American Economic Journal: Applied Economics 2015, 7(1): 54–89
http://dx.doi.org/10.1257/app.20130475

The Impacts of Microcredit: Evidence from Ethiopia[†]

By Alessandro Tarozzi, Jaikishan Desai, and Kristin Johnson*

We use data from a randomized controlled trial conducted in 2003–2006 in rural Amhara and Oromiya (Ethiopia) to study the impacts of increasing access to microfinance on a number of socioeconomic outcomes, including income from agriculture, animal husbandry, nonfarm self-employment, labor supply, schooling and indicators of women's empowerment. We document that despite substantial increases in borrowing in areas assigned to treatment the null of no impact cannot be rejected for a large majority of outcomes. (*JEL* G21, I20, J13, J16, O13, O16, O18)

Beginning in the 1970s, with the birth of the Grameen Bank in Bangladesh, microcredit has played a prominent role among development initiatives. Many proponents claim that microfinance has had enormously positive effects among borrowers. However, the rigorous evaluation of such claims of success has been complicated by the endogeneity of program placement and client selection, both common obstacles in program evaluations. Microfinance institutions (MFIs) typically choose to locate in areas predicted to be profitable, and/or where large impacts are expected. In addition, individuals who seek out loans in areas served by MFIs and that are willing and able to form joint-liability borrowing groups (a model often preferred by MFIs) are likely different from others who do not along a number of observable and unobservable factors. Until recently, the results of most evaluations could not be interpreted as conclusively causal because of the lack of an appropriate control group (see Brau and Woller 2004 and Armendáriz de Aghion and Morduch 2005 for comprehensive early surveys). In this context, randomized controlled trials (RCTs) provide an ideal research design to evaluate the impact of microcredit.

In this paper we present the results of one of the few existing RCTs that evaluate the impact of introducing access to microloans in poor communities in a developing country after the early contribution of Banerjee et al. (2014). We study a large-scale clustered RCT conducted in rural Amhara and Oromiya (Ethiopia) between 2003 and 2006. The main purpose of the RCT was to evaluate whether the

* Tarozzi: Department of Economics and Business, Universitat Pompeu Fabra and Barcelona GSE, Ramon Trias Fargas, 25-27 08005 Barcelona, Spain (e-mail: alessandro.tarozzi@upf.edu); Desai: Health Services Research Centre, School of Government, Victoria University of Wellington, PO Box 600, Wellington, New Zealand (e-mail: jaiki.desai@vuw.ac.nz); Johnson: Harvard Business School, 4229 S. Coors St., Morrison, CO 80465 (e-mail: kristinjohnson36@gmail.com). We are very grateful to Family Health International for granting us access to the data and to Charles Becker, Cristina Czura, Cynthia Kinnan, Simon Quinn, Duncan Thomas, Chris Woodruff, Esther Duflo (the Editor), two anonymous referees and seminar participants at several seminars and workshops for comments and suggestions. All errors and omissions are our own. The trial described in the paper has been registered after the conclusion of the study with the AER RCT Registry, with registry number AEARCTR-0000305.

[†] Go to http://dx.doi.org/10.1257/app.20130475 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

contemporaneous introduction of microcredit and community-driven family planning programs (FPPs) could have a larger impact on contraceptive use than either program operating on its own. The study was conducted using a 2×2 factorial design where 133 local administrative units called *kebeles* or "peasant associations" (PAs) were randomly assigned to one of four groups: microlending only, FPP only, both, or none (control). Despite the primary emphasis on contraceptive use, household surveys conducted before and after the interventions also measured a broad range of socioeconomic outcomes, including income-generating activities, livestock ownership, schooling and measures of women's empowerment. Because the study design assigned a randomly determined subset of community to have access to microloans, the data can thus be used to gauge the impact of increased access to microfinance on the economic lives of households in study areas.

Ours should be a useful addition to a small number of RCTs that evaluate how increasing access to microloans *at the community level* may affect socioeconomic outcomes in poor countries. Other studies evaluate impacts in urban slums in Hyderabad, India (Banerjee et al. 2014), in rural and urban areas of the state of Sonora, Mexico (Angelucci, Karlan, and Zinman 2014), in rural Mongolia (Attanasio et al. 2011) and rural Morocco (Crépon et al. 2011).[1] Other RCTs have estimated the impact of access to microcredit by randomizing at the individual level among microcredit clients close to the threshold of eligibility for loans, in urban South Africa (Karlan and Zinman 2010), urban Philippines (Karlan and Zinman 2011), and Bosnia-Herzegovina (Augsburg et al. 2012).

In our study, we use data collected during a preintervention household survey carried out in 2003 ("baseline") and a postintervention survey completed three years later ("follow-up" or "endline"). Each survey was conducted by interviewing about 6,000 households, with similar sample size from the two regions of Amhara and Oromiya. At follow-up, interviewers revisited the same study villages, but they did not seek to reinterview the same households, so our data constitute a panel of villages but not of households. Baseline and follow-up samples were drawn independently of each other, and independently of program participation.

The RCT was conducted in poor rural areas where agriculture and animal husbandry represented the bulk of the local economic activities. Borrowing was not common and at baseline just above one household in every ten had any outstanding loans. The intervention aimed at increasing access to microcredit in program areas through the entry of two MFIs, the Amhara Credit and Savings Institute (ACSI) in Amhara and the Oromiya Credit and Savings and Share Company (OCSSC) in Oromiya. In several respects, both institutions were typical microlenders that granted small loans to small and self-formed groups of borrowers who took joint responsibility for loan repayment. Loan eligibility was supposed to be determined on the basis of several criteria, of which the presence of a viable business plan and poverty status were the most salient ones. Lending was supposed to especially

[1] Pronyk et al. (2006) randomized access to loans for women (together with a gender and HIV training curriculum) in a randomly chosen half of study villages in rural South Africa. However, only eight villages were included in the study, and key results are not based on a treatment-control comparison but rather on comparing women who self-selected into the program in treated areas with others from control communities matched based on observed characteristics.

target women, although guidelines in this regard were loose, and indeed we find that a majority of loans were initiated by men. In addition, unlike many microfinance institutions (MFIs), OCSSC and ACSI often required some forms of collateral from borrowers (as in the program studied in Attanasio et al. 2011), making it harder for the very poorest households to have access to loans. Both ACSI and OCSSC started lending shortly after the baseline survey and continued to do so until the time of the endline survey, after about three years. Our RCT is thus characterized by one of the longest time spans between preintervention and postintervention surveys among the studies cited earlier.

We will show that borrowing increased substantially more in treated relative to control communities, both on the intensive and the extensive margin. In areas where ACSI/OCSSC had been assigned to enter, borrowing prevalence increased by about 25 percentage points relative to control areas, and the average amount of outstanding loans was almost twice as large. Almost all the increase in borrowing was due to microloans from ACSI/OCSSC, suggesting that rather than displacing other forms of preintervention borrowing, the introduction of microlending led to substantial relaxation of credit constraints. In addition, borrowing from other MFIs changed very little over time, so what we identify is the impact on first-generation borrowers from microlenders, similar to Crépon et al. (2011).

Despite the large increase in borrowing, we find that for a large majority of socioeconomic outcomes the null of no impact cannot be rejected, although in several cases the point estimates are substantively large but imprecisely estimated. For instance, we estimate that in areas assigned to microcredit a 95 percent confidence interval for the value of livestock owned is consistent both with a 25 percent increase and a 10 percent decrease relative to communities assigned to control. Similarly, the 95 percent confidence interval for the impact on revenues from self-employment activities is consistent with a doubling of revenues or a 15 percent decrease relative to control areas. In several cases, the width of the confidence intervals is thus large enough to be consistent with both substantial improvements and large declines. On the other hand, our estimates are sufficiently precise to reject (at the 5 percent level) the null that assignment to the microcredit treatment increased the prevalence of new nonfarm business creation by more than 1.4 percentage points relative to control areas.

Such inconclusive results are at least partly due to insufficient power, especially for outcomes such as revenues and profits that are hard to measure and characterized by large variances. We also highlight that, unlike all the other RCTs cited earlier, our data do not include information on consumption. This is unfortunate, because in other contexts it has been shown that access to microcredit, while leaving aggregate consumption largely unchanged, can increase consumption of durables while decreasing expenditures in "temptation goods," such as cigarettes or alcoholic beverages. Due to data limitations we cannot document if similar patterns emerged in our study areas.

The rest of this paper is organized as follows. Section I describes the details of the intervention and the study design. Here we also discuss how the microlenders partly deviated from the experimental protocol, starting operations in some areas assigned to control while doing the opposite in some PAs assigned to treatment. Because such deviations from protocol potentially invalidate the exogeneity of treatment, we focus on intent-to-treat estimates, interpreted as impacts of assignment to microcredit. We include the details of the estimation strategy in Section II, where we also describe the results. Finally, Section III concludes, after discussing our findings in relation to the literature.

I. Study Design and Baseline Summary Statistics

The study was implemented over a large geographical area in rural western Ethiopia, spread over 133 administrative units called *kebeles* or “peasant associations” (PAs) from two “zones” of the Oromiya region and two zones in the Amhara region, about 300–400 kilometers respectively west and north of the capital Addis Ababa, see the map in Figure 1.[2] The main sources of income in study areas were agriculture and animal husbandry, in some cases supplemented with small-scale retailing activities or day labor. Unlike the arid eastern regions, the study locations usually benefit from plentiful precipitation, with an average of 1,200–2,000 millimeters of rainfall per year in 1971–2000 in both regions.

The study area was identified in the context of the expansion of microcredit and FPPs supported by the David and Lucille Packard Foundation Population Program. The expansion was implemented in Oromiya by the Oromiya Credit and Savings Share Company (OCSSC) and the Oromiya Development Association (ODA) and in Amhara by the Amhara Credit and Savings Institute (ACSI) and the Amhara Development Association (ADA). The research team identified 191 villages in 78 PAs where OCSSC and ODA planned to expand in the coming years in Oromiya, and 162 villages in 55 PAs where ACSI and ADA planned to expand.

In each of the 133 study PAs, interview teams obtained a complete list of all villages with an estimate of the total number of households in each village. If the PA had more than 400 households, three villages were randomly selected. If the PA had fewer than 400 households, two villages were selected at random. Within each selected village, interview teams conducted a complete enumeration of households, and a random sample of households was chosen to participate, with interviews completed between January and May of 2003. In all, 6,412 households were interviewed at baseline, of which 3,196 were in Amhara and 3,216 in Oromiya. The sample is not self-weighted and therefore sampling weights are required to produce unbiased estimates of population statistics. We use sampling weights throughout the paper, although the unweighted results are generally very similar.

A. *Experimental Design*

The data used in this paper were collected *as two independent cross sections from the same villages* in 2003 and 2006 as part of the evaluation of a cluster randomized

[2] Peasant associations are the smallest local unit of government in Ethiopia and comprise a number of villages. PAs are then grouped into “woredas,” which are then aggregated into “zones,” and zones into regions (Ofcansky and LaVerle 1991). The eight study woredas in Oromiya are Anfillo, Ayra Guliso, Haru, Mana Sibu, Nedjo, and Seyo in West Wellega zone, and Metu and Chora in Illubabor zone. In Amhara, they are Bugna, Gidan and Meket Delanta in the Semien (or “North”) Wollo zone, and Metema, Chilga, Alefa Takusa, and Lay Armachiho in North Gonder zone.

Figure 1. Study Areas

Notes: Each contour represents an administrative unit called a *woreda*. The woredas where study PAs were located are shown in black. The northern-most woredas are those in Amhara and the southern-most ones are those in Oromiya.

Source: Geo-spatial data from http://maps.worldbank.org

controlled trial (RCT) conducted by Family Health International. The main focus of the RCT was on fertility choices and contraception, and its primary purpose was to determine whether linking microcredit with family planning services could increase the use of contraception beyond what each program could accomplish separately. As part of this evaluation, after the completion of the baseline survey the 78 PAs in Oromiya were randomly scheduled to see the introduction of microcredit (20 PAs), family planning services (18), both (20), or neither (20). The 55 PAs in Amhara were assigned as follows: microcredit (14), family planning services (13), both (15), or neither (13). Randomization into the three treatment groups and one control group was completed independently in each of the two regions, with no further stratification. Random assignment to experimental arms was conducted at the PA level, so that all sample villages and households from the same PA were assigned to the same group.[3] Randomization produced roughly 800 households in each of the 4 original treatment groups.

[3] The random assignment of PAs to treatment arms was conducted using statistical software by a biostatistician at Family Health International, North Carolina, United States.

