



**INTERNATIONAL FOOD
POLICY RESEARCH INSTITUTE**
sustainable solutions for ending hunger and poverty
Supported by the CGIAR



**ETHIOPIAN DEVELOPMENT
RESEARCH INSTITUTE**

ESSP II Working Paper 43

Assessing the Long-term Impact of Microcredit on Rural Poverty: Does the Timing and Length of Participation Matter?

Guush Berhane and Cornelis Gardebroek

Development Strategy and Governance Division, International Food Policy Research
Institute – Ethiopia Strategy Support Program II, Ethiopia

Ethiopia Strategy Support Program II (ESSP II)

ESSP II Working Paper 43

September 2012

IFPRI-ADDIS ABABA
P.O. Box 5689
Addis Ababa, Ethiopia
Tel: +251-11-646-2921
Fax: +251-11-646-2318
E-mail: ifpri-addis@cgiar.org
<http://essp.ifpri.info/>

IFPRI HEADQUARTERS
International Food Policy Research Institute
2033 K Street, NW • Washington, DC 20006-1002
USA
Tel: +1-202-862-5600
Skype: IFPRIhomeoffice
Fax: +1-202-467-4439
E-mail: ifpri@cgiar.org
<http://www.ifpri.org>

THE ETHIOPIA STRATEGY SUPPORT PROGRAM II (ESSP II)

WORKING PAPERS

ABOUT ESSP II

The Ethiopia Strategy Support Program II is an initiative to strengthen evidence-based policymaking in Ethiopia in the areas of rural and agricultural development. Facilitated by the International Food Policy Research Institute (IFPRI), ESSP II works closely with the government of Ethiopia, the Ethiopian Development Research Institute (EDRI), and other development partners to provide information relevant for the design and implementation of Ethiopia's agricultural and rural development strategies. For more information, see <http://www.ifpri.org/book-757/ourwork/program/ethiopia-strategy-support-program> or <http://essp.ifpri.info/> or <http://www.edri.org.et/>.

ABOUT THESE WORKING PAPERS

The Ethiopia Strategy Support Program II (ESSP II) Working Papers contain preliminary material and research results from IFPRI and/or its partners in Ethiopia. The papers are not subject to a formal peer review. They are circulated in order to stimulate discussion and critical comment. The opinions are those of the authors and do not necessarily reflect those of their home institutions or supporting organizations.

About the Author(s)

Guush Berhane: Post-Doctoral Fellow, Development Strategy and Governance Division, International Food Policy Research Institute – Ethiopia Strategy Support Program II, Ethiopia.

Cornelis Gardebroek: Agricultural Economics and Rural Policy Group, Department of Social Sciences, Wageningen University, The Netherlands.

**Assessing the Long-term Impact of Microcredit on
Rural Poverty:
Does the Timing and Length of Participation
Matter?**

Guush Berhane and Cornelis Gardebroek

Development Strategy and Governance Division, International Food Policy Research
Institute – Ethiopia Strategy Support Program II, Ethiopia

Table of Contents

Abstract.....	vi
1. Introduction.....	1
2. Estimating long-term impacts of periodical participation in microfinance credit	3
2.1. The structure of the dataset used.....	3
2.2. Problems of identifying credit impact in long panels with periodical participation.....	4
3. Empirical method.....	5
3.1. The forward-looking sequential counterfactual method	5
3.2. Propensity score matching method	7
4. Results and discussions	9
4.1. Matched versus unmatched pairwise comparisons	9
4.2. Matched composite versus matched pairwise comparisons	9
4.3. Effects of changes in the composition of treatment and control groups in time- varying treatments	12
5. Conclusions.....	15
References.....	16

List of Tables

Table 2.1—Summary statistics of yearly household consumption expenditure and other variables used in the propensity score matching Method.....	3
Table 4.1—Simple pairwise comparisons between matched and unmatched households: Average participation effects on yearly household consumption expenditure.....	11
Table 4.2—Matched composite versus matched pairwise comparisons: Household consumption expenditure measurement	11
Table 4.3—Effects of repeat-participation and changes in the composition of participants: matched and unmatched pairwise comparisons	14

List of Figures

Figure 2.1—Borrowing participation transition of sample households, 1997–2006	4
---	---

Abstract

In recent years, microfinance institutions are seen as beacons of hope to help eradicate poverty through, among others, providing credit to poor rural households. Availability of small but repeated loans is, in the long-term, expected to impact on poverty. However, decades after the introduction of microfinance institutions into many rural areas, there are still questions as to what extent such long-term benefits are realized. This is because evaluating the long-term impact of microfinance provision on household welfare is difficult due to difficulties in controlling for heterogeneities in the borrower pool and subsequent borrowing dynamics. We use four rounds of panel data on rural households in Ethiopia that had access to microfinance credit to assess the long-term impact of credit by explicitly accounting for such borrowing dynamics. In a panel data setting where only the outcome variable is time-varying, potential future paths of individuals in the control group are considered to account for counterfactuals. Matching is used to adjust for initial differences between participants and controls. These combined methodological innovations enable us to overcome biases due to selection as well as problems of accounting for dropouts and new participants in microfinance impact assessments. Results suggest that the timing of membership matters: the earlier the onset of membership the better the effect. Results are robust compared to standard matched pairwise effects. The overriding implication from this comparison suggests that not accounting for future counterfactuals, for the most part, can lead to biased estimates and overestimates of the impact of credit on household welfare.

Key words: microfinance, impact, dropouts, composite counterfactuals, Ethiopia

1. Introduction

In recent years, microfinance is seen as a beacon of hope to help eradicate poverty and has been at the centre of policy making in many developing countries. A central element in microfinance is providing small but progressively larger loans to those that lack the required collateral to access credit from conventional lenders. These loans are expected to help lift borrowers out of vulnerability and poverty over time. In the case of Ethiopia, repeated loans are primarily intended to bridge shortfalls in key agricultural production inputs such as fertilizer and improved seeds and protect assets that would otherwise be depleted to fill these financing gaps and gradually improve the ability of the beneficiary households to mitigate aggregate shocks. If successful, such loans would eventually trickle down to measurable welfare gains, such as increases in food and non-food consumption (e.g., Menon 2006), or reducing vulnerability to economic hardships (Morduch 1998). Nevertheless, many years since these programs first came into operation, questions still remain if and to what extent the recipients of these successive small loans have been successful in achieving the intended long-term goals of microfinance institutions and policy.

