

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

Heterogeneous impact dynamics of a rural business development program in Nicaragua



Michael R. Carter^a, Emilia Tjernström^{b,*}, Patricia Toledo^c

^a University of California, Davis, 2103 Social Sciences and Humanities, One Shields Avenue, Davis, CA 95616, USA

^b University of Wisconsin, Madison, 1225 Observatory Drive, Madison, WI 53706, USA

^c Ohio University, Bentley Annex 345, Athens, OH 45701, USA

ARTICLE INFO

JEL classification:

I32
O12
O13
Q12
Q16

Keywords:

Agricultural productivity
Impact evaluation
Heterogeneous treatment effects

ABSTRACT

We study the impacts of a rural development program designed to boost the income of the small-farm sector in Nicaragua. Exploiting the random assignment of treatment, we find statistically and economically significant impacts on gross farm income and investment in productive farm capital. Using continuous treatment estimation techniques, we examine the evolution of program impacts over time and find that the estimated income increase persists and that the impacts on productive capital stock continue to rise even after the program concluded. Additionally, panel quantile methods reveal striking heterogeneity of program impacts on both income and investment. We show that this heterogeneity is not random and that there appear to exist low-performing household types who benefit little from the program, whereas high-performing (upper quantile) households benefit more substantially. Analysis using generalized random forests, a machine learning algorithm, points toward greater program impacts for households who were disadvantaged at baseline. Even after controlling for this source of heterogeneity, we find large and persistent differences in how much different types of households benefited from the program. While the benefit-cost ratio of the program is on average positive, the impact heterogeneity suggests that business development programs aiming to engage farm households as agricultural entrepreneurs have limitations as instruments to eliminate rural poverty.

With severe poverty concentrated in rural areas of the developing world, there have been numerous efforts to engage the rural poor as entrepreneurs. The hope is that with the right information, investment and market connectivity, the poor can boost their incomes, invest in their children and work their way out of poverty. However, in contrast to cash transfer programs, which address poverty by “just giving money to the poor” (Hulme et al., 2010), business development programs that treat the poor as incipient entrepreneurs exhibit several characteristics that shape their effectiveness and challenge the evaluation of their impacts:

1. DYNAMICS: By providing new information, incentives and connections, we might expect entrepreneurially-focused programs to induce beneficiaries to learn and to co-invest in their new opportunities, therefore making it likely that impacts will evolve over time.¹
2. PARTICIPATION: While most people can and do accept cash transfers if one is offered, entrepreneurial programs require specialization, investment and risk-taking and are thus unlikely to appeal to all

poor households, limiting their reach as an anti-poverty strategy.

3. HETEROGENEITY: Most entrepreneurial activities generate both winners and losers, based on luck and/or complementary inputs that differ across households (e.g., talents and skills), again limiting the average effectiveness of programs that address the poor as potential entrepreneurs.

While studies of other programs that address the poor as entrepreneurs have noted that partial participation blunts program impacts (see e.g. Banerjee et al. (2011)), this paper uses data from a 5-year study of a Nicaraguan program that was randomly rolled out over time to explore all three of these dimensions of addressing the rural poor as incipient agricultural entrepreneurs.

Nicaragua, one of the poorest countries in the western hemisphere, is no exception to the pattern in which poverty is most severe in rural areas. Beginning in 2007, the government of Nicaragua launched a rural business development program (RBD) in cooperation with the Millennium Challenge Corporation (MCC), the United States govern-

* Corresponding author.

E-mail addresses: mrcarter@ucdavis.edu (M.R. Carter), tjernstroem@wisc.edu (E. Tjernström), toledot@ohio.edu (P. Toledo).

<https://doi.org/10.1016/j.jdevec.2018.11.006>

Received 31 August 2017; Received in revised form 2 November 2018; Accepted 23 November 2018

Available online 14 December 2018

0304-3878/© 2018 Elsevier B.V. All rights reserved.

ment foreign aid agency. The RBD was designed to address a set of constraints that policy-makers believed restricted the productivity and incomes of resource-scarce rural households. Specifically, the RBD offered marketing interventions, temporary input subsidies and/or co-investment incentives, and extension services. Contact with farmers lasted 24 months, after which farmers were expected to continue on with their own knowledge and resources.