The community-based FPPs were based on women from local communities trained and remunerated to make house-to-house visits, provide fertility-related information and offer contraceptives at no cost. In areas where both services were introduced, credit officers also provided information on family planning methods to women borrowers (but did not offer contraceptives). In principle, the FPPs could have had an impact on economic activities via a change in family planning. However, Desai and Tarozzi (2011) show that the programs (both in isolation and when jointly present) failed to modify contraceptive behavior, and were only weakly associated with changes in fertility.[4] For this reason, in this paper we choose to focus only on the impact of increased access to microcredit, although in our preferred estimates we also control for the presence of FPPs, in isolation or in addition to microlending, see Section II for details.

Both ACSI and OCSSC, the two microfinance institutions (MFIs) that partnered with Packard for this evaluation, are development-oriented institutions with strong links to the government. Prior to the end of the civil war in 1991, all banking and insurance activities were monopolized by the Ethiopian government. Proclamation No. 84/94 was later issued allowing private domestic investors to also participate in these activities, but the government maintained a strong involvement in the evolution of Ethiopian MFIs, which overall operate in a noncompetitive environment (Wolday 2002). At the time of writing, government-supported microenterprise lending program encompasses about 30 MFIs registered, licensed, and regulated by the National Bank of Ethiopia, including ACSI and OCSSC.[5] ACSI began as a project of the NGO Organization for the Rehabilitation and Development in Amhara, and was officially established as a microfinance institution in 1997.[6] Its stated mission is to "improve the economic situation of low income, productive poor people in the Amhara region through increased access to lending and saving services." OCSSC was also established in 1997, and was born out of the Oromiya Rural Credit and Savings Scheme Development Project, with the stated mission to "provide need-based microfinancial services to strengthen the economic base of the low-income rural and urban people in Oromiya through increased access to sustainable and cost efficient financial services."

Both ACSI and OCSSC operated on the basis of group lending. Small and self-formed groups of borrowers, who took on collective responsibility for repayment of loans, were selected on the basis of several criteria, of which business plan and poverty status were the more salient ones. Loans were made for one year at interest rates reflecting market conditions. Based on OCSSC and ACSI records, the interest rate in 2002–2003 was 12 percent per year on average. Credit officers helped fill out loan applications and monitored the groups. Borrowers were expected to make regular deposits and repayments. Both OCSSC and ACSI reported repayment rates higher than 95 percent in the years before the intervention. In both regions, the credit program expansion was supposed to target poor women borrowers, but

[4] The most likely reason is that the FPP did not provide injectables (the main contraceptive method in these regions), although referrals to clinics were provided.

[5] See http://www.aemfi-ethiopia.org/site/membership.html.

[6] For more information see http://www.acsi.org.et.

in reality no strict guidelines about the gender of the borrowers was issued. For this reason, as we will show, loans were often granted to individuals of both genders. The guidelines of both microlenders mentioned that no collateral was required in order to have access to loans, although our postintervention data suggest otherwise, with a majority of borrowers indicating that they had been asked for collateral.

The microlenders were directed to start lending in program areas shortly after the baseline survey, and to continue to do so thereafter. Service data collected in study PAs to verify program implementation indicate that by the end of 2003 ACSI/OCSSC were already granting loans in 63 percent of treated PAs, and the proportion grew to 82 percent by the end of 2004. In a large majority of treated communities, program exposure was thus as long as 2–3 years.

Before discussing the baseline summary statistics, it is important to highlight limitations of this study related to statistical power. Although both preintervention and postintervention surveys recorded a wealth of information about households' socioeconomic conditions and income-generating activities, sample size was determined specifically to ensure sufficient power to detect changes in rates of contraceptive usage, which was initially the key outcome of interest.[7] An implication of this is that statistical power was ex ante relatively low for outcomes such as income or wealth indicators, outcomes which are usually characterized by large variability and measurement error. We will return to this point when we discuss the results.

B. *Baseline Summary Statistics*

The randomization was overall successful at producing balance in a broad range of statistics among the four original treatment groups (Desai and Tarozzi 2011, Table 2). Because in this paper we focus on the impacts of microcredit, in Table 1 we show summary statistics calculated separately for communities where microlenders were assigned to start operating ("assigned" to treatment) versus others assigned to receive either FP programs or no intervention ("control"). Overall, the figures show good balance, with differences between arms generally small and significant at standard levels for only one of 35 variables, although the joint null of equality is rejected at standard levels (p-value $= 0.0025$).

The summary statistics document the poor overall socioeconomic status of sample households. Households were large (about five members on average) and most household heads had low levels of schooling. Most study communities were remote, on average more than an hour away from the nearest market or health center. More than a quarter of households used surface water (from rivers, lakes, etc.) as the main source for drinking needs. Food scarcity was also common, with respondents reporting on average more than two months of insufficient food in a typical year.

Agriculture was the main economic activity for almost 90 percent of households. In control areas revenues from crops during the 12 months before the interview were

[7] Sample size was determined to ensure an 80 percent probability of rejecting the null of no effect at the 5 percent significant level, assuming a baseline contraceptive rate of 6 percent (estimated from the 2000 Demographic and Health Survey of Ethiopia), an intra-class correlation of 0.05, and a 12 percentage points difference in contraceptive behavior between any two of the four experimental arms.

Table 1—Baseline Summary Statistics and Tests of Balance

	Control (assigned)		Assigned Treatment − Control	
	Mean (1)	SD (2)	Coefficient (3)	*p*-value (4)
Household composition				
Number of household members	5.22	2.14	−0.046	0.750
Number of adults ($\geq$ 16 years old)	2.43	1.01	0.040	0.522
Number of children ($<$ 16 years old)	2.79	1.78	−0.086	0.490
Male head	0.873	0.333	−0.003	0.857
Head age	40.9	13	−0.556	0.363
Head with no education	0.734	0.442	−0.019	0.697
Access to credit				
Loan from RCA	0.021	0.142	−0.006	0.208
Loan from other MFI	0.005	0.0737	−0.002	0.547
Loan from banks and cooperatives	0.026	0.16	0.000	0.951
Informal loan	0.076	0.264	−0.012	0.592
Any type of loan	0.131	0.337	−0.011	0.623
Any type of loans initiated by a woman	0.017	0.13	0.001	0.882
Amount borrowed from (in 2006 Birr)				
Loans from RCA	11.0	94.8	−3.260	0.377
Other MFI	1.4	26.0	−0.522	0.570
Banks and cooperatives	8.4	79.2	0.199	0.957
Informal loan	14.9	79.2	1.210	0.801
Total	36.6	149.0	−0.488	0.951
Loans initiated by woman	7.3	85.6	−3.790	0.372
Income-generating activities				
Agriculture is main economic activity of household	0.855	0.352	−0.0185	0.369
Total revenue from crop sales last 12 months	203	650	−38.3	0.485
Total expenditure for crops last 12 months	89.9	977	−10.3	0.715
Number of large animals owned	2.84	5.37	−0.374	0.117
Total value of livestock	1,502	2021	−122	0.426
Total revenues from livestock sales last 12 months	160	423	12.8	0.714
Total sales from nonfarm self-employment last 12 months	310	6,804	−147	0.318
Total costs for nonfarm self-employment last 12 months	17.2	144	6.51	0.497
Nonfarm self-employment activities	0.108	0.333	0.017	0.591
Nonfarm self-employment activities managed by women	0.042	0.212	0.011	0.488
Transfers in cash or kind last 12 months	115	443	28.2	0.454
Income from wages last 12 months	174	1,100	−0.709	0.990
Other indicators of socioeconomic status				
Total value of selected assets[a]	36.9	62.7	3.57	0.564
Surface water as main source for drinking	0.264	0.441	0.091	0.093*
Number of months of insufficient food in a typical year	2.4	1.9	−0.2	0.224
Distance to nearest health facility (minutes)	89.7	90.6	11.5	0.507
Distance to nearest market (minutes)	79.0	68.6	4.6	0.661

Notes: Data from baseline (2003) survey. Sample size is $n = 6{,}412$, of which 3,216 assigned to treatment and 3,196 assigned to control. Columns 1 and 2 report statistics for households in PAs where ACSI/OCSSC were not randomly assigned to start lending. Column 3 shows the difference between the mean for households in areas assigned to ACSI/OCSSC and the means in column 1. Column 4 shows *p*-values for the test of equality of means, robust to intra-PA correlation. The number of clusters (PAs) is 133. Asterisks denote statistics significant at the 10 (*), 5 (**), or 1 (***) percent level. The joint null of equal means is rejected at standard levels ($F(34, 99) = 2.10$, *p*-value = 0.0025). When testing the joint null we exclude household size because of collinearity with the variables that describe the demographic household composition. All figures expressing monetary values have been converted into 2006 Birr using region-specific consumer price indexes (CPIs) constructed by the Central Statistical Agency of Ethiopia. In Amhara, the CPI increased from 114.6 in January–May 2003 to 158.1 in March–July 2006 (a 38 percent increase), while in Oromiya the increase was from 122.8 to 156.8 (a 28 percent increase). The PPP exchange rate according to the latest World Bank figures is 2.25 Birr/US$1 (World Bank 2008).

[a] Estimated from the resale value of the following items owned by the household: radio, electric stove, lamps, beds, tables and chairs, bicycles, motorcycles, cars, and trucks.

203 Birr on average (or about US$90 using the PPP exchange rates in World Bank 2008), while total expenditures were 90 Birr. Animal husbandry was also important, both as an income-generating activity and as a form of asset accumulation: in control areas, on average households owned almost three large animals such as cows or oxen, the total value of livestock was 1,500 Birr (about US$670, more than 7 times the average revenues from crop cultivation) and sales of animals amounted to 160 Birr in the previous 12 months. Other sources of income included wage labor (174 Birr per household on average during the previous 12 months), transfers in cash or kind (115 Birr) and sales from nonfarm business activities (310 Birr), although households only owned 0.1 such activities on average.

Baseline information also shows that borrowing was very limited in the area. Only 13 percent of households borrowed in Control areas, and the average amount of outstanding loans (including zeros for nonborrowing households) was less than 40 Birr. Most households borrowed from informal sources, while less than 3 percent had loans from formal institutions such as banks and cooperatives. Microfinance institutions or revolving credit associations were also rare, with about 2 percent of sample households having funds from such sources.

In principle, the very low prevalence of borrowing at baseline may also be consistent with low demand for credit, but several indicators suggest that limited access to credit had negative implications for households' income generating activities and consumption smoothing ability. First, we have seen that on average households experienced more than two months of insufficient food. Indeed, only 27 percent of respondents said that their household always had enough to eat, while 45 percent stated that food was not sufficient for 2–3 months in a typical year and about 1/4 said that food scarcity was a problem for 4 months or longer. Access to credit may have helped households to smooth consumption seasonality. Second, limited access to credit was mentioned among the three most important factors limiting income growth in agriculture and nonfarm business activities by about 20 percent of households involved in such activities (figures not reported in the table).[8] Limited access to credit may also have contributed to the fact that only 11 percent of households had any nonfarm business, although our data do now allow us to test this conjecture.

C. *Deviations from the Experimental Protocol*

We have seen that the results in Table 1 show a good degree of balance in observed characteristics between areas assigned to treatment and control. However, the implementation agencies did not always comply with the experimental protocol. In fact, actual treatment coincided with the randomly assigned treatment in only 104 of the 133 PAs, that is, in 78 percent of cases. Specifically, 8 of the 69 PAs where microcredit was supposed to be introduced did not see it happening, while

[8] Insufficient credit was not listed as a possible option limiting income growth in livestock activities, so we do not have clear information about the role of credit constraints for this activity, although we know that insufficient grazing land was mentioned as the key limiting factor in 83 percent of cases.

ACSI/OCSSC started to operate in 15 of 64 areas where they had not been assigned to do so.[9]

In Appendix Table B1 we show the results of a linear probability model where we regress a dummy equal to one if the household resided in a PA where microlending was actually introduced on a dummy equal to one when the PA was assigned to microlending and a list of observed household characteristics. Although as expected assigned treatment is the strongest predictor of treatment, a number of other coefficients are large and significant at standard levels. In particular, borrowing from informal sources decreases the predicted probability of treatment by 16.6 percentage points (p-value < 0.001), and income from wages or transfers also decreases significantly such probability, although the corresponding slopes are very small. A joint test of significance of all slopes except assigned treatment is rejected at the 5 percent level (p-value $= 0.011$).

The results of the regression do not show an overall unambiguous pattern of purposely selective program placement, for instance toward areas that were richer or poorer, or toward areas with more or less access to formal credit. Despite this, in order to avoid potential bias from endogenous deviations from the assigned program placement, in the rest of the paper we rely on intent-to-treat estimates, see Section II for details. In the online Appendix we also show the results of two-stage least square estimates where we use assignment to treatment of PAs to instrument for actual (potentially endogenous) treatment of study PAs. Random assignment ensures the exogeneity of the instrument, while the relatively limited departures from the experimental protocol rule out concerns related to weak instruments.