Existing studies focus on evaluating before and after effects, regardless of the timing of participation and dynamics between participation and outcome measurement periods. There are however differences among target households when it comes to benefiting from availability of credit. One major difference is that not all targeted households start to use credit at the same time and in the same intensity. For some reasons, some join earlier than others; some keep their membership for a long time while others drop out quickly. As a result, the effect of microfinance credit varies across different client pools and therefore evaluating the trickled down effects of credit requires measuring the relative impacts across these pools (Karlan and Goldberg 2007).

The aim of this paper is to evaluate the long-term impact of microfinance institution (MFI) credit by overcoming heterogeneities across periodical participant pools. Of particular interest is whether and how differences in the timing of membership (e.g., early versus late) impacted overall livelihoods, particularly in the face of economic distresses such as droughts. Our outcome variable, household consumption expenditure, is composed of food and non-food expenditures, including expenditures on health and schooling of children, and captures the broader dynamics of borrowing effects over time. Moreover, since we explicitly allow for heterogeneity in the timing of borrowing, we identify the overall social impact of credit including that which is due to heterogeneous socio-economic responses to early versus late borrowing. Such an explicit consideration of heterogeneity of timing gives an insight into how MFIs can maximise their social impacts by improving the design and timing of loan products based on heterogeneities in entry and loyalty of clients (Karlan and Goldberg 2007). Moreover, recent studies (e.g., King and Behrman 2009) emphasise that issues of *timing* and *duration* in evaluating social programs are crucial but relatively understudied, and our paper contributes to this literature.

Evaluating such effects is however a difficult task not only because pertinent data spanning sufficiently long periods is scarce but also because obtaining an appropriate 'control group' to identify effects over time, given heterogeneities due to the timing of participation, is difficult (Karlan 2001). Microfinance impact studies thus far have focused on either experimental (e.g., Karlan and Zinman 2007) or quasi-experimental cross-sectional designs (e.g., Pitt and Khandker 1998, Coleman 1999), or classic two-period panel data fixed-effects methods (e.g., Tedeschi 2008) to investigate the causal effects of credit. Cross-sectional designs are useful to tease out biases due to individual borrowers as well as MFI selection characteristics inherent in such programs (e.g., program placement). They, however, lack the time-dimension required to capture lasting effects of credit. Likewise, although the classical fixed-effects method captures long-term effects when panel data is available, it has fundamental flaws when applied to repeated observations that go beyond the classic two-period case.

First, in the case of dichotomous treatment, the fixed-effects model works on the condition that individuals change treatment status (i.e., assumes treatment status is reversible) across time (Wooldridge 2002, 637–638). However, reversibility of treatment status means that impact estimates are biased because dropouts from previous borrowings are included in the control group (contamination-effect) and excluded from the treated group (attrition-effect). An appropriate impact assessment includes dropouts in the comparison group and the control group includes only untreated individuals (Karlan 2001). Second, in time-varying treatments with more than two-period observations, counterfactuals are needed to account for potential future paths of individuals in the control group (Brand and Xie 2007). The fixed-effects method simply compares participants and non-participants pairwise and does not account for such future paths.

The classical parametric evaluation literature is thus silent about these effects and time dimensions in treatments—agents simply are exposed to one of two possible conditions of the treatment at a given time and outcomes are measured subsequent to exposure (Brand and Xie 2007). Owing to the difficulty of establishing appropriate counterfactuals when both the treatment and outcome variables are time-varying, long-term impact evaluations in many applications remain challenging even when panel data is available. Recent developments in nonparametric methods offer alternative ways to handle such identification problems. A body of literature in epidemiology (e.g., Robins, Hernan, and Brumback 2000) and sociology (Brand and Xie 2007) exploits the conceptual apparatus of the ‘potential outcome approach’ in experimental causal inference and extends it to non-experimental panel data. The strategy involves establishing a composite of ‘forward-looking composite counterfactuals’ for each group in each loan-cycle, considering participation as an *irreversible* treatment. These composite counterfactuals combine weighted averages of those who will borrow later between the treatment and the outcome measurement period and those who will never borrow by the end of the outcome measurement periods. This alternative method is used in this paper to overcome the difficulty of identifying credit impact from a four-wave panel data that covers ten years. The data comes from borrower and non-borrower households residing in the zone of operations of a rural MFI in Tigray, northern Ethiopia.

Results indicate that, compared to later participants, early participants are better off even after accounting for initial as well as future counterfactuals. Comparative results show that not accounting for future counterfactuals overestimates impact. It also suggests that the timing of participation matters when it comes to the capacities credit enables households to develop to overcome economic distresses: the earlier, the better. The contributions of this paper are threefold. First, we measure credit impact from long-term panel data accounting for counterfactuals in future pathways, thereby reducing biases due to dropouts and time dimension. This contrasts with conventional methods that compare participants and non-participants pairwise. Second, the propensity score matching method is used to establish appropriate controls for participant groups in each period. Although this non-parametric method is not new, we believe applying it to panel data to establish causality from sequential counterfactuals adds a new dimension to microfinance impact evaluations. Third, this paper provides additional evidence on selection bias, mainly due to the timing of decision to participate and changes in the composition of participants and non-participant, and underlines the danger of employing parametric impact assessments that naively compare participants and non-participants pairwise.

The rest of the paper is organised as follows. By way of describing the structure of the data set used, section 2 discusses the challenges to identify impact using conventional methods in time-varying settings. Section 3 introduces the concept of ‘forward-looking composite counterfactuals’. The techniques to overcome identification problems using this concept and its empirical implementation in this paper are discussed. Section 4 presents and discusses the results. Section 5 concludes.

2. Estimating long-term impacts of periodical participation in microfinance credit

The main objective of this paper is to assess the long-term impact on household welfare of credit with periodical participation of households. Doing so in a more than two-period panel data setting involves solving identification problems due to differences in the *timing of participation*. Before proceeding to discuss why these heterogeneities cause identification problems, it is essential to elaborate on the dynamics of participation using the structure of the dataset at hand. Therefore, the following sections present the dataset structure and discuss the challenges of identifying credit impact using conventional methods in more than the classic two-period settings. An alternative method proposed by Brand and Xie (2007) that exploits the timing of treatment and outcome measurement to mitigate these identification problems is discussed in section 3.

2.1. The structure of the dataset used

Panel data used in this paper come from farm households in Tigray, northern Ethiopia, whose livelihoods largely depend on rainfall based agriculture where production is only once a year. Data were collected during 1997–2006, in four-waves with a three-year interval. Sample households include participants and non-participants of the Dedebit Credit and Saving Institution (DECSI), a MFI that provides credit services in the Tigray region since 1994. First it worked only on trial basis in selected villages, but since 1997 it is present in almost all villages.