While none of these interventions are novel, earlier non-experimental efforts to evaluate similar programs' effectiveness have confronted identification problems because of endogenous program placement and participation (see e.g. Evenson (2001) and Anderson and Feder (2003)). Several recent studies employ experimental designs to solve these identification problems: Bardhan and Mookherjee (2011) and Carter et al. (2013) find positive impacts of subsidized agricultural inputs to farmers in West Bengal and Mozambique, respectively. Cole and Fernando (2016) find that farmers respond to mobile-phone based agricultural information delivery in Gujarat, while Ashraf et al. (2009) estimate that over the short term at least, extension services positively impact incomes.² However, unlike Carter et al. (2014) who find that positive impacts evolve but persist over time, the impacts in Ashraf et al. (2009) dissipate over time, reinforcing the importance of paying attention to impact dynamics.

To evaluate the impacts of the Nicaraguan RBD program, we worked with program implementers to select a random subset of program-eligible households for inclusion in the study. These study households were in turn randomly split into early and late treatment groups, as the treatment could not be rolled out to all households at once due to capacity constraints on the implementer side. Early-treatment households were offered the program in 2007, shortly after completing a baseline survey. Late-treatment households were offered the program some 20 months later, after the second (mid-line) survey. A third (end-line) survey took place two years later, in 2011. The result is a 3-round panel data set, in which final exposure to the program randomly varies across households from as much as almost 4 years to as little 18 months.³ We exploit the fact that the late-treatment households made their program participation decisions after the mid-line survey, which allows us to realize statistical efficiency gains by focusing the analysis only on those who participate in the program (a double-complier sample).⁴ Further, while the baseline and mid-line data have a conventional binary-treatment/control structure, the full 3 rounds of the panel data allow us to use fixed-effect continuous treatment estimators to trace out program impacts over time.

Using this design, we explore the RBD's impacts on three key outcome variables: income in targeted agricultural activities, investment in productive capital stock, and per-capita household consumption expenditures. We find significant average impacts of the RBD on income and capital stock investments, but not on household consumption expenditures. Our estimates show that the impacts evolve over time and suggest that the standard binary treatment estimates based on the mid-line

¹ As King and Behrman (2009) point out, programs with significant learning and adoption components are unlikely to attain steady-state effectiveness soon after an intervention begins. In this study, we therefore pay particular attention to how the observed impacts evolve over time.

² See also Feder et al. (1987) for an earlier study of extension service intensification using a quasi-experimental research design, which uncovers positive but diminishing effects of extension services.

³ While most of the variation in treatment duration is between early and late groups, a small portion also results from variation within treatment groups. For example, a hiring delay for a livestock program trainer led early treatment livestock farmers to receive a shorter treatment duration than farmers in other production rubrics. While we did not randomize within-group treatment order, there is no evidence (qualitative nor quantitative) that the ordering was anything but random. Please see Section 1.2 for more details on the roll-out.

⁴ The validity of this double complier sample is discussed extensively below. The full sample is used to test the robustness of the two-sided complier results.

data present an incomplete picture of long-term impacts. In particular, the average impacts of the RBD program on farm-level capital stocks continue to grow after the mid-line survey, suggesting that longer time frames may be necessary to appropriately evaluate these types of programs. The failure of consumption expenditures to respond to the RBD program appears to reflect households' decisions to reinvest income increases rather than consume them.

Looking beyond average impacts, we employ the panel quantile regression techniques developed by Abrevaya and Dahl (2008) to determine the extent to which estimated average impacts represent the range of impacts experienced by program participants. The analysis reveals quite striking heterogeneity in impacts. Beneficiaries in the 75th conditional quantile of incomes enjoy much larger impacts than those in the lower quantiles, and a similar pattern holds for investment in farm capital. Indeed, program impacts on income are estimated to be three to four times greater for households in the top conditional quantiles compared to the lowest quantile, and top quantile impacts on capital investment are almost twice those in the lowest quantile. The average impact paths appear steeper than those estimated by a median regression, as the former estimates are driven up by the OLS regression's sensitivity to extreme values.