D. *Endline Survey and Attrition Concerns*

The endline survey was completed in March–July 2006, approximately three years after the preintervention survey. At this time, a total of 6,263 households were interviewed, of which 3,059 were in Amhara and 3,204 in Oromiya. As explained earlier, the endline sample was drawn independently from baseline, drawing randomly from a new census conducted in the same study villages in 2006. This does not allow us to evaluate attrition in a straightforward way, with potentially important implications for the interpretation of the results. On the one hand, our focus on intent-to-treat estimates (see Section II) still allows us to interpret the results as causal impacts of assigning communities to receive specific interventions, regardless of the actual implementation of such interventions (at the PA level) and regardless of actual take up conditional on implementation (at the household level). On the other hand, the interpretation of the results would change if the interventions had an impact on the composition of the surveyed sample. This may have happened if the interventions led to differential migration into or out of study areas, or to differential survey response rates, with unclear consequences for the results. For instance, if assignment to microcredit reduced out-migration due to improvements in local economic conditions, the estimates would confound impacts on outcomes conditional

[9] In the end, both microfinance and FPP were introduced in 43 PAs, microfinance only was introduced in 33, FPP only in 20, and neither in 37.

on residing in a treatment community with impacts on the probability of residing in such community.

Despite these considerations, our data suggest (albeit indirectly) that differential migration or response rates are unlikely to raise important concerns for the interpretation of the results. As a first check, we use information about the length of time spent by sample households in their village of residence. Only 80 of 6,263 respondents in the postintervention survey (1.3 percent) reported having lived in their village for less than 4 years (recall that the baseline was completed 3 years earlier). Information on the reason for migration is available for only 46 of these households, but in no case was the reported reason directly related to the availability of microcredit, although 23 respondents reported migration was due to "work." Of these 23, 18 moved into areas where microcredit was not introduced, so most of these re-locations were not due to earning opportunities opened up by the increased availability of credit.

Next, in Appendix Table B2 we evaluate whether either actual or assigned treatments were systematically associated with a few household characteristics—such as duration of current residence, the number of adults in the households or the gender or education of the head—that were least likely to have changed directly as a consequence of the interventions. Significant changes over time in these characteristics could signal differential migration or survey response, but we find little evidence of it. The null of equal changes across treatment arms is only rejected (at the 5 percent level) for the number of adults in the household, and even in this case the differences across arms are not large, with the means ranging from 2.4 to 2.7.

Overall, we thus conclude that migration or differential survey responses are unlikely to be key drivers behind our results, although the lack of a true panel of households does not allow us to rule out these concerns conclusively.

II. Estimation Methodology and Results

The lack of a panel of households does not allow us to look at changes in outcomes at the household level, but because the same villages were surveyed before and after the intervention we can control for the presence of time-invariant location-specific fixed effects. Throughout the paper, we define treatment at the level of the unit of randomization, that is, the PA. As we mentioned earlier, in a number of PAs the actual program implementation differed from that scheduled to take place according to the randomized assignment. In the rest of the paper we focus on intent-to-treat (ITT) estimates, by regressing outcomes of interest on PA-specific dummy variables for randomly assigned treatment. For a given outcome y, the equation being estimated is thus the following:

$$(1)\quad y_{pi,t} = \left[\beta_{Post} + \beta_{MF} MF_p + \beta_{FP} FP_p(1 - MF_p) + \beta_{Both} FP_p MF_p\right] Post_t + \alpha_p + u_{pi,t},$$

where $y_{pi,t}$ denotes the outcome for household (or individual) i in PA p and time t (where $t = 0$ denotes baseline and $t = 1$ follow-up), α_p is a PA fixed effect and

$Post_t$, MF_p and FP_p denote binary variables equal to 1 when, respectively, $t = 1$ and when microcredit or FPP were *randomly assigned to be introduced* in PA p. The residual $u_{pi,t}$ is allowed to be correlated within PA, so all standard errors and tests will be robust to intra-PA correlation. The intent-to-treat parameter β_{MF} is thus the main object of interest. Equation (1) also controls for assignment to FPP, either in isolation or in addition to microlending. The coefficient β_{FP} measures the impact of assigning a community to receive FPPs without microlending (that is, when $FP_p = 1$ and $MF_p = 0$), while β_{Both} measures any differential impact of assignment to microlending when FPPs were assigned to be introduced as well (that is, $FP_p = 1$ and $MF_p = 1$). In light of the fact that the FPPs were not effective at changing contraceptive usage (the main outcome initially targeted by the study, see Desai and Tarozzi 2011) our prior was that $\hat{\beta}_{FP}$ and $\hat{\beta}_{Both}$ would be generally small and not statistically significant. This prediction is mostly but not always borne out in the data. Still, because the main objective of the paper is to gauge the impact of microfinance we choose to focus on $\hat{\beta}_{MF}$, although we report the full results in the online Appendix, where we also show the estimates when the presence of family planning services is ignored (that is, when we impose $\beta_{Both} = \beta_{FP} = 0$).

If the experimental design had been followed perfectly (a condition that fails in our empirical context), the ITT would identify the average impact of giving access to microlending in the study area, regardless of actual household borrowing.[10] However, the potentially endogenous deviations from experimental protocol mean that the ITT in (1) only identifies the impact of *assigning* communities to the treatment, regardless not only of actual borrowing but also of actual program implementation in the field. The impact of treatment assignment remains of policy interest, given that any program expanding access to microfinance will likely have to contend with the unwillingness or inability of the MFI to enter certain markets. Throughout the paper we thus focus on the reduced form impact of program assignment.

As an alternative, we could have estimated a model analogous to (1) using two-stage least squares (2SLS), replacing the dummies for assigned treatment with dummies for actual treatment and using the former exogenous variables to instrument for the potentially endogenous latter ones. We do not pursue this strategy here to make our results more easily comparable to the other RCTs in the special issue of this Journal, but the interested reader can find the 2SLS estimates in the online Appendix.[11]

When estimating program impacts on a large number of outcomes, as in our context, the probability of rejecting the null of no impact for at least some of the

[10] Heckman, LaLonde, and Smith (1999, 1903) define ITT as the "mean effect of the offer of treatment."

[11] Because randomly assigned treatment status is an exogenous and strong instrument for treatment, 2SLS will estimate the ITT impact of providing access to microcredit if program impact is homogeneous across areas. However, in the common situation where program impacts are heterogeneous, and under the plausible assumption that assignment to a given treatment (weakly) increases the probability of that treatment in all PAs, 2SLS will only estimate a local average treatment effect (LATE, Imbens and Angrist 1994). The LATE can thus be interpreted as the impact only for "complier" PAs, defined as those that saw the introduction of microlending only because they were assigned to it. Under the assumption of no externalities or general equilibrium effects across PAs, one could also estimate the average impact on households who actually borrowed by dividing the 2SLS estimate by the fraction of borrowers. However, the introduction of microfinance is likely to generate general equilibrium effects that will affect nonborrowers as well (Kaboski and Townsend 2012; Buera, Kaboski, and Shin 2013) so for this parameter assigned treatment, while still a strong instrument for borrowing behavior, would likely fail the exclusion restriction.

outcomes can be very large even when the null is true for all outcomes, if for each outcome one uses standard critical values that are only designed to control the probability of a Type I error for individual tests. We address this issue in two ways. First, for each "family" of relatively homogenous outcomes we construct an index of standardized outcomes in the spirit of Kling, Liebman, and Katz (2007). Each index is constructed as the simple average of the outcomes within the family, standardized using the mean and the standard deviation of the outcome estimated from control areas at endline. Second, in order to take into account the multiplicity of tests when doing inference, for each index we also report a corrected (more conservative) p-value estimated as described in Hochberg (1988). The details of the construction of the indexes as well as of the corrected p-values can be found in Appendix C.

A. *Impact on Borrowing Behavior*

To estimate the impacts on borrowing behavior we rely on household survey data collected before (2003) and after (2006) the intervention that reflect self-reported information on loans at the time of the surveys. In contrast, we have no detailed information provided directly by the MFIs about variables such as total loans disbursed or clients served by PA/year. However, service data collected in all PAs to verify the extent to which program implementation deviated from the experimental protocol indicate that by the end of 2003 ACSI/OCSSC were already granting loans in 63 percent of the PAs where they eventually entered before the follow-up survey, and the proportion grew to 82 percent by the end of 2004. In a large majority of treated communities, program exposure was thus as long as two to three years.

In Table 2, we demonstrate that the intervention led to substantive and statistically significant increases in borrowing. The figures in column 1, panel A, show that assignment to MF increased the fraction of households with any outstanding loans from ACSI/OCSSC by 25 percentage points relative to control areas, where less than 6 percent of households borrowed from these sources.[12] The estimate is precise and the null of no impact is rejected at the 1 percent level. In columns 7 and 8, we show the corresponding estimates when we use assigned treatment status dummies (interacted with *Post*) as instruments for actual treatment status (again interacted with *Post*). As expected, the instruments are strong and the 2SLS estimate is substantively larger, indicating a 36 percentage point increase in borrowing prevalence in treated areas. As we indicated earlier, in the rest of the paper we only focus on the OLS-ITT estimates, while we show the 2SLS estimates in the online Appendix.

We find no evidence of crowding out of other forms of borrowing (columns 2–4): the frequency of loans from NGOs, banks and cooperatives increased by 2–3 percentage points relative to control areas, that of loans from informal sources was barely affected, and neither ITT is significant at standard levels. As a consequence,

[12] The figure of 0.06 represents the fraction of households borrowing from "revolving credit associations" (RCA) at endline in areas assigned to receive neither MF nor FPP. In the section of the postintervention questionnaire where outstanding loans were listed, borrowing from ACSI/OCSSC was coded as borrowing from RCAs, and so we use the two terms as identical, although, especially at baseline, in some cases the loans may have been obtained from sources different from ACSI/OCSSC. In any case, our data indicate that the large majority of loans from RCAs at follow-up were indeed from ACSI/OCSSC.

Table 2—Impacts on Borrowing

	ACSI & OCSSC (1)	NGOs (2)	Banks & coops (3)	Informal sources (4)	All sources (5)	All sources: women (6)	$MF = 1 \times (t = 1)$ 2SLS 1st stage (7)	ACSI & OCSSC 2SLS 2nd stage (8)
Panel A. Credit access	Any loan from							
Microcredit × Post	0.252***	0.026	0.019	−0.006	0.252***	0.081***	0.753***	0.357***
	(0.058)	(0.030)	(0.039)	(0.022)	(0.064)	(0.025)	(0.009)	(0.093)
Observations	12,675	12,675	12,675	12,675	12,675	12,675	12,675	12,675
Endline mean in control	0.0597	0.0138	0.081	0.052	0.223	0.056		0.0597
Panel B. Loan amounts	Loan amounts (in 2006 Birr)							
Microcredit × Post	368***	21	−0.6	3.6	389***	134***	0.753***	458***
	(84)	(29)	(54)	(8.5)	(90)	(32)	(0.009)	(120)
Observations	12,675	12,675	12,675	12,675	12,675	12,675	12,675	12,675
Endline mean in control	61	14	102	18	200	38		61
Panel C. Index of dependent variables				Any loan		Loan amounts		
Microcredit × Post				0.380***		0.408***		
				(0.085)		(0.078)		
Hochberg-corrected p-value				< 0.001***		< 0.001***		
Observations				12,675		12,675		

Notes: Data from 2003 and 2006 surveys. Standard errors (in parentheses) and tests are robust to intra-PA correlation (there are 133 clusters/PA). All regressions also include PA fixed effects. The coefficients in columns 1–6 are OLS estimates of β_{MF} in model (1) in the text. The dependent variables in columns 1–6 of panel A are defined as follows: a dummy for whether the household had an outstanding loan from ACSI/OCSSC (column 1), or from NGOs (column 2), or from a bank or cooperative (column 3), or from informal sources such as money lenders or other individuals (column 4), or from any sources (column 5), or if a woman in the household had a loan from any source (column 6). The dependent variables in columns 1–6 of panel B are the amounts corresponding to the loans defined in the column headers. In panel C we show the results of the estimation of model (1) using an index of dependent variables as outcome. The index is the simple average of the standardized outcomes in columns 1–6 of either panel A ("Any loan") or panel B ("Loan amounts"), for details see Appendix C, where we also describe the calculation of the Hochberg-corrected p-values. The coefficients in columns 7 and 8 are from the 2SLS estimation of model (1) with MF and FP defined in terms of PA-level actual treatment status, with the corresponding PA-level randomly assigned variables (interacted with $Post_t$) used as instruments. In column 7 we show the first-stage coefficient for (actual) MF status interacted with $Post_t$, while column 8 shows the second stage estimates of β_{MF} for the same dependent variable as in column 1. The endline means reported at the bottom of each panel are calculated for areas that were randomly assigned to receive neither microcredit nor family planning: that is, the means are conditional on $MF = FP = 0$, $t = 1$. All statistics are calculated using sampling weights. All figures expressing monetary values are in 2006 Birr. The PPP exchange rate according to the latest World Bank figures is 2.25 Birr/US$1 (World Bank 2008).