The official launch of the MFI program coincided with the first wave of this study. Respondent households were selected using a multi-level sampling procedure. First, 16 villages were selected to represent regional differences, including access to credit. Second, 25 households were randomly selected from each village. Of the total 400 households, 351 are included in the analysis, providing a total of 1404 observations for the balanced four-year panel. The remaining households were excluded because they either became non-targets (outliers) over time, mainly due to old age, or dropped out from the survey. Respondents were surveyed on several characteristics at household-level (e.g., annual consumption levels, participation in MFI credit), village-level (whether or not the village had access to basic infrastructure) and MFI-level (e.g., ease of access to MFI credit). Expenditure values (in local prices) on major household (food and non-food) consumption items were collected¹. Table 2.1 provides summary statistics of annual household expenditure and other variables used in the analysis.

Table 2.1—Summary statistics of yearly household consumption expenditure and other variables used in the propensity score matching Method.

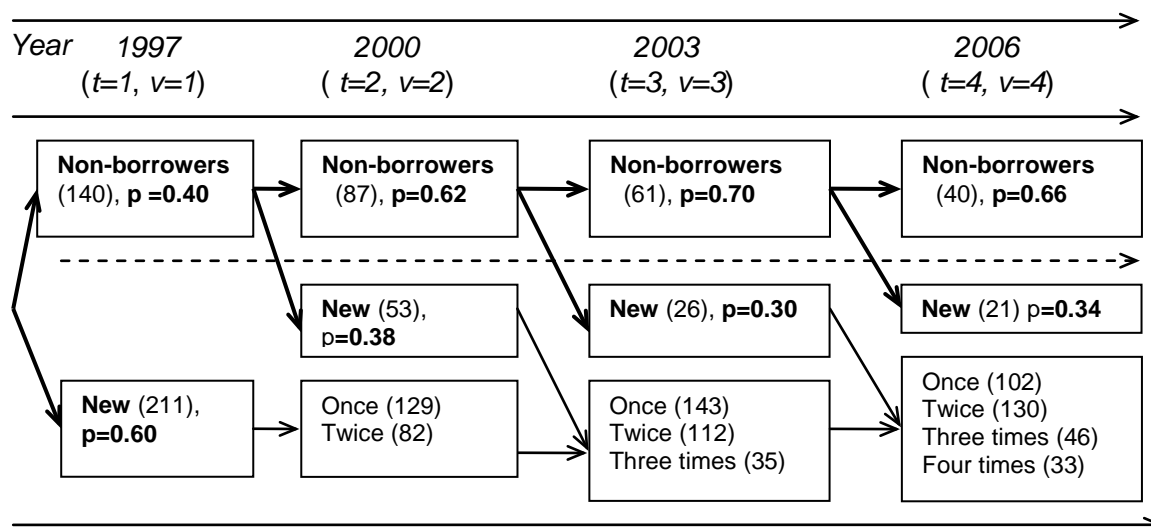
Variable name	Mean	Std.dev.	Min	Max
Participation in program credit	0.450	0.498	0	1
Annual household consumption expenditure	3514.0	4271.0	230	72064
Family size	5.282	2.352	2	13
Male-headed (yes=1)	0.244	0.429	0	1
MFI Branch office close enough (yes=1)	0.762	0.426	0	1
Poorer village (yes=1)	0.239	0.427	0	1
Age of household head	52.9	14.9	19	92
Size of land owned (in <i>tsimad</i> =0.25 ha.)	4.4	3.8	0.25	10.5
Non-farm income dummy	0.742	0.438	0	1
Participation in extension programs (yes =1)	0.165	0.371	0	1
Micro-dam availability	0.514	0.499	0	1

Source: Authors' calculations

¹ A consumer price index for the region is used to adjust for price changes over time (CSA 2008).

An important issue in this dataset for empirical identification is that households joined the MFI at different times during the survey. Of the 351 sample households, 211 participated early (in 1997), while all others joined in later years. So, participation status of members fluctuated over the years. Some members dropped out after participating only once and others continued up to four times. Figure 2.1 presents the number (in parentheses), proportion and frequency of participation in each year. The fact that households switched their status over the survey years makes it possible to apply conventional (e.g., fixed-effects) methods to estimate the impact of participation. However, as will be made clear in section 2.2, these switches also make it difficult to establish appropriate controls for participants in each year.

Figure 2.1—Borrowing participation transition of sample households, 1997–2006



Source: Authors' compilation

Notes: number of participants in parentheses; t = participation period; v = outcome measurement period; p = transition probabilities

2.2. Problems of identifying credit impact in long panels with periodical participation

A straightforward method to estimate impact of periodical participation would have been to exploit the panel nature of the data and use the fixed-effects method. This method uses the time-varying nature of the treatment variable (i.e., participation) across observation periods to identify impact pairwise. Particularly, it exploits the fact that units change their treatment status across observation periods (Chamberlain 1984: 1247–1317). In the case of borrowing, such status changes involve two-way transitions of members between participation and non-participation status over time, which essentially assumes *reversibility* of the treatment. In other words, when previous borrowers drop out at some point, they become part of the control group as if their previous participation does not have an effect on their outcome. Moreover, the fixed-effects method does not use variations of households that do not change status (i.e., always participants and never participants) in explaining outcome variables.

There are therefore three types of biases that arise in using standard parametric methods, such as the fixed-effects method, with panel datasets of more than two periods with dichotomous treatment, as in our setting. First, not all participant households became members of the MFI at the same time. For whatever reason, some joined earlier than others (see Figure 2.1). In other words, there are heterogeneities among the four *new* participant groups due to the timing of membership. One possible reason is that selection criteria by both the MFIs and participants might have changed over time (Tedeschi and Karlan 2010). Moreover, even the degree of participation over the years varies across members; some

continued to participate while others dropped out. Second, estimates are biased if dropping out is non-random (Karlan 2001)², i.e. if dropouts are those who became worse off or better off due to the program. Third, since they were once in the treated (participant) group, dropouts may contaminate the control group carrying over the effect of the treatment to the control group, particularly when the effect of the treatment lasts longer than the treatment period as in borrowing used to acquire durable inputs, e.g., oxen or draft animals. Thus, although long-term panels can be advantageous to capture long-term impacts in many applications, in this particular application the use of such periodical participation and the resulting two-way transition yields biased estimates. As such, partly owing to these problems, most credit impact assessments are limited either to the two-period classic panel data methods or quasi-experimental cross-sectional methods that compare ever-participants to never-participants on pooled data.

3. Empirical method

This section presents an alternative empirical method that overcomes the challenges of identifying impact of periodical participation in a microcredit program, as discussed in section 2.2. The following section briefly introduces a method proposed by Brand and Xie (2007) and its implementation. The Propensity Score Matching (PSM) technique that is used in the implementation of this method is also discussed.

