel data from Bangladesh, *The World Bank Economic Review* **19**:263-286
- Khandker, S. R., Khalily, B and Samad, H (2016). Dynamics of Microfinance Benefits, a book chapter in: Beyond ending poverty: The dynamics of microfinance in Bangladesh, World Bank, Washington, D.C.
- Khandker, S.R. (2005). Microfinance and poverty: evidence using panel data from Bangladesh, *The World Bank Economic Review*, **19**:263-86.
- Kremer, M (2003). Randomized evaluations of educational programs in developing countries: Some lessons, *American Economic Review Papers and Proceedings*, **93**:102-06.
- Lensink, R., Mersland, R., Vu, N.T & Zamore, S (2018). Do microfinance institutions benefit from integrating financial and nonfinancial services? *Applied Economics*, **50**:2386-2401
- Levitt, S and List, J (2009). Field experiments in economics: The past, the present, and the future, *European Economic Review*, **53**:1-18
- Martinez, C (2015). Doing well by doing good? Empirical evidence from microfinance, CFD Working Papers 06-2015, Centre for Finance and Development, The Graduate Institute.
- Mazumder, M. S & Lu, W. (2015). What impact does microfinance have on rural livelihood? A comparison of governmental and non-governmental microfinance programs in Bangladesh, *World Development*, **68**:336-354.
- Menkhoff and Rungruxsirivorn (2011). Do village funds improve access to finance? Evidence from Thailand, *World Development Volume* **39**:110-122
- Morduch, J. (1999). The microfinance promise, *Journal of Economic Literature* **37**:1569–1614.
- Mosley, P (2001). Microfinance and poverty in Bolivia, *Journal of Development Studies*, **37**:101-132
- Muriu, P.W (2012). What explains the low profitability of Sub-Sahara Africa microfinance institutions? *African Journal of Social Sciences*, **2**:85-115
- Muriu, P.W (2016). Do microfinance profits converge? Evidence from a dynamic panel analysis, *Journal of Economics and Sustainable Development*, **7**:2222-2855
- Muriu, P.W., Murinde, V and Mullineux, A.W (2017). Reflections on microfinance book chapter In: Biekpe N., Cassimon D., Mullineux A. (eds) *Development Finance* pp 109-159. Palgrave Macmillan, Cham
- Nghiem, S., Coelli, T & Rao, P (2012). Assessing the welfare effects of microfinance in Vietnam: Empirical results from a quasi-experimental survey, *The Journal of Development Studies*, **48**:619-632.
- Pearlman, S (2014). Dropouts, defaulters, and continuing bor-rowers: client exit from microfinance, *The Developing Economies* **52**:301-21
- Pitt, M. and Khandkher, S (1998). The impact of group based credit programs on poor households in Bangladesh: Does the gender of the participants matter? *The Journal of Political Economy*, **106**:958-996.
- Roodman, D and Morduch, J (2014). The Impact of microcredit on the poor in Bangladesh: Revisiting the evidence, *The Journal of Development Studies*, **50**:583-604
- Samer, S., Majid, I., Rizal,S., Muhamad, M. R., Sarah-Halim, and Rashid, N (2015). The impact of microfinance on poverty reduction: Empirical evidence from Malaysian perspective, *Procedia-Social and Behavioural Sciences* **195**:721-728
- Tarozzi, A., Desai, J and Johnson, K (2015). The impacts of microcredit: Evidence from Ethiopia, *American Economic Journal: Applied Economics* **7**:54-89.
- Tedeschi, G (2008). Overcoming selection bias in microcredit impact assessments: A case study in Peru, *Journal of Development Studies*, **44**:504-518
- Tedeschi, G. A., & Karlan, D. (2013). Cross-sectional impact analysis: Bias from dropouts. In *The Credibility of Microcredit: Studies of Impact and Performance* (pp. 83-104). Brill.
- Wooldridge, J (2002). *Econometric analysis of cross section and panel data*, 2nd. ed. Massachusetts: The MIT Press.
- You, J & Annim, S (2014). The impact of microcredit on child education: Quasi-experimental evidence from rural China, *The Journal of Development Studies*, **50**:926-948.
- Zohir S, Simeen M, Sen B, Asaduzzaman M, Islam J, Ahmed N, Al-Mamun A (2001). Monitoring and evaluation of microfinance institutions. BIDS, Dhaka
- Prina, S (2015). Banking the poor via savings accounts: Evidence from a field experiment, *Journal of Development Economics*, **115**:16-31