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

the impact on loans from the microlenders is virtually identical to the overall impact on the frequency of borrowing (column 5). We also see that, despite the lack of strict guidelines about targeting women for loans, female borrowing saw very substantial impacts: while in control areas less than 6 percent of households had loans initiated by women, the ITT estimates show an 8 percentage points larger prevalence in treatment areas.[13]

[13] Note also that, because the follow-up was conducted in the same villages as the baseline, model (1) can be estimated using village, rather than PA, fixed effects. In the online Appendix Table D1 we show that the two sets of

The estimates in the bottom panel B of Table 2 show that not only the frequency of borrowing increased, but the amounts involved were substantial. The estimated ITT was 368 Birr (standard error 84), that is, more than one-and-a-half times the average revenues from crop sales in control areas at baseline and about one-fourth of the average value of livestock owned (see Table 1). Consistent with the results on borrowing prevalence, the impact on borrowing from ACSI/OCSSC is almost identical to that on total borrowing, with no evidence of important changes in loans from other sources. The mean amount of women-initiated loans increased by 134 Birr relative to control areas, from an average close to 0 at baseline.

When we look at the indexes of dependent variables (panel C) we find, again, strong evidence of increases in borrowing in areas assigned to treatment. Both the index of measures of borrowing prevalence and that of amounts borrowed indicate an average increase of about 0.4 standard deviations relative to control areas, and both are significant at the 1 percent level, even when we use the more conservative Hochberg criterion to take into account multiple testing.

When we look only at households who borrowed from ACSI/OCSSC (results not in table), we estimate that the median loan at follow-up in treated areas was approximately 1,200 Birr (about $500 USD), with only about 10 percent of loans smaller than 700 Birr and about 10 percent larger than 2,000 Birr. To put these figures in perspective, the official poverty line, expressed in total consumption per adult/year in 2006, was close to 1,500 Birr, while the mean amount of total outstanding loans among households who borrowed (from any sources) at baseline in control areas was about 300 Birr (in 2006 units). Our data thus show considerable increases in both the extensive and the intensive margin of borrowing.

As we discussed earlier, in a subset of PAs the microlending operations were accompanied by the introduction of community-based FPPs, conducted by independent organizations. As indicated in model (1), we always control for the presence of such programs, either in isolation or in addition to microcredit. We show the full results in online Appendix Table D1. As expected, in all regressions the coefficients β_{FP} and β_{Both} are not significant at standard levels, although in some cases the point estimates are not small. In online Appendix Table D1 we also show that ignoring the presence of such programs leaves the results substantively unchanged, although the estimated ITTs become smaller for borrowing prevalence and larger for amounts borrowed.

In Appendix Table B3 we show the results of a regression where we analyze what predicts borrowing from microlenders in areas where access to this form of credit was actually introduced. Of course these estimates are not to be interpreted causally because all regressors are likely endogenous, being possibly correlated with sources of unobserved heterogeneity such as entrepreneurship, impatience and risk aversion among the others. Still, we chose regressors that were unlikely to have been affected by the availability of credit, so that at least we can limit concerns about reverse causality from borrowing to the covariates. Recall also that we do not have a panel of households, so we do not have any household-specific data recorded for these

results are almost identical in terms of both point estimates and standard errors, and so in the rest of the paper we will only focus on the PA-fixed effects results.

same households before the intervention. We also included in the regressions PA fixed effects, so that the estimates control for all PA-level characteristics that may enter in an additive, linear way in the prediction. The results show that a number of variables are very strong predictors of borrowing from microcredit. In particular we find evidence that households with low socioeconomic status were less likely to borrow from ACSI/OCSSC: everything else being the same, the presence of a head with no formal schooling reduces the likelihood of borrowing by 7 percentage points (p-value 0.002), while the predicted probability increases by 2 percentage points for every additional hectare of cultivable land ($p < 0.01$) and for every additional sleeping room in the household's dwelling ($p = 0.103$). Households that have resided in the current location for less than 4 years were 14 percentage points less likely to borrow ($p = 0.034$). Overall, this is consistent with ACSI/OCSSC lending preferentially to households more likely to be able to repay the loans and possibly more likely to offer collateral. One additional interesting observation is that households that had a nonfarm business that started at least 4 years before the follow-up (that is, before the intervention) were 6 percentage points more likely to borrow ($p = 0.013$).

Respondents reported that a large majority of loans from ACSI/OCSSC were utilized for productive purposes. Of a total of 1,682 microloans at follow-up, 1,388 (83 percent) were reported as having been initiated to pay for "working capital" or "basic investment." Such categories encompassed production-related items such as wages for hired labor, rents for land and equipment, cost of seeds, fertilizers and pesticides, fees for veterinary services, purchase of animals, land, equipment, etc. In contrast, only 25 loans were initiated to repay other loans, and a total of 144 (9 percent) were used to pay for consumption, schooling, or ceremonies. We also find that most loans were initiated to fund crop cultivation or animal husbandry, with 80 percent of the 1,388 loans used for working capital or investment in these sectors and only 235 (17 percent) used for "trading and services." The remaining 40 loans were used for hard to categorize agricultural and nonagricultural "processing" and for "production."

B. *Impact on Households' Economic Activities*

Next, we turn to the analysis of impacts of ACSI/OCSSC operations on households' economic activities. Both baseline and follow-up surveys included information on sales as well as input purchases for farm and livestock activities and for nonfarm self-employment businesses. No information was collected on family employment or home consumption, so we cannot estimate a measure of profit. We focus then on measures of "net sales," calculated as differences between yearly revenues and input purchases. Expenditures and sales related to these activities were recorded with a 12-month recall period, see Appendix A for details. The one-year reference period likely reduced concerns about seasonality, but it may have exacerbated recall errors that are common in contexts such as ours, where record-keeping is rare. de Mel, McKenzie, and Woodruff (2009) find that in a sample of microenterprises in Sri Lanka reports were similar when using a 12-month recall as compared to the sum of monthly data collected four times during the year, although they also find evidence of misreporting with both methodologies.

TABLE 3—IMPACTS ON SELF-EMPLOYMENT ACTIVITIES: EXPENSES AND REVENUES

	Has nonfarm business (1)	Has female-led nonfarm business (2)	Started business last 3 years (3)	Revenues last 12 months (4)	Investment last 12 months (5)	All expenses last 12 months (6)	Net revenues last 12 months (7)	Index of dependent variables (8)
Panel A. Nonfarm self-employment activities								
Microcredit × Post	−0.006	−0.027	−0.018	573	10	47	526	0.068
	(0.043)	(0.032)	(0.016)	(414)	(9)	(129)	(403)	(0.072)
Hochberg-corrected p-value								0.706
Observations	12,675	12,675	12,675	12,675	12,675	12,675	12,675	12,675
Endline mean in control	0.254	0.146	0.073	438	8	291	146	
	Cash revenues from crops (1)	Expenses for crop cultivation last 12 m (2)	Net revenues from crop last 12 m (3)	Land cultivated last 12 m (hectares) (4)	Value of livestock owned (5)	Value of large animals owned (6)	Livestock sales last 12 m (7)	Index of dependent variables (8)
Panel B. Crop cultivation and animal husbandry								
Microcredit × Post	63	154**	1	0.08	206	198	77**	0.070
	(219)	(75)	(196)	(0.13)	(258)	(243)	(38)	(0.046)
Hochberg-corrected p-value								0.378
Observations	12,675	12,675	12,381	12,348	12,675	12,675	12,675	12,675
Endline mean in control	727	185	698	1.36	27,72	2,372	317	

Notes: Data from 2003 and 2006 surveys. Standard errors (in parentheses) are robust to intra-PA correlation. The rows labeled "Microcredit × Post" display OLS estimates of β_{MF} in model (1) in the text. All regressions use sampling weights. The control means at the bottom of each panel are calculated from endline data from areas where neither microcredit nor FPP were introduced: that is, the means are conditional on $MF = FP = 0, t = 1$. Net revenues from crop cultivations are estimated as the difference between revenues and the fraction of costs imputed for the crop share sold by the household, see Appendix A for details (so net revenues from crops are not the simple difference between columns 1 and 2). All figures expressing monetary values are in 2006 Birr. The PPP exchange rate according to the latest World Bank figures is 2.25 Birr/US$1 (World Bank 2008). In column 5 of panel A, "investment" refers to expenditures in "equipment, machinery, assets" for nonfarm businesses. The figures in column 6 ("all expenses") add to such expenditures the amounts spent for hired labor, raw material, transport, storage and "other items." The dependent variables in column 8 are indexes constructed as the simple average of the standardized outcomes in columns 1–7 of the same panel (all entered with their own sign), see Appendix C, where we also describe the calculation of the Hochberg-corrected p-values.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

In panel A of Table 3, we show that none of the impacts on nonfarm business activities was significantly different from zero at standard levels, although some of the estimates are large in magnitude. A key result is that we find no evidence that the substantial increase in borrowing documented earlier led to nonfarm business creation in areas assigned to treatment. The ITT is actually negative for the fraction of households with a nonfarm business, or with a female-led nonfarm business, or for the fraction that started a business in the previous three years (columns 1–3). Although the confidence intervals for these variables always include positive values, the estimates are sufficiently precise to reject the null of an ITT larger than 1.4 percentage points for the probability of nonfarm business creation during a three-year period that encompassed approximately the time interval between baseline and endline surveys (the confidence interval is −0.049 to 0.013).

When we look at nonfarm business-related monetary outcomes we find that, while again not significant at standard levels, the impacts are large in magnitude, see columns 4 to 7. The ITT for "net revenues" (the difference between sales and monetary costs in the previous 12 months) is 526 Birr with a standard error of 403, so that the upper bound of the 95 percent confidence interval (1,316 Birr) is almost as large as the average value of livestock in control areas at baseline, while the lower bound (−264) is almost one-fifth of such value in magnitude.

In panel B we show estimated ITT impacts on crop cultivation and livestock activities during the 12 months before the survey. The ITT for revenues from the sale of crops is 63 Birr, or 9 percent of the endline mean in control areas, while the standard error is more than three times as large, so the 95 percent confidence interval is very wide (−366 to 492 Birr). In contrast, the ITT for the total monetary costs incurred for crop cultivation is significant at the 5 percent level, and the point estimate is very large (154 Birr), more than 80 percent of the average for this value in control areas at follow-up. However, this result is difficult to interpret because we do not have information on total quantities produced or consumed, so such costs refer to inputs used to produce crops destined to both the market and self-consumption. In column 3, we estimate the ITT for "net sales" estimated as the difference between total cash sales and total crop-related expenditures multiplied by the share sold, see Appendix A for details. Such impact is close to 0, although the 95 percent confidence interval is very wide (about $\pm$ 400 Birr). The ITT for the amount of land cultivated is similarly close to zero and imprecisely estimated, although we can reject the null that assignment to microcredit led to increases larger than one-quarter of the average area cultivated in control areas at follow-up.

Columns 5 to 7 of panel B show that areas assigned to microloans saw relatively larger increases in the stock of animals owned as well as in the value of their sales. The estimated impact on the value of livestock owned is large (206 Birr), but the standard errors are even larger so the 95 percent confidence interval includes negative ITTs as large as 20 percent of the mean livestock holdings at baseline in untreated areas, and positive values almost half as large as that. The ITT for the value of sales is moderately large (77 Birr) and is significant at the 5 percent level, with the lower bound of the 95 percent confidence interval positive but close to 0 and the upper bound equal to about 10 percent of baseline holdings in control areas.

When we look at the indexes of standardized dependent variables (column 8), the impact is small and close to 0.07 standard deviations for both nonfarm self-employment and crop and livestock-related activities. Both estimates are not significant at standard levels, and not surprisingly the null of no impact is also not rejected using the more conservative Hochberg criterion.