3.1. The forward-looking sequential counterfactual method

A central issue in our setup is that treatment effects are nonreversible and therefore our interest is to identify the causal effect from the timing of participation in credit on outcomes, measured at several points in time subsequent to the treatment. However, when outcomes are measured at different periods other than just subsequent to treatment, it is no longer clear what the appropriate control group should include because the control group varies depending on the gap between the timing of the two events. The concept of 'forward looking sequential counterfactuals' proposed by Brand and Xie (2007) provides a framework to construct appropriate counterfactuals for treatments with lasting effects that vary over time. This concept is useful in applications where (a) exposure to treatment can take place at any point in time but once treated the effect stays on, and (b) the effect of the treatment varies over time subsequent to treatment, regardless of whether or not the treatment is repeated.

In this approach, the control group for individuals treated at t whose outcome is measured at v is composed of individuals that are (a) untreated by t but are treated any time up to v , (b) never treated up to the end of the observation period but can be potentially treated any time in the future. Thus, individuals that participated prior to t can no longer serve as controls at t and the only controls for individuals that have participated at t are therefore those that have never participated up to t . However, control individuals at t may or may not participate after t . Likewise, participants at t would have had the same potential paths after t had they not participated at t . For example, a household that participated in 1997, in our case, would have had two possible paths (i.e., participate or not) in subsequent loan cycles had this household not participated in 1997. This means, if outcome is measured in 2006, this household would

² Karlan 2001 discusses the nature of the bias in cross-sectional designs that exclude dropouts from the treatment group. Tedeschi and Karlan (2010) measure the extent of this bias using quasi-experimental method for two-period panel data from Peru.

have had three other participation chances up to 2006. Brand and Xie (2007) argue not accounting for such counterfactual possibilities biases impact measurement and therefore an appropriate counterfactual outcome must include all potential outcomes of the future paths of the control units considered. This is done by assigning transition probability weights to each potential future path of the controls (see Figure 2.1). A forward-looking sequential approach therefore composes a ‘composite’ of counterfactuals that are weighted combinations of those individuals later treated and those individuals never treated.

However participant and non-participant groups may also differ in their pre-treatment characteristics and participation may be based on these pre-treatment characteristics, in which case estimates are biased even after accounting for future paths. This problem is dealt using the propensity score method assuming the conditional independence assumption discussed in section 3.2. Under this assumption, the Average Treatment Effect on the Treated (*ATT*) of participation in borrowing at time t , $d=t$, on annual household consumption expenditure, C , measured at time v , $v \geq t$, is calculated using forward-looking counterfactual approach as:

$$E(\Delta C_v^{d=t} | X) = E(C_v^{d=t} | X) - E(C_v^{d>t} | X) \quad (1)$$

where the left-hand side (LHS) variable is the *ATT* of participation in borrowing at $d=t$ conditional on X , a vector of exogenous (or pre-treatment) characteristics that determine participation. The terms on the right-hand side (RHS) are expected annual consumption expenditure after $d=t$, conditional on X , for participant and non-participant groups at v , respectively. In our case, these two terms are calculated using PSM, given as the average treatment effects on the participants and non-participants had the latter participated (more on this in section 3.2). The difference between the two provides the *ATT*. Note that $d>t$ in $E(C_v^{d>t} | X)$ indicates that those in the control group are not treated until t but may or may not be treated up to v and this term is a composite of non-participants’ future outcomes expected between t and v . Depending on the causal question asked, this term is decomposed into its future components. That is, decomposition depends on the treatment ($d=t$) as well as the outcome measurement, v periods considered. Note that decomposition is needed only if $v>t$. Otherwise, (1) is reduced to standard pairwise comparison between participants and non-participants.

We illustrate this decomposition using two causal research questions in this paper. First, consider ‘what is the effect of *early* (1997) participation ($t=1$) in MFI borrowing on annual household consumption measured in 2006 ($v=4$)?’ As shown in Figure 2.1, in this specific problem, there are three possible paths for households who did not participate by $t=1$: to participate at $t=2$, to participate at $t=3$, to participate at $t=4$ or not to participate at all, $d>4$. Let the probability to participate at any future path t be given by $p(t)$, otherwise $q(t)=1-p(t)$. Ignoring X for now, the composite term can be decomposed into its path dependent additive components as follows (Brand and Xie 2007):

$$E(C_v^{d>t} | X) \cong E(C_4^{d>1} | X) = [p(2) \cdot E(C_4^{d=2})] + [q(2) \cdot p(3) \cdot E(C_4^{d=3})] \\ + [q(2) \cdot q(3) \cdot p(4) \cdot E(C_4^{d=4})] + [q(2) \cdot q(3) \cdot q(4) \cdot E(C_4^{d>4})] \quad (2)$$

where the terms in square brackets of the *RHS* give the weighted *ATT* of participation at $t=2$, $t=3$, and $t=4$ and non-participation by $t=4$, respectively, measured at $v=4$. Note that the transition probabilities $p(t)$ and $q(t)$ in (2) assign the likelihood of individuals to transit to one of the two paths (participate or not participate) at t conditional on being non-participant prior to t , i.e., $p(t) = \text{prob}(d = t | d \geq t)$ and $q(t) = 1 - p(t)$. We therefore use the probability of being in the participant group in each path such that $p(t) + q(t)=1$ (Brand and Xie 2007).

Combining eq. (1) and eq. (2) gives the *ATT* of *early* participation in MFI borrowing on household consumption expenditure measured in 2006.

$$\begin{aligned}
ATT_1 &= E(\Delta C_v^{d=t} | X) = E(C_v^{d=t} | X) - E(C_v^{d>t} | X) \\
&= E(C_4^{d=1} | X) - \left\{ p(2) \cdot E(C_4^{d=2}) + q(2) \cdot p(3) \cdot E(C_4^{d=3}) \right. \\
&\quad \left. + q(2) \cdot q(3) \cdot p(4) \cdot E(C_4^{d=4}) + q(2) \cdot q(3) \cdot q(4) \cdot E(C_4^{d>4}) \right\}
\end{aligned} \tag{3}$$

Second, consider the causal question, ‘what is the effect of *late* (e.g., in 2003) participation ($t=3$) on annual consumption expenditure measured in 2006 ($v=4$)?’ As before, the composite term in (1) is decomposed into its components³. For non-participants at $t=3$ where $v=4$ (see Figure 2.1), there is only one chance to participate before the survey period ends: to participate or not to participate at $t=4$.

$$E(C_v^{d>t} | X) \cong E(C_4^{d>3} | X) = [p(4) \cdot E(C_4^{d=4})] + [q(4) \cdot E(C_4^{d>4})] \tag{4}$$

Combining eq. (1) and eq. (4) gives the *ATT* of *late* participation (in 2003) in MFI borrowing on household consumption expenditure measure at $v=4$ (i.e., in 2006).

$$\begin{aligned}
ATT_3 &= E(\Delta C_4^{d=3} | X) = E(C_4^{d=3} | X) - E(C_4^{d>3} | X) \\
&= E(C_4^{d=3} | X) - \left\{ p(4) \cdot E(C_4^{d=4}) + q(4) \cdot E(C_4^{d>4}) \right\}
\end{aligned} \tag{5}$$

This procedure is implemented for all ten possible counterfactual constructions in this paper. Comparative results are summarised in the results section from which the causal effect of e.g. *early* versus *late* participation can be made by comparing results obtained from (3) and (5).