While Bandiera et al. (2017) find a similar pattern of heterogeneity in their analysis of BRAC's asset transfer and business development program in Bangladesh, they are unable, in their own words, to "uncover the root causes" of this heterogeneity. Many potential explanations exist for why the impacts of anti-poverty business development programs may be nil in the lower quantiles, and we are able to bring our multi-period data to bear on this question. The analysis in Section 4 shows that there is relatively little movement of households across quantiles over time. That is, there appear to be "lower quantile type" households who benefit little from the RBD program, and high types who benefit substantially.

We further employ a Generalized Random Forest (GRF) to search for the source of this heterogeneity. Those households who seem to benefit most from the program were among the least advantaged at baseline. Yet even after controlling for baseline disadvantage, we still observe substantial residual heterogeneity. This suggests that the programs may work best for households that enjoy some as yet unidentified entrepreneurial talent. This finding, along with a 70% program participation rate suggests that the RBD is an effective tool for raising incomes for some: it places a substantial minority of households on an upward economic trajectory. However, it also appears to be an ineffective tool for many others. These observations do not imply that programs like the RBD are bad policy, but suggest that by themselves they may be unable to raise the living standards of all targeted households.⁵

The remainder of this paper is organized as follows. Section 1 introduces the RBD and its roll-out, describes the data, and presents basic descriptive statistics and balance tests between the early and late treatment groups. Section 2 presents our empirical approach. Section 3 shows the average impact estimates for income, investment and consumption and explores the validity of our two-sided complier estimator. Section 4 looks beyond average impacts and estimates the extent and meaning of impact heterogeneity using both generalized quantile estimation and GRF. Section 5 concludes.

1. Background

Agriculture has played an important role throughout Nicaragua's history, but multiple constraints have conspired to prevent agriculture

⁵ By way of comparison, Banerjee et al. (2011) find that approximately one-third of intended beneficiaries declined participation in a business development program that offered a free asset transfer. These authors do not, however, break down the distribution of benefits across household types.

from reaching its productive potential—examples include a lack of basic infrastructure, low education levels, and low access to credit and technology. Nicaragua’s National Development Plan identified the Western Region of Nicaragua, which includes the departments of León and Chinandega, as having particularly high potential for agricultural growth. While high-potential, the area is also quite poor: the World Bank (2008) determined that more than 50 percent of households in the Western Region live in poverty.

In July 2005, the Millennium Challenge Corporation (MCC) signed a five-year, \$175-million compact with the Government of Nicaragua to develop a set of projects in the Western Region, with the objective of relaxing the aforementioned constraints. The compact had three components: a transportation project, a property regularization project, and the one we focus on here: a rural business development (RBD) project.⁶ This latter component aimed to raise incomes for farms and rural businesses by helping farmers develop and implement a business plan built around a high-potential activity.

1.1. Program description and research design

The Nicaraguan implementing agency (the Millennium Challenge Account, or MCA) identified the productive activities most suitable for inclusion in the program: beans, cassava, livestock, sesame, and vegetables. In order to be eligible, farmers had to own a small- or medium-sized farm, have some experience with one of these crops, be willing to develop a business plan together with extension agents, and contribute 70% of the cost of investments identified in the business plan. In addition, MCA and the implementers developed and applied activity-specific eligibility criteria (the precise rules are shown in Appendix A).⁷ Once farmers enrolled in the program and got their business plan approved, the RBD program worked with them for 24 months. While the exact benefits varied across the productive activities, all farmers received technical and financial training as well as supplies based on their individual business plan. Participating farmers also enjoyed co-investment benefits, either in the form of partial subsidies for improved agricultural inputs, or in terms of shared cost for individually or cooperatively owned equipment or installations (e.g., milking sheds or cooling tanks).

The research team grouped farmers into small geographical clusters of approximately 25 farmers, with a lead farmer identified for each. The randomization exploited implementer capacity constraints, which meant that not all eligible farmers could be enrolled in the project immediately. The research team worked with the RBD implementers to identify all the geographical clusters that would eventually be offered RBD services. The evaluation team then selected a subset of these clusters (146 in total) for random assignment to either early or late treatment status. These clusters were identified by professionals from the Millennium Challenge Account in Nicaragua (MCA-N), who worked under time constraints to identify a sufficient number of potential clusters within the target crops. While these clusters were not randomly selected from a larger set of clusters, they actually constituted the universe of potential clusters at the time that the study was rolled out. MCA-N later on had to seek out additional clusters to fulfill their program goals, but we have no reason to believe that MCA-N professionals had any incentive to include or exclude areas based on the expected outcomes of these farmers.