In Table 4 we analyze the ITTs on aggregate measures of income from self-employment activities as well as from other income sources. Overall, areas assigned to microcredit saw substantively larger increases in both revenues and expenses for self-employment activities, although the estimates are noisy and the null of no impact cannot be rejected at standard levels. The point estimates of the ITT is 712 Birr, almost exactly half of mean revenues in control areas. The confidence interval is very wide and ranges from −227 to 1,651 Birr. The ITTs for total costs and "net revenues" for self-employment activities (the latter calculated as the difference between total revenues and total costs) are also large but not significant. In particular we estimate an average impact on net revenues of 513 Birr (or 68 percent

Table 4—Impacts on Income Indicators

	Household income last 12 months						
	Totals from self-employment activities						
	Revenues (1)	Costs (2)	Net revenues (3)	Wages (4)	Transfers (5)	Other income (6)	Index of dependent variables (7)
Treated village	712	199	513	49	−11	28	0.083*
	(479)	(156)	(431)	(84)	(17)	(19)	(0.048)
Hochberg-corrected p-value							0.352
Observations	12,675	12,675	12,675	12,675	12,675	12,675	12,675
Endline means in control	1,482	727	755	294	23	31	

Notes: Data from 2003 and 2006 surveys. Standard errors (in parentheses) are robust to intra-PA correlation. The rows labeled "Microcredit × Post" display OLS estimates of β_{MF} in model (1) in the text. All regressions use sampling weights. The control means at the bottom of each panel are calculated from endline data from areas where neither microcredit nor FPP were introduced: that is, the means are conditional on $MF = FP = 0, t = 1$. Revenues (column 1) is the sum of revenues from sales of crop or livestock and products from nonfarm businesses. Costs (column 2) is the sum of all costs for crop cultivation or nonfarm businesses plus costs for animals purchased. Net revenues (column 3) is the difference between the two. See Appendix A for additional details. All figures expressing monetary values are in 2006 Birr. The PPP exchange rate according to the latest World Bank figures is 2.25 Birr/US$1 (World Bank 2008). The dependent variable in column 7 is an index constructed as the simple average of the standardized outcomes in columns 1–6 (all entered with their own sign), for details see Appendix C, where we also describe the calculation of the Hochberg-corrected p-value.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

of the average in control areas) but once again the 95 percent confidence interval is wide (−332 to 1,358 Birr). The magnitude of the ITT for total production costs (199 Birr) is 54 percent of the ITT for amounts borrowed from ACSI/OCSSC. Although the estimate is noisy (the standard error is 156 Birr), this is consistent with loans being initiated in large part for productive purposes, as indicated by respondents.

The impact on wages was also positive (49 Birr) but not significant. Given the standard error (84) we can reject negative impacts larger than 116 Birr or positive impacts larger than 214 Birr. Impacts on transfers from any source or "other income" are also not significant although for these variables the point estimates and the standard errors are small enough that substantive impacts can be ruled out. That transfers were not affected is interesting because it suggests that the large increase in borrowing documented in Table 2 did not come at the expense of reduced transfers from any existing transfer network that predated the program. This finding is also consistent with the earlier result that the introduction of microfinance did not appear to crowd out existing borrowing from informal sources.

The composite index of standardized outcomes show a 0.08 standard deviation increase (column 8). This is significant at the 10 percent level, but when we use the more conservative Hochberg criterion to account for multiple testing the null of no impact cannot be rejected.

Because we do not have access to panel data our ability to evaluate impacts heterogeneous by preintervention characteristics is limited. Despite this, we can use quantile regression to study whether different parts of outcome distributions were affected differently. This is potentially important because other evaluations have found that impacts on income-generating activities were concentrated in the upper tail of the distribution, see Banerjee et al. (2014) or Angelucci, Karlan, and Zinman (2014). We focus on outcomes for which the OLS regressions showed substantive (if not statistically significant) impacts. For each outcome, we estimate quantile regressions at 9 equally spaced quantiles from 0.1 to 0.9, using the same right-hand side variables as in model (1), but replacing the PA fixed effects with a constant and three arm-specific dummies.[14] We show the results in the 8 graphs in Figure 2, together with 90 percent confidence bands estimated using 250 block-bootstrap replications using PAs as blocks.

When we look at nonfarm business net sales (panel A), we find that the low prevalence of this kind of business causes most ITTs to be close to zero. The only exception is for the ninth decile, where the impact is large and negative (close to -200) but not significant. It may be recalled that the OLS regressions in Table 3 showed large and positive (although not significant) average impacts. It turns out that the latter were completely driven by a handful of outliers: if we exclude 8 observations larger than 50,000 Birr, the average impact from estimating model (1) with OLS becomes negative as well (-76, standard error 96, hence not significant). Overall, this confirms that ACSI/OCSSC did not substantively help the creation of the growth of nonfarm businesses in study areas.

When we look at crop cultivation (panels B to D), we find that impacts on crop sales (either total or net of costs) were relatively small and not significant throughout the distribution. The large positive average ITT estimated with OLS for crop-associated costs was driven by increases in the part of the distribution to the right of the median: the quantile effects are all positive and increasing above the fourth decile, and they are all significant above the median at the 10 percent level (recall that we always display 90 percent confidence bands). The impact at the ninetieth percentile is close to 300 Birr, although the 90 percent bands range from close to zero to above 500 Birr.

In panels E and F we look at quantile regressions for the value of livestock owned and the value of sales. The OLS regressions in Table 3 showed that both ITTs were positive and relatively large, although only the latter was significant, at the 5 percent level. The quantile regressions for the value of livestock show positive impacts for all deciles but the first, with larger impacts for higher deciles. However, the estimates for the highest deciles are also noisier, so the null of no impact is only rejected (at the 10 percent level) for the deciles from the fourth to the seventh. In contrast, impacts on livestock sales are more clearly concentrated in the right tail of the distribution, also because the fraction of households with any sales is sufficiently low that impacts on deciles up to the median remain equal to zero. The ITTs are in a 150–200 Birr range and not significant for deciles 6 and 7 and between 200 and 300 Birr for the eighth and ninth decile of the distribution, although for both these latter deciles the lower bound of the 90 percent confidence interval is close to 0.

[14] The inclusion of the many PA-specific dummies generates problems of convergence in the solving algorithm for the quantile regression in some cases, so we focus on the simpler model, which anyway estimates consistently the ITT of program assignment.

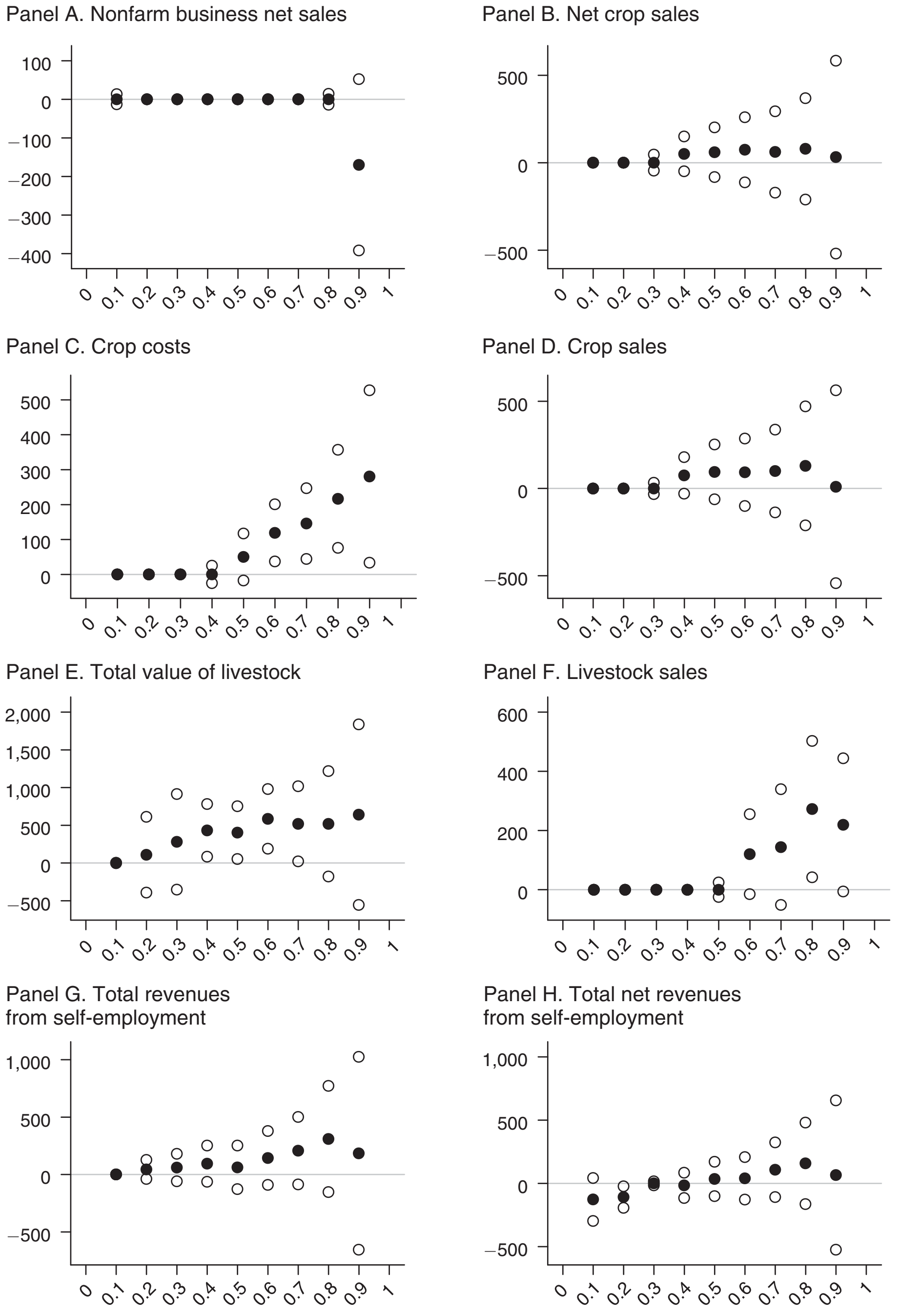

Figure 2. Results of Quantile Regressions

Notes: Each graph shows point estimates (black circles) and 90 percent confidence bands (white circles) for the results of 10 quantile regressions for the outcome indicated at the bottom left of each figure. In each quantile regression there are eight right-hand side variables, that is, the constant and the three arm-specific dummies *MF*, *FP* and *Both*, interacted or not with the endline dummy *Post*. All estimates use sampling weights and the standard errors are estimated using 250 block-bootstrap replications using PAs as blocks. The quantile is indicated along the horizontal axis, while all estimates are measured in 2006 Birr. The PPP exchange rate according to the latest World Bank figures is 2.25 Birr/US$1 (World Bank 2008).

Source: Data from baseline (2003) and endline (2006) surveys

Finally, in panels G and H we look at aggregate measures of revenues (gross or net) from all self-employment activities. The estimates are almost all positive but noisily estimated, with overall evidence of larger impacts above the median. To sum up, the quantile regressions suggest that areas assigned to the microcredit experimental arm saw overall increases in earnings from self-employment activities, which, however, mostly affected the right tail of the distributions. It should also be kept in mind that, as for the case of the ITTs estimated with OLS, the impacts on the quantiles are also imprecisely estimated, so that the confidence bands are wide and the null of no impact cannot be rejected in a large majority of cases.

Impacts on Labor Supply.—Next, in Table 5 we show the estimated ITTs for labor supply, separately for adults of age 16–75 (in the top panel A) and for teenagers (13–19, bottom panel B). For each individual of age 10 and above, the survey recorded the two most important activities the individual was involved in during the previous 12 months. For each activity, records were then taken about the number of weeks spent in such activity, the number of days usually spent per week as well as the number of hours spent per day. Hours were counted as "work" if they were related to one of the following activities: crop cultivation, care of livestock, fishing, mining, manufacture and processing, retail and wholesale trade, finance, public administration, education, health, and social services or other services. Hours spent in school or in domestic work were listed separately.

When we look at adults, none of the estimates is significant at standard levels and all are small. In areas assigned to microcredit, the ITT is 1.1 hours per week, almost completely explained by an increase in time spent in self-employment activities. The standard errors are not small, however, so that substantively important impacts cannot be ruled out and the 95 percent confidence interval for total hours is -1.7 to 3.9 hours per week. The ITT for women's time spent in self-employment activities is -1.1 (standard error 1.04) suggesting that access to microcredit did not lead to substantive changes in women's time spent in economic activities.[15]

Moving now to teenagers, we find no evidence of substantive changes in the total number of hours spent working, with a point estimate of -0.7. However, the estimated impacts on hours in self-employment activities is negative and large (-1.7) relative to the endline mean in control areas (12.4), and for girls the magnitude is even larger (-2.6) and significant at the 10 percent level. For girls, the lower bound of the 95 percent confidence interval (-5.6) is almost as large as the endline average in control areas (6.9). Time spent by teenagers in activities outside the household was on average limited (2.2 hours per week at endline in control areas), but the ITT is relatively large (1 hour per week) and the null of no impact is rejected at the 10 percent level.