Note however that although non-reversible, participation in borrowing is repeatable. The time-varying treatment proposed by Brand and Xie (2007) does not give a way to incorporate the *ATT* of *repeated-borrowers* because treatment is assumed non-reversible and non-repeatable, or simply the treatment state is an *absorbing state* and once treated, individuals remain in the treatment state. The method discussed so far therefore provides the gross effect of credit after the onset of participation. To substantiate results from this method, effects of repeated-borrowing and dropping out are analysed pairwise. These results are provided in sub-section 4.3.

3.2. Propensity score matching method

The composite counterfactual method discussed in section 3.1 gives a way to construct appropriate *future controls* for each participant group in each treatment period. However, participant and non-participant households in each treatment period may not be directly comparable because participant households may self-select (or, be selected) into the program based on initial differences, including the outcome of interest, in which case the mean outcome of the two groups differ even in the absence of the program. Therefore, before proceeding to future counterfactuals, initial comparability must be established to avoid initial selection bias, at least, based on some common observable characteristics.

To deal with this problem, we use the Propensity Score Matching (PSM) technique that has gained popularity in recent years for its potential to remove substantial amount of bias from non-experimental data (e.g., Dehejia and Wahba 1999). This technique helps to adjust for

³ Analysing the effect of participation at $d=t, t+1, T$ on outcomes measured at $v=t, t+1, \dots, T$ follows the same procedure. For a general formula that can be used in many other applications, the interested reader is referred to Brand and Xie (2007).

initial differences between a cross-section of participant and non-participant groups by matching each participant unit to a non-participant unit based on 'similar' observable characteristics. An advantage of PSM is that it summarises all the differences in a single dimension, the propensity score, which is then used to compute treatment effects non-parametrically. The propensity score conveniently summarises the conditional probability of participation given pre-treatment or exogenous characteristics (Rosenbaum and Rubin 1983). An important assumption on which this technique builds on is the Conditional Independence Assumption (CIA), which states that selection is solely based on observable characteristics and potential outcomes are independent of treatment assignment⁴. Meaning, under this assumption, the participants' mean outcome had they not participated can be identified by non-participants' mean outcome. Besides CIA, another condition in PSM is the Common Support requirement, which ensures that individuals compared from the participant and non-participant groups are, to begin with, comparable. Specifically, it ensures individuals with the same observable characteristics have a positive probability of being in both participant and non-participant groups (Heckman, LaLonde, and Smith 1999: 1865). This requirement can be imposed such that estimation is performed on individuals that have common support. The average treatment effect on the treated (*ATT*) is therefore given by the difference in mean outcome of matched participants and non-participants that have common support conditional on the propensity score.

The next practical steps are followed to implement the PSM technique in this paper. The first step is to predict the propensity score for each group in each period using a probit model. Justifying the CIA requires that only variables that simultaneously influence the participation decision and consumption outcome but that themselves are not affected by participation are included (Heckman, Ichimura, and Todd 1998). As such, variables included in our specifications are either measured before treatment or carefully selected exogenous characteristics. Specifications vary across participant groups, accounting for heterogeneities due to 'timing of participation'. Since there are four observed treatments (loan cycles) whose outcomes are measured four times, a total of thirty distinct matching specifications, one for each cross-section, are needed to construct the ten composite counterfactuals (see section 4).

The second step is to choose a method by which weights are assigned for matching. Four different matching algorithms are available in the literature (see Caliendo and Kopeinig 2005). Throughout the paper Kernel Matching (KM) is used. A major advantage of the KM method is that it ensures low variance because it uses weighted averages of all individuals in the control group to construct the counterfactual outcome. Its drawback is however the possibility of bad matches because it uses full information. A recommended solution is to properly impose the common support upon implementation (Heckman, Ichimura, and Todd 1998). Once the propensity score is estimated and used to compute the matching, the third and critical step is to perform a 'balancing test' to check if the matching procedure was effective, i.e. to test if matching balanced observable covariates across treated and control groups. A t-test on equality of means in the treated and control households suggests the extent to which the difference in the covariates between the treated and control groups have been eliminated so that any difference in outcome variable between the two groups can be inferred as coming mainly from the treatment (Heckman and Smith 1995).

⁴ Obviously, this is a strong assumption and there may be bias due to unobservables. Given the complexity of the problem we try to handle in this paper, we hope that this bias may be empirically less important compared to the magnitude of bias that this technique eliminates.

4. Results and discussions

This section presents results based on methods discussed in section 3. Before proceeding to the main results, a pairwise comparison between matched and unmatched households highlights the bias due to initial differences. The main results are provided in the sub-section 4.2 where, after matching among participants and non-participants in each loan cycle, average participation effects using composite counterfactuals are compared to average participation effects from simple pairwise comparisons. Finally, as a robustness check, sub-section 4.3 provides comparative results between matched composite and unmatched pairwise comparisons.

4.1. Matched versus unmatched pairwise comparisons

The effect of participation for both matched and unmatched households is given in Table 4.1. This table gives a simple pairwise comparison of average treatment effect of participants as compared to non-participants (controls) in each year with and without matching. Note that in this table each new participant group's (i.e., $t=1, \dots, 4$) annual consumption is observed over subsequent observation years (i.e., $v=1, \dots, 4$). For example, average annual consumption of new participants in the first observation year ($t=1$) is given under $v=1$ to $v=4$, hence, the table is diagonal. There are two columns under each observation year, which give average impact estimates for the same participant group with and without matching. Comparing the matched against unmatched estimates for the same participant group (in each t) gives the bias reduction after appropriately accounting for initial differences using the matching method. It can be observed that the average effect on each *new* participant's annual consumption (see diagonal) is higher for the unmatched than for the matched. For example, for new participants at $t=1$, average annual consumption increased by 476 Ethiopian Birr (ETB) before matching but after matching, the increase is reduced to 398 ETB. As such, comparing the new participants (on the diagonal of the table) against their appropriately matched controls reduces the average impact. One implication is that, compared to an average non-participant that constituted the unmatched controls, better off households have self-selected into the program in each year. Clearly, comparing the new participants in each year against non-participants, without accounting for initial differences, in our case, overestimates impact. The bias remains even after entry (off-diagonal). However, the direction of bias differs from year to year.