⁶ The MCC terminated a portion of the compact in June of 2009, reducing compact funding from \$175 million to \$113.5 million. While this action cut off the property regularization part of the program, the RBD Program was not affected by this partial project termination.

⁷ The impact of these eligibility criteria on the characteristics of the eligible population is described in Toledo and Carter (2010) who show that the RBD beneficiaries are found in the middle deciles of the rural income distribution of the areas where the program was implemented (the Departments of León and Chinandega).

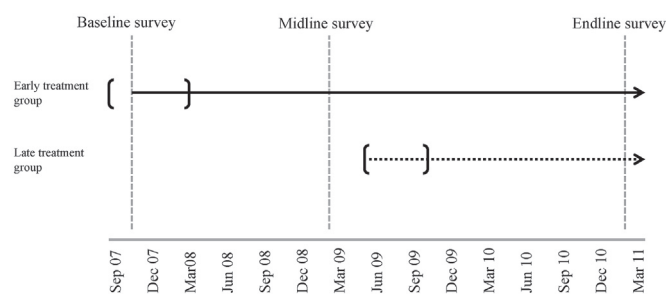


Fig. 1. Timeline of Received Treatment and Timing of Surveys.

Brackets around treatment start denote the range of treatment start dates for each group.

Once the researchers had randomly assigned clusters to early and late treatment status, 1,600 households were sampled from the roster of all eligible producers in these clusters, split equally between early and late areas. The randomization was blocked at the level of the crop to ensure that early and late groups had equal representation of the different production activities. Approximately 12 farmers were randomly selected for the study from each cluster. The 1,600 sampled households completed a baseline survey in late 2007, just as the RBD program was rolling out in the early treatment clusters. The mid-line survey took place approximately 18 months later, right before the late treatment group was offered the program. As illustrated in Fig. 1, the randomization and the timing of the surveys meant that the late treatment group functioned as a conventional control group at the time of the mid-line survey. Both early and late treatment clusters were then surveyed a third time in 2011. This roll-out strategy also provided quasi-random variation in the duration of time that households spent in the program, a feature that will prove important in the continuous treatment estimates presented below.

Two consultant firms, Chemonics and Technoserve, carried out the bulk of treatment roll-out. Since the consultants became engaged later than anticipated, professionals from the Millennium Challenge Account in Nicaragua (MCA-Nicaragua) initiated the roll-out in the beginning. MCA-Nicaragua staff were skilled professionals and rolled out the program according to the protocol agreed upon with the consultants, but due to limited capacity this resulted in a more spread-out treatment start than anticipated. Other than the bean clusters that we exclude from the analysis due to manipulation (they were treated early because the consultants believed they were better), we have no evidence that the roll-out timing was strategic.

For most of the program crops, we have all the price and quantity information needed to analyze target activity income, which is our primary outcome variable. The exception for this are the vegetable farmers. Due to the sheer number of crops that they farm, we were unable in our surveys to collect adequate information on these farmers. In the analysis to follow, we therefore drop the vegetable clusters (2 of these clusters were assigned to the early treatment and 2 to the late treatment). The consumption and investment results do not change if we include these farmers.

In addition, midway through the research process, the research team found that the randomization protocol had been violated for bean farmers in one sub-region (León), as local program implementers treated early farmer groups that had been randomized into late treatment status. We eliminate all bean clusters from this sub-region.⁸ Fortunately, the original research design was blocked at the sub-region level (in order to study a land titling program, which ultimately never took

⁸ Had the protocol violation been a random mistake, we might have reclassified and kept the offending clusters. However, both qualitative and quantitative evidence suggests that this violation of the protocol was driven by the program implementer’s desire to cherry-pick strong groups for early treatment.

Table 1
Summary statistics and baseline balance checks.