The composite index of standardized outcomes, in panel C, show an average impact very close to zero and not significant at standard levels, even when we do not adopt the more conservative Hochberg criterion.

[15] The low numbers of hours worked, on average, at endline in control areas are unconditional means that also include individuals who did not work, most of whom were women. If we estimate this statistic using only individuals for whom work was the primary activity, the averages become 38 hours/week for men and 25 for women.

Table 5—Impacts on Labor Supply

	All adults all activities (1)	All adults self employment (2)	All adults outside employment (3)	Women self employment (4)	Women outside employment (5)
Panel A. Hours work/week: members 16–75 years old					
Microcredit × Post	1.1	0.9	0.1	−1.1	1.0
	(1.43)	(1.17)	(0.68)	(1.04)	(1.03)
Observations	31,769	31,769	31,769	16,051	16,051
Endline means in control	22.9	18.1	4.7	8.8	3.9
	All teens all activities (1)	All teens self employment (2)	All teens outside employment (3)	Girls self employment (4)	Girls outside employment (5)
Panel B. Hours work/week: teens 13–19 years old					
Microcredit × Post	−0.7	−1.7	1.0*	−2.6*	0.5
	(1.42)	(1.53)	(0.54)	(1.52)	(0.66)
Observations	10,537	10,537	10,537	5,372	5,372
Endline means in control	14.6	12.4	2.2	6.9	2.3
Panel C. Index of dependent variables					
Microcredit × Post	0.018				
	(0.013)				
Hochberg-corrected *p*-value	0.360				
Observations	12,675				

Notes: Data from 2003 and 2006 surveys. Standard errors (in parentheses) are robust to intra-PA correlation. The rows labeled "Microcredit × Post" display OLS estimates of β_{MF} in model (1) in the text. All regressions use sampling weights. The control means at the bottom of each panel are calculated from endline data from areas where neither microcredit nor FPP were introduced: that is, the means are conditional on $MF = FP = 0, t = 1$. Hours of work are estimated from recall data about time allocation in the previous 12 months; see Appendix A for details. The dependent variable in panel C is an index constructed as the simple average of the ten standardized outcomes in panels A (entered with their own sign) and panel B (each multiplied by -1), for details see Appendix C, where we also describe the calculation of the Hochberg-corrected *p*-values.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

C. *Impacts on Child Schooling and Other Socioeconomic Indicators*

Next, in panel A of Table 7 we turn to the analysis of child schooling. Ex ante, it is not clear that improvements in such indicators should have been expected, even if our results had shown (as they did not) clear evidence of improvements in income-generating activities.[16] In fact, while income effects and reductions in credit constraints will ceteris paribus typically raise child schooling, the opposite effect may arise if access to credit increases sufficiently the returns to child labor (Wydick

[16] Note that Table 6 is missing to keep table numbering consistent among the different microcredit evaluations in this Journal. In our study this table is skipped because we do not have any information on household consumption.

Table 7—Impacts on Child Schooling and Other Socioeconomic Indicators

	Fraction of children 6–15 in school (1)	Average hours/week worked by children 10–15: Self employment (2)	Average hours/week worked by children 10–15: Outside activities (3)	Fraction of 10–15 girls for whom housework is primary activity (4)	Fraction of 16–20 in school (5)
Panel A. School attendance and time allocation of children					
Microcredit × Post	0.025	−0.6	−0.014	0.047	0.009
	(0.046)	(1.71)	(0.59)	(0.038)	(0.034)
Observations	22,071	11,774	11,774	5,924	7,234
Endline mean in control	0.554	12.7	2.2	0.138	0.455
	Empowerment: Fraction of decisions with woman's involvement: All (1)	Empowerment: Fraction of decisions with woman's involvement: Economic (2)	Value of selected assets (3)	Someone seriously ill last 3 years (4)	Number of months of food insecurity (5)
Panel B. Other indicators					
Microcredit × Post	−0.043	−0.038	−5	−0.015	0.52**
	(0.030)	(0.032)	(13)	(0.034)	(0.26)
Observations	10,500	10,497	12,675	6,263	12,675
Endline mean in control	0.814	0.784	73	0.47	1.29
Panel C. Index of dependent variables					
Microcredit × Post	−0.071**				
	(0.034)				
Hochberg-corrected p-value	0.21				
Observations	12,675				

Notes: Data from 2003 and 2006 surveys. Standard errors (in parentheses) are robust to intra-PA correlation. The rows labeled "Microcredit × Post" display OLS estimates of β_{MF} in model (1) in the text, except in column 4, panel B, because the dependent variable was measured only at follow-up (in this case we exclude the PA-specific fixed effects and the *Post* dummy and we include a region fixed effect). All regressions use sampling weights. The unit of observation is a child in all regressions of panel A, a woman in columns 1 and 2 of panel B, and a household in columns 3 to 5 of panel B. The dependent variable in panel C is an index constructed as the simple average of the five standardized outcomes in panels A and the outcomes in columns 1–3 and 5 in panel B (illness episodes are not included in the index because the outcome was only measured at endline). The outcomes in columns 2–4 of panel A and in column 5 of panel B are multiplied by -1 before the inclusion in the index. For details see Appendix C, where we also describe the calculation of the Hochberg-corrected p-values. The control means at the bottom of each panel are calculated from endline data from areas where neither microcredit nor FPP were introduced: that is, the means are conditional on $MF = FP = 0, t = 1$.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

1999). For instance, in the RCT described in Augsburg et al. (2012) the authors found that increased borrowing from microfinance was associated with an increase in labor supply and a decrease in schooling among teenagers in poorer families.

In Ethiopia, public primary and secondary schools are nominally free, although associated costs such as textbooks and uniforms must be born by the families, see Appendix A for additional details. In rural areas it is common for children to start school late. At baseline, only 15 percent of 7-year old children were attending school, while the proportion increased monotonically to 54 percent at age 14 and declined afterward. Such staggered start of school leads to the lack of clear jumps in enrollment at specific ages, and to a large variation in the age of children enrolled in the same grade. We thus analyze separately impacts for children 6 to 15 and for individuals 16–20. For each individual in the household roster, enumerators recorded years of school completion as well as school attendance at the time of the interview. Attendance was measured in a binary way, simply asking about whether the individual was "currently attending school," while no information was collected on learning outcomes such as test scores. For individuals ten or older, the questionnaire also included separate questions about time allocation, so we will also look at labor supply and domestic work among children 10–15.

In column 1, we see that the estimated program impact on schooling attendance of children 6–15 is positive but very small (0.025) and not significant. This compares to a very large increase in schooling attendance that took place in this age group in untreated areas, where the indicators increased from 37 to 55 percent.[17] The standard error is relatively large, however, so that a 95 percent confidence interval for the ITT ranges from a 6.5 percentage point decline to an 11.5 percentage point increase in attendance. When we look at time allocation among the subset of these children for whom such data were collected (that is, children of age 10–15), we find similarly small and not significant impacts in the number of hours worked, regardless of whether we look at outside or self-employment activities. Among girls in the same age range the proportion for whom domestic chores was the primary activity increased by 5 percentage points relative to control areas. This is a large figure relative to the level at endline in control areas (14 percent), but it is estimated imprecisely and is thus not significant. In column 5, we also show that we do not find any evidence of changes in schooling attendance among older cohorts (16 to 20).

Next, we look at indicators of women's empowerment. When a woman of age 15–49 was present in a male-headed household (about 90 percent of households were, see Table 1), surveyors asked the woman about who in the household was involved in decision making related to 20 different topics, ranging from children's health and education, to contraceptive use, savings and the woman's involvement in the labor market, see Appendix A for details. We thus construct a measure of empowerment as the fraction of decisions the woman stated she was involved in, and a separate measure calculated taking into account only seven domains—such as savings decisions or labor market participation—that we categorized as being more clearly "economic" in nature. Interviewers were instructed to complete the corresponding section of the questionnaire without the spouse being present during the interview, to enhance the truthfulness of the responses.

At follow-up, both indicators show that in control areas women were on average involved in about 80 percent of decisions, suggesting a high degree of participation in the study areas. The estimates in columns 1 and 2 of panel B show no evidence of impacts in either indicator. Both estimates are relatively small and actually negative

[17] Such overall large increase in school attendance was broadly consistent with nationwide trends. For instance, data from the Ethiopia Demographic and Health Surveys show that between 2000 and 2005 the percentage of 10–14 year olds with some primary education increased from 37 to 55 percent among girls and from 47 to 58 percent among boys, see Macro International Inc. (2007).

(−0.043 and −0.038, respectively) although neither is statistically significant at standard levels. For both indicators the 95 percent confidence intervals are wide, although we can reject the null of positive ITTs larger than 3 percentage points. The lack of a beneficial program for these indicators is perhaps not surprising given that, unlike what is often observed in many microfinance institutions worldwide, ACSI/OCSSC did not lend exclusively or predominantly to women.

In column 3 of panel B, we also find no significant impact on the resale value of a list of assets owned by households, including radios, electric stoves, kerosene or pressure lamps, beds, tables, chairs, bicycles, motorcycles/scooters, cars and trucks. The ITT is negative but close to zero (−5 Birr), and the standard error (13) sufficiently small that we can reject the null of economically important impacts on this variable. However, the list of assets is clearly nonexhaustive, so these estimates are only a very coarse approximation of the impacts on household wealth.

In column 4, we see that microcredit had barely any impact on the fraction of households where at least one member was "seriously ill" during the previous three years (that is, in the period of time between baseline and follow-up surveys). This is of course a very coarse health indicator, and in addition it was only measured at follow-up, so in this case we estimate the ITT with a version of model (1) in levels.[18]

Finally, in column 6 of panel B, we see that microcredit was associated with a relative *increase* in the number of months of food insecurity. The estimate is large (0.5 months) and significant at the 5 percent level. Given the coarse way this variable was measured it is possible that this result is spurious, also because the results on economic activities, while mostly not significant, suggested if anything some improvements in areas with increased access to credit.

When we look at the composite index of standardized outcomes, in panel C, we find an average *negative* impact of 7 percent of a standard deviation, and this is significant at the 5 percent level. However, when we use the more conservative Hochberg criterion to take into account the multiplicity of tests the null of no impact cannot be rejected.

III. Discussion and Conclusions

In this paper we have evaluated the impact on several socioeconomic outcomes of the introduction of microfinance in rural areas of Amhara and Oromiya (Ethiopia) in the context of a clustered randomized field trial carried out between 2003 and 2006. Our results should be a useful addition to a still limited number of RCTs that evaluate the impact of *introducing* microloans in communities that previously had no access to it. Our empirical framework is perhaps closer to that in Crépon et al. (2011), who describe an RCT conducted in Morocco: the study area was rural and with little access to credit at baseline, the interest rate charged by the microlenders was relatively low, loans were granted to both men and women and used mostly to

[18]The estimates are similarly negative but close to zero and not significant if we look at the probability of a child under six being seriously ill in the previous three years, or at health expenditures for serious illness during the same period (results not reported).

fund crop production and livestock activities. In addition, in both of these studies (and unlike, for instance, in Banerjee et al. 2014) there was very little activity of other MFIs before or during the study, so that the estimates can be interpreted as impacts of "first generation" microcredit, in areas where access to this form of credit was novel and largely limited to borrowing from the partner MFIs.

ACSI and OCSSC, the two MFIs involved in our evaluation, did not always comply with the experimental protocol so that actual and randomly assigned treatment coincide in only 78 percent of the "peasant associations" (PAs) included in the study. Throughout the paper we thus focused on intent-to-treat estimates, interpreted as impacts of residing in a PA randomly assigned to see the introduction of microlending. Three years after the start of the trial we estimate that in areas where the two MFIs were assigned to start operating, the fraction of households with loans was 25 percentage points higher than in areas assigned as controls. These were very large increases, especially relative to the preintervention borrowing rates of 13 percent throughout the study area. The average loan size among borrowers was large, close to 80 percent of the poverty line in Ethiopia in terms of adult consumption per year. We also do not find evidence that the microloans supplanted preexisting sources of credit.

The magnitude of the increase in borrowing is large also when compared to other recent RCTs that evaluated impacts of microfinance in other locations. The intervention evaluated in Crépon et al. (2011) increased access to credit by 13 percentage points relative to control area; in Hyderabad (India) Banerjee et al. (2014) found that after about 1.5 years program areas saw 9 percentage points more borrowing from MFIs but also 5 percentage points less borrowing from informal sources; credit usage increased by 12 percentage points in a Mexico-based RCT (Angelucci, Karlan, and Zinman 2014) and by 24 percentage points in Mongolia (Attanasio et al. 2011), although in this latter case the study population was composed of women who had expressed an interest in borrowing. Our respondents reported that a large majority of loans were used to fund production activities although, unlike in the most common narrative of microfinance, investment in nonfarm small businesses played only a minor role, with most funds reported as being invested for crop cultivation and livestock-related activities.