4.2. Matched composite versus matched pairwise comparisons

The pairwise comparison of participation effects between matched and unmatched groups in the previous section suggests that matching reduces a significant part of bias due to differences prior to entry. However, it does not account for how participating households would have fared in subsequent years if they had not participated. This effect is captured by the composite counterfactual estimated according to methods provided in section 3.2. Main results are given in Table 4.2. For comparison purposes, matched pairwise effects are also provided for each outcome measurement period along with the composite effects. The columns in Table 4.2 provide these comparative results for each group of participant ($t=1, \dots, 4$) and in each outcome measurement period ($v=1, \dots, 4$). Note that standard errors are hard to generate for composite results because they are calculated from several matching results according to methods in section 3.1 and 3.2.

There are two important findings in this exercise. First, regarding the main causal question of comparing the effect of early versus late participation, the composite counterfactual results suggest that early participants have consistently fared better than late participants. Specifically, after accounting for both initial differences and potential future changes in the composition of participants and their controls, long-term participants have enjoyed relatively

higher average annual consumption than short-term participants. In Table 4.2, this can be seen by comparing the composite effects for each new participant group (i.e., at each t) against its preceding participant column wise. Note that the composite effect, for the most part, declines going from top (early participants) to bottom (late participants) in each column. One reason is that the effect of borrowing lasts longer than the specific period it refers to and that long- rather than short-standing participants are more likely to enjoy higher effects in terms of capacity to smooth consumption over time. Another reason is that, since participation is state dependent, chances to repeat participation will result in further increases in consumption for the early participants in comparison to those whose participation in MFI borrowing began later.

Second, in contrast to simple pairwise effects, the composite effects provide conservative results in all comparisons except for the initial year⁵. This is because the composite effects take future potential counterfactuals into account whereas the pairwise estimates do not. In other words, not accounting for future potential counterfactuals overestimates impact. This is so because not participating in any earlier year does not preclude the possibility of participating in any later year and, given positive effects of participation, not accounting for these chances of later participation overestimates impact of early participation. This can be elaborated using the most early participants (i.e., $t=1$) whose outcome is measured in four periods. The composite effect in the last outcome measurement period ($v=4$) takes into account the fact that some of their matched controls (i.e., non-participants at $t=1$) have been able to participate at $t=2$, $t=3$ or at $t=4$. This reduces the average effect from ETB 1488 to ETB 1238. Clearly, the difference is the counterfactual for early participants had they not participated at $t=1$. Thus, failing to account for the different future pathways between participation and outcome measurement periods overstates the effect of (early) participation.

⁵ Note that in each initial year, the pairwise effect is the same as the composite effect because $t=v$ and there is no need to account for future potential counterfactual.

Table 4.1—Simple pairwise comparisons between matched and unmatched households: Average participation effects on yearly household consumption expenditure

Timing of participation	1997 (v=1)		2000 (v=2)		2003 (v=3)		2006 (v=4)	
	Matched	Unmatched	Matched	Unmatched	Matched	Unmatched	Matched	Unmatched
t=1	398.045*** (109.118)	475.671*** (112.671)	529.716* (281.947)	506.302* (282.546)	371.162** (146.650)	268.145* (1.820)	1487.966*** (546.147)	1663.482*** (540.301)
t=2			133.156 (428.498)	258.850 (380.228)	-296.506 (263.564)	-439.124* (247.017)	1457.280* (871.109)	1135.41** (751.849)
t=3					717.044* (462.315)	739.875** (378.987)	497.059 (1178.491)	1398.763** (918.440)
t=4							429.150 (1591.532)	2041.472*** (934.478)

Source: Authors' calculations

Note: ***, **, *, significant at the 1%, 5%, and 10% levels, respectively; Standard errors in parentheses.

Table 4.2—Matched composite versus matched pairwise comparisons: Household consumption expenditure measurement

Timing of participation	1997 (v=1)		2000 (v=2)		2003 (v=3)		2006 (v=4)	
	Composite	Pairwise	Composite	Pairwise	Composite	Pairwise	Composite	Pairwise
t=1	398.045	398.045*** (109.118)	388.733	529.716* (281.947)	859.674	371.162** (146.650)	1238.704	1487.966*** (546.147)
t=2			133.156	133.156 (428.498)	-568.872	-296.506 (263.564)	524.370	1457.280* (871.109)
t=3					717.044	717.044* (462.315)	-292.578	497.059 (1178.491)
t=4							429.150	429.150 (1591.532)

Source: Authors' calculations

Note: ***, **, *, significant at the 1%, 5%, and 10% levels, respectively; Standard deviations in parentheses.

Obviously, many factors other than borrowing dictate changes in consumption levels over time and, with a slight downturn in 2003, average consumption increased between $v=1$ and $v=4$ for both participants and non-participants, albeit at different pace. Specifically, except in the bad year 2003 in which case there was a consumption downturn, the pairwise causal effects, for the most part, overestimate impacts because the counterfactual paths are not taken into account.

Evidently, conventional parametric impact assessments that compare 'ever participants' to 'never participants' without considering the timing of the decision to participate and the different potential future pathways an individual household might have followed in the absence of the program would yield biased estimates.

Finally, given the relatively long period the dataset covers, including two drought years (1999/2000 and 2003), it is interesting to see the implications of the effects of these differences in timing of participation on household consumption and hence relative capacity to cope with vulnerability during and after the drought years. Composite effects of participation in the first three periods i.e., at $t=1, 2,$ and $3,$ on annual consumption during the last three outcome measurement periods, i.e., $v=2, 3,$ and $4,$ are of interest here. Compared to controls, results suggest that the average annual consumption of the earliest ($t=1$) participants has increased steadily, including during and post drought years. Intuitively, sufficient time is needed for the cumulative impact of credit to take effect (King and Behrman 2009). This is however not the case for later ($t=2$ and $t=3$) participants. In fact, although both participant groups have seen increased average consumption in the year they participated (which happened to be the drought years for both), in both cases, it has declined a year after participation (post drought years). A possible explanation for this is that households might have diverted loans to smooth consumption in the drought years, a common phenomenon despite DECSI's claims of 'productive' use of credit. A study on the same MFI by Borchgrevink et al. (2005: 68–69) finds indications of use of credit given for production purposes being diverted to consumption during drought periods. Moreover, the fact that loans are repaid after one year seems to explain the relative decline in participant households' consumption in the post drought periods. Nevertheless, for the $t=2$ participants, the result suggests this decline had been reversed in 2006⁶. It can therefore be concluded that relative to non-participants, earlier participants gained better capacities to cope with shocks and the earlier the better. This conclusion has to be taken with caution though, because the results explicitly compare variations of average consumption due to credit and not overall consumption variability due to shocks.