Variable	Full sample			Complier sample		
	Early Mean/SE	Late Mean/SE	Difference (early)–(late)	Early Mean/SE	Late Mean/SE	Difference (early)–(late)
Household characteristics						
Program farmer: age	51.123 (0.670)	53.037 (0.600)	–1.915**	50.476 (0.841)	52.773 (0.655)	–2.297**
Program farmer: education	4.456 (0.246)	4.004 (0.221)	0.452	4.818 (0.276)	3.996 (0.246)	0.822**
Program farmer: years of experience	20.934 (0.637)	21.311 (0.684)	–0.378	20.726 (0.676)	21.279 (0.889)	–0.553
Program farmer: gender (=1 for female)	0.137 (0.013)	0.132 (0.012)	0.005	0.137 (0.015)	0.119 (0.014)	0.018
Household members	5.251 (0.094)	5.482 (0.125)	–0.231	5.403 (0.118)	5.555 (0.144)	–0.152
Per capita expenditures	4157.200 (203.026)	4192.418 (281.960)	–35.218	4219.047 (236.150)	4067.813 (246.538)	151.234
Credit constrained (=1 if constrained)	0.393 (0.021)	0.405 (0.022)	–0.012	0.365 (0.024)	0.389 (0.028)	–0.024
Technical efficiency	0.597 (0.011)	0.602 (0.014)	–0.005	0.611 (0.012)	0.604 (0.017)	0.007
Farm characteristics						
Value of total capital (\$)	8515.709 (758.177)	7126.357 (632.196)	1389.351	9370.442 (844.159)	6998.305 (507.500)	2372.137**
Landholdings: owned (manzanas)	37.542 (3.868)	44.321 (6.589)	–6.779	42.575 (5.050)	46.143 (7.546)	–3.568
Landholdings: amt. planted in target crop	5.768 (0.901)	5.274 (0.514)	0.493	5.643 (1.028)	6.317 (0.777)	–0.674
Landholdings: amt. planted in maize	3.132 (0.188)	3.013 (0.114)	0.120	3.301 (0.256)	3.012 (0.133)	0.289
Target activity income	7511.586 (694.095)	8338.627 (802.124)	–827.041	8339.753 (789.080)	9064.863 (915.839)	–725.110
Share of seasons used improved seeds	0.131 (0.026)	0.145 (0.025)	–0.013	0.118 (0.027)	0.175 (0.031)	–0.057
<i>N</i>	695	701		454	453	

The asterisks in columns (3) and (6) denote significance levels of *t*-tests of the differences in the means across early and late treatment groups. Standard errors are clustered at the cluster level.

***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

place), so the elimination of these bean clusters does not damage the integrity of the experimental design. We do, however, lose some 150 observations from this excision of the bean cluster.⁹ In aggregate the elimination of vegetable and León bean clusters reduces the total sample size from 1600 farm households to 1396. Attrition was under 2 percent by the time of the third wave.

1.2. Data

Table 3 shows summary statistics and two separate balance checks. Because our control group (late treatment) was eventually offered the program, we check for statistical balance in two ways: columns (1) and (2) contain the means in the full sample across randomization status, with column (3) showing the difference between the two and the results of a *t*-test of equality. Columns (4)–(6) present the same means and differences for the subset of households who complied with their treatment status and enrolled in the program. Section 3 examines the comparability of the complier sample in the early and late groups in more detail. As can be seen in Table 1, the proportion of farmers who took up the treatment is 65.3% in the early treatment group, and 64.6% in the late treatment group.