Despite the remarkable increase in borrowing, we do not find clear evidence of widespread improvement in socioeconomic indicators in treated areas, although most estimates are imprecise and most confidence intervals are so large that both substantial improvements and large declines in a number of indicators cannot be rejected by the data. Of a total of 40 outcomes, including input and output measures of income-generating activities, labor supply, child school attendance, indicators of women's empowerment and (coarse) indicators of health and food adequacy, only five are significant at standard levels (and none is at the 1 percent level), and even these do not point univocally to a clear improvement in household welfare. The null of no impact can only be rejected for expenses for crop cultivation (at the 5 percent level, with the point estimate indicating an increase), livestock sales (10 percent, increase), hours of work for teens' activities outside of the households (10 percent, increase), girls' hours of work in self-employed activities (10 percent, decrease) and months of food insecurity (5 percent, increase). But given the very large number of outcomes analyzed, mere chance may have produced such rejections of the null hypothesis of no impacts even if the null had been true for all of them. Indeed, when we use more conservative, Bonferroni-type tests for the indexes of dependent variables constructed for each "family of outcomes," following the procedure described in Hochberg (1988), the null of no impact can only be rejected for borrowing behavior.

Note that this latter result also shows that the lack of significant changes in socioeconomic outcomes was not a mere result of low demand for the loans. On the other hand, our results cannot distinguish between a scenario where the loans were not adapted to the needs of rural households versus an alternative one where the returns to investment were anyway low in these areas, although the high take up rates suggest that a large fraction of households valued the loans sufficiently to justify borrowing.

When we look at income-generating self-employment activities, we find that increased access to loans was not associated with more nonfarm business creation, while we find some evidence of impacts on the scale of farm activities, although the estimates are imprecise. The intent-to-treat for total revenues from self-employment activities is very large (712 Birr, about half of the poverty line expressed as total consumption per adult/year) but the 95 percent confidence interval is large enough to include both a 110 percent increase and a 15 percent decrease relative to control areas. Similarly, the confidence interval for the impact on net revenues from the same activities (513 Birr) includes both an 80 percent increase and a 40 percent decrease relative to control areas. Results from quantile regressions show that these impacts were driven by changes in the distributions above the median, although these estimates are equally imprecise. The number of hours worked by adults remained similar in treatment and control areas, and we also do not find impacts on time spent by teenagers in self-employment. Changes in schooling attendance were also generally similar across experimental arms for both children 6–15 and for older 16–20 year-olds, although the corresponding confidence intervals once again include substantively large positive and negative figures. As in Crépon et al. (2011) and Banerjee et al. (2014), we find that increased access to loans was not associated with significant improvements in indicators of women's empowerment (our point estimates actually indicate a small but not statistically significant decline), but it should be recalled that in our study (as in Crépon et al. 2011) microlenders did not target exclusively women borrowers.

Finally, one of the few statistically significant results indicates a one-half month increase in the number of months of food insecurity in treatment relative to control areas (where the average was 1.3 months per year). Unfortunately this very coarse indicator was the only proxy of consumption in our data, so we cannot do much to probe whether this finding (which appears to be at odds with the noisy but generally positive estimates for income) was spurious. More generally, the lack of consumption data is a clear drawback in our analysis because, in this respect, we cannot compare our findings with those of other evaluations that have found that access to microcredit, while leaving aggregate and nondurable consumption largely unchanged, increased consumption of durables while decreasing expenditures in "temptation goods" such as cigarettes or alcoholic beverages.

One relative strength of our study is that the time interval between preintervention and postintervention surveys was relatively long, approximately three years. This is potentially important because the literature has highlighted that lumpy investments may actually decrease certain welfare indicators in the short term, before investments have paid off (see for instance Banerjee et al. 2014 or Fulford 2013). On the other hand, and unlike Banerjee et al. (2014), we do not have data from the *interim* period between baseline and endline surveys, so we cannot gauge to what extent the immediate impacts differed from those observed at endline.

An additional shortcoming of our study is that, although households from the same villages were surveyed before and after treatment, it was not the same households that were surveyed. We thus have a panel of villages but not of households. Having baseline data is still useful because it allows us to gauge to what extent observed household characteristics were balanced across different experimental arms at baseline, and also allows us to control in the estimation for any time-invariant community-level confounder. However, the lack of baseline data limits severely our ability to estimate heterogeneous treatment effects. This is potentially very important because the existence of heterogeneous impacts is a common theme among RCTs evaluating other microfinance programs. We are also unable to study attrition explicitly, although our data suggest overall that the composition of the population among experimental arms was not endogenously affected by the program.

A final but crucial cautionary note is that the failure to identify statistically significant impacts on key outcomes such as net revenues or livestock ownership may also have been the result of measurement error or insufficient statistical power. Recall that the data used in this paper come from a randomized controlled trial for which the primary purpose was the evaluation of FPPs and microloans on contraceptive choices. The power calculations were thus conducted in relation to such fertility-related outcomes. Revenues and costs are notoriously difficult to measure in household surveys (de Mel, McKenzie, and Woodruff 2009) and measurement error of a nonbinary dependent variable, while not causing estimation bias under certain conditions, will increase the standard errors of the estimates. The figures in Table 1 clearly show that standard deviations were large for outcomes expressed in monetary terms. For instance, mean revenues from nonfarm business in control areas at baseline was 310 Birr with a standard deviation about 20 times as large, while the mean value of livestock was 1,502 Birr, with a standard deviation of 2,021 and a very high intra-PA correlation of about 0.20. For this latter outcome, the point estimate for the ITT in Table 3 shows impacts of about 200 Birr, which correspond to a small effect size of about 0.1. Under such a scenario, and taking into account a total of 60 households in each of 60 clusters, the probability of rejecting the null of no impact would be only 21 percent using a 10 percent significance level.[19]

While these important caveats need to be kept in mind, our results are overall consistent with the broad framework described in the survey by Banerjee (2013), with increased access to microfinance associated with some improvements in living standards of beneficiary communities, but without compelling evidence of a true

[19] We calculated power using the Optimal Design software, version 1.56.

transformative power of microfinance. Our study thus provides additional power to the cautionary note sounded early on by Morduch (1999) in relation to the potential of microfinance as a development tool against poverty.

Appendix A: Detailed Description of Outcomes

Outstanding Loans: A household is labeled as having an outstanding loan if any household member owes money or goods to anyone at the time of the interview. Women's borrowing is identified by specific questions about the individual who contracted the loan.

Revenues and Costs from Nonfarm Self-employment: The respondent was directed to include under this label "nonagricultural enterprise, which produces goods or services (for example, artisan, metalworking, tailoring, repair work; also include processing and selling your outputs from your own crops if done regularly)," shops or trading business. Respondents were also asked to report for how many days/weeks/months the business operated in the previous 12 months, and the total revenues from sales per unit of time. Total monetary costs incurred during the previous 12 months were also recorded separately for hired labor, raw materials, equipment/machinery, transport/packing/storage and other items. As an example, suppose that a business operated for 3 months, with weekly earnings of 100 Birr/week, and with total costs of 500 Birr. Then we estimate "net revenues" to be equal to 700 Birr ($= 100 \times 4 \times 3 - 500$). Separate information was collected for each existing business separately. The survey did not collect information on family labor or self-consumption of goods produced by the business.

Animal Husbandry: The value of animals owned is derived from questions about the expected revenues from their hypothetical sale at the time of the interview. The value of sales (in Birr) is the total revenue from actual sales of animals over the previous year. Separate information was collected for each animal type separately (types included cows, oxen, calves, bulls, camels, horses, donkeys, mules, sheep, goats, chickens, and "others"). No information was collected about costs for hired labor, veterinarian services, feed, etc. so for this economic activity we cannot calculate a value for "net sales."

Revenues and Costs from Crop Cultivation: Information was collected separately for each crop type (the principal crops were wheat, barley, teff, maize, sorghum/millet, beans and—in Oromiya—coffee). For each crop, the questionnaire recorded the total revenues from sales (in Birr) over the last 12 months, the share of the total crop sold, and the total amount of expenses incurred for cultivation and sales. We calculate net revenues from sales as the difference between revenues and the corresponding imputed costs, estimated as total costs multiplied by the fraction of the crop sold. So, for instance, if a household incurred a total cost of 500 Birr for cultivation and sold 50 percent of the quantity produced for a total of 300 Birr, net crop sales are calculated as $300 - 0.50 \times 500 = 50$ Birr.

Other Sources of Income: Income from wages is reported as total earnings (in Birr) in monetary or in-kind terms, for work conducted for someone else over the previous 12 months. Income from transfers include transfers in cash or kind received from relatives or friends. Other income includes any inflow in cash or kind from pensions, interests, rents, gambling, inheritances, etc.

Hours of Work: For each individual of age ten and above, the survey recorded the two most important activities the individual was involved in during the past 12 months. For each activity, records were then taken about the number of weeks spent in such activity, the number of days usually spent per week as well as the number of hours spent per day. Hours were counted as "work" if they were related to one of the following activities: crop cultivation, care of livestock, fishing, mining, manufacture and processing, retail and wholesale trade, finance, public administration, education, health, and social services or other services. Hours spent in school or in domestic work were listed separately.

Schooling in Ethiopia: Primary school covers grades 1 to 8, and by the end of eighth grade pupils must pass a national examination before they are allowed to start secondary school (grades 9 and 10). Students that pass another national examination at the end of the tenth grade are allowed to enroll in two years of "preparatory" school (grades 11 and 12) and those who also pass a 12th grade exam are eligible to enroll in public universities.[20]

Indicators of Women's Empowerment: When a woman of age 15–49 was present in a male-headed household (almost 90 percent of households were male-headed, see Table 1), interviewers asked the woman to list all members involved in decision-making related to 20 different issues. Such issues included the following: 1. Food eaten at home; 2. Routine purchases for household items such as cleaning supplies; 3. The woman's own clothes; 4. The clothes of the woman's spouse; 5. Children's clothes; 6. Children's education (to attend, and then continue); 7. The woman's health; 8. The health of the spouse; 9. Children's health; 10 Large expensive purchases for the household; 11. Giving money to the woman's parents/family; 12. Giving money to the spouse's parents/family; 13. Gifts for special occasions; 14. Monthly savings; 15. Sale of cattle; 16. Time the woman spends socializing; 17. Whether the woman works outside the household; 18. Number of children to have; 19. Contraceptive use; 20. When daughters can marry.

We construct two indicators of empowerment. The first is the fraction of domains for which the woman is included as one of the decision-makers, while the second is calculated in the same way but including only domains with a more distinct "economic" content, which we roughly identify to be items 10–15 and 17 in the list above. In several instances a specific decision was not relevant for the household, in which case the decision was not considered in calculating the indicator. For instance, in a family with no children or cattle, items 5, 6, 9, 15, and 20 would not be relevant,

[20] See Section P of the 2008 Statistical Abstract of Ethiopia, http://www.csa.gov.et/surveys/National statistics/national statistics 2008/SectionP-Education.pdf.

so the indicator would be calculated as the fraction of 15 decisions where the woman was involved as a decision maker.