4.3. Effects of changes in the composition of treatment and control groups in time-varying treatments

In sections 4.1 and 4.2 interest centred on the effect of timing of first-time participation in MFI borrowing on annual household consumption in subsequent years. Thus, the analysis was mainly based on entry, considering borrowing as an irreversible regime, i.e., once households participate, they remain members thereafter. However, there are borrowing dynamics after this entry event took place. Particularly, once in the borrowing regime, some households participated repeatedly, and others participated occasionally or never repeated at all. Details about these dynamics are given in Figure 2.1. This section presents effects of such borrowing dynamics, mainly effects of dropping out and repeated-borrowing in a particular year. Due to the complexity of applying composite counterfactuals, effects are

⁶ For $t=3$ participants, the effect in subsequent years is not known because the observation period does not allow for this.

analysed pairwise. However, comparative results are provided such that the biases in handling dropouts discussed in section 2 are also evaluated.

Table 4.3 presents comparative results of (a) including dropouts in control groups but not in treatment groups (comparison 1) in contrast to including them in treatment groups but excluding them from the control groups (comparison 2). This comparison shows the bias in using the standard fixed-effects method that would naively include dropouts in control groups; (b) excluding dropouts from both and excluding new participants from treatment groups (comparison 3). In comparison 3, only repeated participants are compared to their periodical non-participant counterparts. This can be used as reference for comparisons 1 and 2. Moreover, this last comparison provides the effect of repeated participation on outcomes measured in different periods.

Comparing comparison 1 and 2 in Table 4.3, shows that including dropouts in control (but not in treatment) groups biases impact even after adjusting for initial conditions using matching. However, the direction of the bias depends on the relative condition of dropouts in each year, which in turn may depend on the economic conditions in the year before. That is, in years when dropouts are those that have become worse off after participation, excluding them from the treatment group overstates impact. On the other hand, in years when dropouts are better off households, excluding them from treatment groups overstates impact both due to attrition and contamination effects. For participants in 2000, the results show that excluding dropouts from the treated groups and including them in the control group understates impact in all outcome measurement periods (see matched columns, Table 4.3). On the contrary, excluding dropouts from the treatment group (including them in the control group) overstates impact because not only were relatively worse off participants selectively dropped out in the bad year 2003, but also the new participants in the same year were relatively better off. The latter can be seen by comparing the average participation effect of ETB 717 for matched new participants (Table 4.2) and ETB 302 for matched new and repeat participants (Table 4.3) in 2003. The same is true for participants in 2006. This means there is not only a strong selection processes at work, but also the direction of selection depends on the underlying economic conditions households face in each period. Note that in the unmatched case, the same comparison provides a different picture, concealing the selection processes.

Lastly, the last part of Table 4.3 (comparison 3) provides the effect of repeated participation, excluding dropouts and new participants. Once again, comparing the veteran participants to their non-participant counterparts by excluding dropouts overstates impact even when new participants are excluded, e.g. for participants in 2000, because only worse off participants dropped out systematically. Thus, for the most part, results in comparison 3 are higher than in comparison 1, which are in turn, for the most part higher than results in comparison 2, reflecting the consistency of the results obtained.

Table 4.3—Effects of repeat-participation and changes in the composition of participants: matched and unmatched pairwise comparisons

Average participation effects (ATT) on yearly household consumption expenditure

Composition of treatment & control groups	Treatment periods	1997 (v=1)		2000 (v=2)		2003 (v=3)		2006 (v=4)	
		Matched	Unmatched	Matched	Unmatched	Matched	Unmatched	Matched	Unmatched
Comparison 1:	<i>t</i> =1	398.045*** (109.118)	475.671*** (112.671)	529.716* (281.947)	506.302* (282.546)	371.162** (146.650)	268.145* (1.820)	1487.966*** (546.147)	1663.482*** (540.301)
<u>Treatment group:</u> Repeat and new participants at <i>t</i>	<i>t</i> =2			443.185 (371.215)	590.496* (363.088)	-191.687 (223.447)	-240.763 (179.716)	2350.142*** (695.203)	2382.831*** (694.100)
<u>Control group:</u> Non-participants at <i>t</i> (dropouts excluded)	<i>t</i> =3					-32.794 (294.988)	492.149** (211.200)	2113.532*** (708.526)	2886.439*** (722.815)
	<i>t</i> =4							1058.209 (970.112)	1656.060 (1347.406)
Comparison 2:	<i>t</i> =1	398.045*** (109.118)	475.671*** (112.671)	529.716* (281.947)	506.302* (282.546)	371.162** (146.650)	268.145* (1.820)	1487.966*** (546.147)	1663.482*** (540.301)
<u>Treatment group:</u> Repeat and new participants at <i>t</i>	<i>t</i> =2			388.838 (307.474)	305.734 (285.217)	-235.219 (143.111)	-326.401 (148.135)	1371.132*** (591.648)	1548.644*** (544.875)
<u>Control group:</u> Non-participants at <i>t</i> (dropouts included)	<i>t</i> =3					301.201** (150.605)	386.784*** (149.853)	1177.210** (581.238)	1620.214*** (552.194)
	<i>t</i> =4							1129.253* (771.022)	1311.087* (756.857)
Comparison 3:	<i>t</i> =1	398.045*** (109.118)	475.671*** (112.671)	529.716* (281.947)	506.302* (282.546)	371.162** (146.650)	268.145* (1.820)	1487.966*** (546.147)	1663.482*** (540.301)
<u>Treatment group:</u> Only repeat participants at <i>t</i>	<i>t</i> =2			790.598* (488.708)	804.851* (426.080)	-130.691 (240.557)	-112.555 (215.603)	3148.094*** (838.369)	3189.090*** (776.011)
<u>Control group:</u> Non-participants at <i>t</i> (dropouts excluded)	<i>t</i> =3					-208.260 (318.011)	348.360* (208.593)	2193.486*** (752.784)	3273.235*** (733.314)
	<i>t</i> =4							1190.914 (2177.474)	1837.332 (1410.308)

Source: Authors' calculations

Note: ***, **, *, significant at the 1%, 5%, and 10% levels, respectively; Standard errors in parentheses.