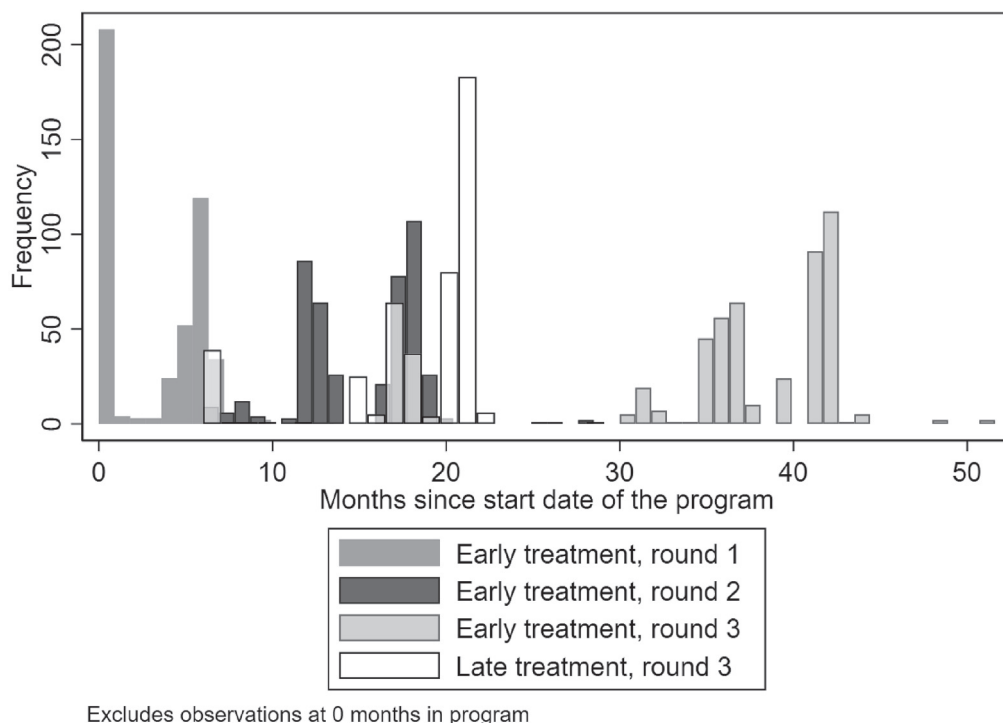
Most of the variable averages for the full sample suggest that the groups represent the same population, (i.e., they are not statistically significantly different from each other) and that the randomization was

⁹ Because the study was originally powered to detect the additional treatment arms defined by the land titling program, we should have ample power even after this loss of observations.

successful. An *F*-test of the joint significance of the baseline covariates fails to reject the null of no effect (*F*-stat: 1.72). Participating farmers in the late treatment group are on average two years older than the early treatment farmers, but we believe this arose by chance and it seems unlikely that these differences will interact substantially with the treatment. The difference in age is significant in the “complier sample” as well, and we now additionally see statistically significant differences in education. While these imbalances seem unlikely to be economically relevant, we might be more concerned by the imbalance in baseline capital ownership as it could interact with the treatment and affect treatment impacts. The heterogeneity analysis in Section 4.4 reveals no significant heterogeneity of treatment impacts with respect to this variable.¹⁰ This provides at least suggestive evidence that these baseline impacts are not driving the bulk of our results. Other key productive inputs such as amount of land owned and farmed do not differ significantly across the two groups, and the fraction of farm households that are credit constrained is nearly identical between the two groups.¹¹

¹⁰ We also estimate binary heterogeneity analyses with baseline capital. In results available from the authors, we interact the treatment dummy with a binary indicator for being above or below median capital levels. Capital does not interact with treatment based on these results either.

¹¹ Following Boucher et al. (2009) a farm is classified as credit constrained if they have positive demand for a loan at the current rate of interest but indicate that either they are quantity-rationed in the sense that they cannot qualify for a loan (e.g., they lack required collateral assets), or they are risk-rationed in the sense that they are afraid to risk the collateral required by the loan (see Boucher et al., 2008).



Excludes observations at 0 months in program

Fig. 2. Distribution of the duration of RBD treatment (dual complier sample) – excluding pre-treatment observations.

Fig. 2 shows a histogram detailing the distribution of months in the RBD program for the sample of compliers across all three survey rounds. The figure excludes observations with zero months of treatment since this group (comprised of the early and late treated households at baseline, plus the late treated households at the mid-line), dwarfs the other categories. Despite some bunching, the data show reasonable dispersion: the data contain households observed with as little as 1 month in the program up to as much as 50 months in the program. The largest overlap is between early treatment farmers in round 2, and the late treatment farmers in round 3. The last group, with 30–40 months of exposure, is comprised exclusively by early treatment households in round 3. The variation in length of program exposure comes from a combination of the variation of program start and survey timing, and is the variation that we exploit in the continuous treatment estimators explained in the next section.¹²

2. Econometric methodology

Our three outcome variables of interest—target activity income, investment, and household consumption—capture both direct and indirect channels of impact. The small-farm intervention was designed to enhance the access of small farmers to improved technologies and to markets, so we begin by examining program impacts on income in the target crops. We define income as the total value of production in the target crop, calculated using the prices that the household obtained for the part of their harvest that was sold.¹³ This measure is likely to overstate the actual impacts on household income as it ignores any reallocation of fixed inputs such as owned land and family labor to the target crop, reallocation of labor away from other

¹² We can explain more than 90% of the total variation in treatment duration by the randomized assignment to early and late groups and time dummies.