Appendix B: Supplementary Tables

Table B1—Predictors of Actual Treatment Status

Dependent variable: treatment = 1	Coefficient	SE
Randomly assigned to treatment	0.645	0.071***
Number of adults in households	0.012	0.012
Number of children in households	0.004	0.005
Household head is male	0.002	0.021
Age of household head	−0.001	0.001
Household head has no formal education	−0.031	0.031
Household has loans from banks/cooperatives	0.012	0.042
Household has loans from informal lenders	−0.166	0.045***
At least one loan initiated by a woman	−0.060	0.046
Agriculture is main economic activity of household	−0.043	0.028
Total revenue from crop sales last 12 months	0.002	0.002
Total expenditure for crops last 12 months	0.000	0.001
Number of large animals owned	0.004	0.002*
Total value of livestock	0.001	0.001
Total revenues from livestock sales last 12 months	−0.002	0.001**
Total sales from nonfarm self-employment last 12 months	0.000	0.000
Total costs for nonfarm self-employment last 12 months	0.002	0.002
Number of nonfarm self-employment activities	−0.051	0.042
Number of nonfarm self-employment activities	0.033	0.035
Transfers in cash or kind last 12 months	−0.003	0.002
Income from wages last 12 months	−0.002	0.001*
Total value of selected assets	0.002	0.004
Surface water as main source for drinking	−0.030	0.032
Number of months of insufficient food	−0.007	0.007
Distance to nearest market (minutes)	0.001	0.001
Constant	0.254	0.070***
Observations	6,410	
R^2	0.46	
p-value, H_0: all slopes equal to zero	0.011**	

Notes: Data from baseline (2003) survey. The figures are regression coefficients from an OLS regression of a dummy for actual treatment status (whether ACSI/OCSSC actually started operating in the PA) on a dummy for the randomly assigned treatment status (whether they were randomly assigned to start operating in the PA) and a series of predictors. Standard errors and tests are robust to intra-PA correlation. All statistics are calculated using sampling weights. All figures expressing monetary values are quartic roots of values in 2006 Birr.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table B2—Changes in Sample Composition

	Means at follow-up				Test of equality	Sample
	Both (1)	Credit (2)	FP (3)	Control (4)	(p-value) (5)	size (6)
Panel A. Randomly assigned treatment						
Lived in current village of residence < 4 years	0.032	0.023	0.041	0.038	0.5463	12,674
Number of adults (16 or above)	2.6	2.7	2.5	2.7	0.3818	12,675
Age of household head	40.9	41.5	41.2	42.4	0.5866	12,647
Household head is male	0.871	0.887	0.856	0.875	0.7275	12,647
Household head has no formal schooling	0.663	0.662	0.703	0.597	0.8788	12,647
Household head ≥ primary schooling	0.075	0.069	0.064	0.084	0.4226	12,647
Panel B. Actual treatment						
Lived in current village of residence < 4 years	0.022	0.025	0.065	0.036	0.5536	12,674
Number of adults (16 or above)	2.6	2.7	2.4	2.6	0.0500**	12,675
Age of household head	41.2	41.8	40.0	42.5	0.5693	12,647
Household head is male	0.880	0.889	0.847	0.863	0.7103	12,647
Household head has no formal schooling	0.687	0.585	0.708	0.649	0.6881	12,647
Household head has no formal schooling	0.059	0.091	0.085	0.067	0.8394	12,647

Notes: Data from baseline (2003) and endline (2006) surveys. The figures in columns 1–4 are means of the variables calculated for each treatment arm, defined according to the random assignment (panel A) or according to actual implementation (panel B). The figures in column 5 are p-values for the test of equal changes between baseline and endline across the four arms. The tests are performed by estimating OLS regressions of each household-specific outcome on PA-fixed effects, a dummy equal to one for endline observations, and this same dummy interacted with three treatment-specific dummies equal to one if the household lived in a PA where assigned (panel A) or actual (panel B) intervention was as indicated in columns 1–3. The test is robust to intra-PA correlation of residuals. All statistics are calculated using sampling weights.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table B3—Predictors of Borrowing

Household head is male	0.0656	(0.0221)***
Age of household head	−0.0004	(0.0006)
Head has completed at least primary schooling	−0.0040	(0.0278)
Head has no formal schooling	−0.0664	(0.0202)***
Number of adults in household	0.0091	(0.0071)
Number of 6–15 year old children in household	0.0220	(0.0055)***
Number of children below age six in household	0.0199	(0.0083)**
Household has resided in village for less than four years	−0.1384	(0.0642)**
Cultivable land owned (Hectares)	0.0205	(0.0042)***
Number of sleeping rooms in dwelling	0.0239	(0.0145)
Household has nonfarm business more than three years old	0.0622	(0.0245)**
Intercept	0.1119	(0.0416)

Notes: Endline data (2006) from PAs where ACSI/OCSSC operated during the study period. Standard errors (in parentheses) are robust to intra-PA correlation. The figures correspond to OLS estimates with sampling weights of a linear probability model where the dependent variable is a dummy = 1 if the household had an outstanding microloan at follow-up. The regression also includes PA fixed effects. None of the predicted probabilities of borrowing lie outside the unit interval (min = 0.034, max = 0.808, mean = 0.268). Sample size n = 3,528, with 23 observations (< 1 percent) dropped because of missing values in one or more of the predictors.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Appendix C: Construction of Indexes of Dependent Variables and Correction for Multiple Inference

In Tables 2–7, for each "family" of outcomes we construct a household-level index of standardized outcomes in the spirit of Kling, Liebman, and Katz (2007). To illustrate, the results labeled "Any loan" in panel C of Table 2 are estimated as follows. First, each of the outcomes listed in columns 1–6 of panel A of the same table is standardized by subtracting the mean and dividing by the standard deviation of the variable calculated for control areas ($MF = FP = 0$) at endline. Next, the index is calculated as the average of the six standardized variables. Finally, the coefficient β_{MF} is estimated as usual from model (1) using the index as dependent variable. All the five indexes in Tables 2–4 are calculated in a similar way. In Tables 6 and 7 several outcomes are defined at the individual level and are often missing by construction (for instance, the fraction of children 6–15 in school is only defined for households where any child in this age group is present). In such cases, we adopt the following (admittedly ad-hoc) procedure: when the variable is missing (for instance because the household does not include any individual in a given demographic group) we impute a value equal to the mean over all nonmissing observations recorded at the same time *t* and in the same randomly assigned group. The calculation of the index and the estimation proceeds then as described earlier, with the provision that the standardization is done using mean and standard deviations calculated without including the imputed values.

All variables included in the index are transformed in a way that larger, positive values are loosely "desirable." For instance, net profits and child schooling are easily recognized as desirable and are thus not transformed, while the number of months of food insecurity or the fraction of girls whose primary activity is domestic chores enter the index with a negative sign. Admittedly, in some cases such classification is ambiguous, for instance in the case of adult labor supply, or of costs incurred in self-employment activities. The caption in each table specifies which outcomes are multiplied by -1 before the inclusion in the index, but in general we follow the convention that outcomes that signal an increase in economic activity led by adults are "desirable."

To account for the multiplicity of tests in inference, for each index we also report a more conservative *p*-value estimated as described in Hochberg (1988). This proceeds as follows. Suppose that *m* tests are carried out using an α significance level, and let $p_1, \ldots, p_m$ denote the *m* individual *p*-values *ranked in increasing order*. Then, for any $i = m, m-1, \ldots, 1$ if $p_i < \alpha/(m-i+1)$ then the null is rejected for all $i' \leq i$. The modified *p*-values in the tables are thus calculated as $p_i(m-i+1)$. In our empirical context $m = 7$ and the ranked p-values are $p_1 < 0.001$, $p_2 < 0.001$, $p_3 = 0.042, p_4 = 0.088, p_5 = 0.126, p_6 = 0.180$, and $p_7 = 0.353$. It is easy to see that while $p_1 < 0.05/(7-1+1)$ and $p_2 < 0.05/(7-2+1)$, the *p*-values are always larger than $\alpha/(m-i+1)$ for all $i > 2$. Hence, following this procedure the null of no impacts can only be rejected for the two indexes in Table 2, that is, for the prevalence of borrowing and the amounts borrowed.

REFERENCES

Angelucci, Manuela, Dean Karlan, and Jonathan Zinman. 2015. "Microcredit Impacts: Evidence from a Randomized Microcredit Program Placement Experiment by Compartamos Banco." *American Economic Journal: Applied Economics* 7 (1): 151–82.

Armendáriz de Aghion, Beatriz, and Jonathan Morduch. 2005. *The Economics of Microfinance.* Cambridge: MIT Press.

Attanasio, Orazio, Britta Augsburg, Ralph De Haas, Emla Fitzsimons, and Heike Harmgart. 2015. "The Impacts of Microfinance: Evidence from Joint-Liability Lending in Mongolia." *American Economic Journal: Applied Economics* 7 (1): 90–122.

Augsburg, Britta, Ralph De Haas, Heike Harmgart, and Costas Meghir. 2015. "The Impacts of Microcredit: Evidence from Bosnia and Herzegovina." *American Economic Journal: Applied Economics* 7 (1): 183–203.

Banerjee, Abhijit Vinayak. 2013. "Microcredit Under the Microscope: What Have We Learned in the Past Two Decades, and What Do We Need to Know?" *Annual Review of Economics* 5: 487–519.

Banerjee, Abhijit, Esther Duflo, Rachel Glennerster, and Cynthia Kinnan. 2015. "The Miracle of Microfinance? Evidence from a Randomized Evaluation." *American Economic Journal: Applied Economics* 7 (1): 22–53.

Brau, James C., and Gary M. Woller. 2004. "Microfinance: A Comprehensive Review of the Existing Literature." *Journal of Entrepreneurial Finance and Business Ventures* 9 (1): 1–26.

Buera, Francisco J., Joseph P. Kaboski, and Yongseok Shin. 2013. "The Macroeconomics of Microfinance." Federal Reserve Bank of St. Louis Research Division Working Paper 2013-034A.

Crépon, Bruno, Florencia Devoto, Esther Duflo, and William Parienté. 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 (1): 123–50.

de Mel, Suresh, David J. McKenzie, and Christopher Woodruff. 2009. "Measuring microenterprise profits: Must we ask how the sausage is made?" *Journal of Development Economics* 88 (1): 19–31.

Desai, Jaikishan, and Alessandro Tarozzi. 2011. "Microcredit, Family Planning Programs and Contraceptive Behavior: Evidence from a Field Experiment in Ethiopia." *Demography* 48 (2): 749–82.

Fulford, Scott L. 2013. "The effects of financial development in the short and long run: Theory and evidence from India." *Journal of Development Economics* 104: 56–72.

Heckman, James, Robert LaLonde, and Jeffrey Smith. 1999. "The economics and econometrics of active labor market programs." In *Handbook of Labor Economics*, Vol. 3A, edited by Orley Ashenfelter and David Card, 1865–2097. Amsterdam: North Holland.

Hochberg, Yosef. 1988. "A sharper Bonferroni procedure for multiple tests of significance." *Biometrika* 75 (4): 800–802.

Imbens, Guido W., and Joshua D. Angrist. 1994. "Notes and Comments: Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (2): 467–75.

Kaboski, Joseph P., and Robert M. Townsend. 2012. "The Impact of Credit on Village Economies." *American Economic Journal: Applied Economics* 4 (2): 98–133.

Karlan, Dean, and Jonathan Zinman. 2010. "Expanding Credit Access: Using Randomized Supply Decisions to Estimate the Impacts." *Review of Financial Studies* 23 (1): 433–64.

Karlan, Dean, and Jonathan Zinman. 2011. "Microcredit in Theory and Practice: Using Randomized Credit Scoring for Impact Evaluation." *Science* 332 (6035): 1278–84.

Kling, J. R., J. B. Liebman, and L. F. Katz. 2007. "Experimental analysis of neighborhood effects." *Econometrica* 75 (1): 83–119.

Macro International. 2007. *Trends in Demographic and Reproductive Health Indicators in Ethiopia: Further analysis of the 2000 and 2005 Demographic and Health Surveys Data*. United States Agency International Development. Calverton, MD, January.

Morduch, Jonathan. 1999. "The Microfinance Promise." *Journal of Economic Literature* 37 (4): 1569–1614.

Ofcansky, Thomas P., and Berry LaVerle. 1991. *Ethiopia: A Country Study*. Washington, DC: Federal Research Division, Library of Congress.

Pronyk, Paul M., James R. Hargreaves, Julia C. Kim, Linda A. Morison, Godfrey Phetla, Charlotte Watts, Joanna Busza, and John D. H. Porter. 2006. "Effect of a structural intervention for the prevention of intimate-partner violence and HIV in rural South Africa: a cluster randomised trial." *Lancet* 368 (9551): 1973–83.

Tarozzi, Alessandro, Jaikishan Desai, and Kristin Johnson. 2015. "The Impacts of Microcredit: Evidence from Ethiopa: Dataset." *American Economic Journal: Applied Economics*. http://dx.doi.org/10.1257/app.20130475.

Wolday, Amha. 2002. "Product Development in the Ethiopian Microfinance Industry: Challenges and Prospects." Association of Ethiopian Microfinance Institutions (AEMFI) Working Paper 4.

World Bank. 2008. *Global Purchasing Power Parities and Real Expenditures, 2005*. Washington, DC: International Comparison Program.

Wydick, Bruce. 1999. "The Effect of Microenterprise Lending on Child Schooling in Guatemala." *Economic Development and Cultural Change* 47 (4): 853–69.

This article has been cited by:

1. Abhijit Banerjee, Esther Duflo, Rachel Glennerster, Cynthia Kinnan. 2015. The Miracle of Microfinance? Evidence from a Randomized Evaluation. *American Economic Journal: Applied Economics* **7**:1, 22-53. [Abstract] [View PDF article] [PDF with links]