5. Conclusions

This paper dealt with the methodologically challenging question of assessing the impact of differences in the timing of membership in a microfinance credit program, using a four-wave panel data covering ten years that comes from a rural microfinance institution in Ethiopia. Specifically, the main questions addressed include (a) whether or not early participation, as opposed to late participation, matters in terms of increases in average annual household consumption, (b) a methodological issue of consideration of potential future pathways of control groups when the timing of participation and outcome measurement are different, and (c) overcoming impact estimation biases due to the dynamics of borrowing, mainly dropouts and new entrants.

In the empirical methodology section of the paper, we argue that parametric impact assessment methods such as the fixed-effects method may yield biased estimates when the treatment variable is dichotomous and units are observed more often than in the classic two-period panel data case. This is because such methods exploit individuals who are 'on' and 'off' the treatment over time, which contaminates the impact estimate. An alternative method is used that treats participation in credit as an irreversible regime and identifies impact from the timing of onset of participation. As such, only non-participant households up to the time of entry of participants are considered as candidates for control. The propensity score matching is used to balance potential initial heterogeneities among participants and non-participant controls.

The results of the propensity score matching indicate that matching participants and non-participants on some basic pre-treatment characteristics reduces substantial amount of selection bias. Comparisons between matched and unmatched average treatment effects suggests that over the years 'better off' households tend to participate. The results from the main analysis indicate that early participants enjoyed higher average annual consumption over time, compared to late participants. Comparing the composite effects against simple pairwise effects, the composite effects provide conservative results because they take future counterfactuals into account. Thus, not accounting for potential future pathways overstates impact. Note that since we explicitly take timing heterogeneities into account and the impact indicator incorporates household expenditure in a broader sense, including health and schooling, the impact estimates provide an overall assessment of the social impacts of credit. Thus, the results can be taken as suggesting that participants have benefited in terms of broader socio-economic aspects, such as health and education of their children. Moreover, credit has a better shock-cushioning effect for earlier participants.

The analysis has also considered effects of the dynamics of borrowing once in the borrowing relationship. Particularly, pairwise effects of repeated-borrowing and dropping out are considered in a comparative way. Results suggest including dropouts in control (but not in treatment), as in a fixed-effects approach, biases impact even after adjusting for initial conditions using matching. However, the direction of the bias depends on the relative condition of dropouts in each year. Further, the comparative analysis indicates that, because of the selection processes at work when individuals drop out or repeat, impact assessments that compare participants and non-participants over a long period but that only adjust for initial conditions and not for the dynamics of participation over the study period are likely to be biased.

References

- Brand and Xie. 2007. "Identification and estimation of causal effects with time-varying treatments and time-varying outcomes." *Sociological Methodology* 37(1): 393–434.
- Borchgrevink, A., T. Woldehanna, G. Ageba, and W. Teshome. 2005. *Marginalised groups, credit and empowerment: the case of Dedebit Credit and Saving Institution (DECSI) of Tigray, Ethiopia*. AEMFI Occasional Paper 14. Addis Ababa, Ethiopia: Association of Ethiopian Microfinance Institutions.
- Caliendo, M. and S. Kopeinig. 2005. *Some practical guidelines for the implementation of Propensity Score Matching*. IZA Discussion Paper 1588. Bonn, Germany: Institute for the Study of Labor (IZA).
- CSA (Central Statistical Agency of Ethiopia). 2008. *Consumer Price Index at Country and Regional Levels*. Addis Ababa: CSA, The Federal Democratic Government of Ethiopia.
- Chamberlain, G. 1984. "Panel Data." In *Handbook of Econometrics*. Vol 2, edited by Z. Griliches, and M.D. Intriligator. Cambridge, MA: MIT Press.
- Coleman, B. 1999. "The impact of group lending in northeast Thailand." *Journal of Development Economics* 60 (1): 105–141.
- Dehejia, R.H., and S. Wahba. 1999. "Causal effects in non-experimental studies: reevaluating the evolution of training programs." *Journal of American Statistical Association* 94: 1054–1062.
- Heckman, J., H. Ichimura, and P. Todd. 1998. "Matching as an Econometric Evaluation Estimator." *Review of Economic Studies* 65: 261–294.
- Heckman, J., R. LaLonde, and J. Smith. 1999. "The Economics and Econometrics of Active Labor Market Programs." In *Handbook of Labor Economics*. Vol. III, edited by O. Ashenfelter, and D. Card. Amsterdam: Elsevier.
- Heckman, J., and J. Smith. 1995. "Assessing the case for social experiments." *Journal of Economic Perspectives* 9: 85–110.
- Karlan, D., and N. Goldberg. 2007. *Impact Evaluation for Microfinance*. Doing Impact Evaluation Series 7. Washington, DC: Thematic Group on Poverty Analysis, Monitoring and Impact Evaluation, World Bank.
- Karlan, D., and J. Zinman. 2007. *Expanding Credit Access: Using Randomised Supply Decisions to Estimate the Impacts*. CEPR Discussion Paper DP6407. London: Centre for Economic Policy Research.
- Karlan, D. 2001. "Microfinance impact assessments: the perils of using new members as a control group." *Journal of Microfinance* 3 (2): 75–85.
- King, E.M., and J.R. Behrman. 2009. "Timing and duration of exposure in evaluations of social programs." *The World Bank Research Observer* 24 (1): 55–82.
- Menon, N. 2006. "Long-term Benefits of membership in Microfinance Programs." *Journal of International Development* 18 (4): 571–595.
- Morduch, J. 1998. "Does microfinance really help the poor? New evidence from flagship programs in Bangladesh." Manuscript, Princeton University.
- Pitt, M., and S. Khandker. 1998. "The impact of group-based credit programs on poor households in Bangladesh: does the gender of participants matter?" *Journal of Political Economy* 106 (5): 958–996.

- Robins, J.M., M.A. Hernan, and B. Brumback. 2000. "Marginal Structural Models and Causal Inference in Epidemiology." *Epidemiology* 11: 550–560.
- Rosenbaum, P., and D. Rubin. 1983. "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70 (1): 41–55.
- Tedeschi, G. A. and D. Karlan. 2010. "Cross-Sectional Impact Analysis: Bias from Dropouts." *Perspectives on Global Development and Technology* 9 (3–4): 270–291.
- Tedeschi, G.A. 2008. "Overcoming Selection Bias in Microcredit Impact Assessments: A Case Study in Peru." *Journal of Development Studies* 44 (4): 504–518.
- Wooldridge, J.M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.