¹³ Note that the RBD was intended to allow farmers to receive better prices for their produce, hence it is important that we value output based on prices actually received. When a farmer did not sell any part of their crop, we valued output using the mean price in their geographical cluster by season and crop.

income-generating activities, as well as the costs of purchased inputs. We believe that it still provides important insights into the program’s impacts. We then examine two domains where we would expect to see impacts only if the program actually enhanced total household income, namely consumption expenditures and investment in productive capital.

We evaluate the impacts of the program using two main econometric approaches. First, we estimate local average treatment effects (LATE) using Analysis of Covariance (ANCOVA) estimation on two different samples. Our standard LATE approach uses randomly assigned treatment assignment to instrument for treatment status to compute the Treatment on the Treated (ToT) estimates. We compare this approach to a two-sided complier (2SC) estimator, which is similar to the standard ToT approach, but allows us to gain power. Second, we employ a continuous fixed effect treatment estimator to examine the evolution of impacts over time.

To motivate our focus on continuous treatment effects, note that the workhorse impact evaluation estimators assume that program participation is a binary state—either a household receives the treatment or it does not. While this approach deals well with treatment heterogeneity across treated units (hence the derivation of local average treatment effects), it is not equipped to deal with impacts that evolve over time. Programs like the RBD that provide information, improve market access and enhance investment incentives might be expected to achieve their full impact over a medium-term time period of unknown duration (especially for credit-constrained households that must self-finance investments). In the extreme case, they may even cause short-term decreases in key indicators as households switch livelihood strategies or even cut consumption to fund investments (Keswell and Carter, 2014), for example, find evidence of these short-term dips in the case of land redistribution in South Africa).

To better frame these issues, consider the hypothetical impact relationships for the RBD intervention illustrated in Fig. 3. The solid step-wise line illustrates what we might expect to see for the early treatment group, while the dashed step-wise line illustrates the same for the late treatment group. The horizontal axis shows roughly where

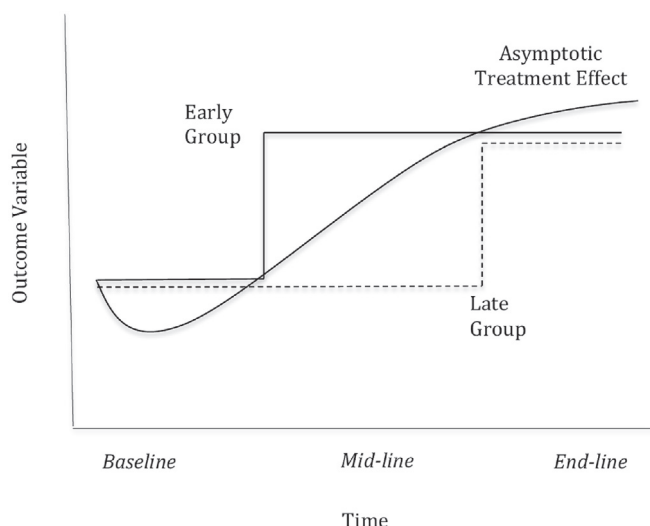


Fig. 3. Hypothetical impact patterns.

the different survey rounds were undertaken relative to the treatment. If the program had reached its full long-term impact on the early (late) treatment group by the time of the second (third) round survey, then conventional binary estimators would work well. In this case we would expect the data to trace out impact patterns similar to the step functions.

On the other hand, if the impact of the program evolves more slowly over time (for example, with an initial dip followed by a slow rise toward a long-run or asymptotic treatment effect), then our data would be generated by a non-linear impact or duration response function in which impact depends on the duration of time in the program. Impacts measured at mid-line using standard binary treatment estimators (which work well when the data follow a pattern shown by the step functions in Fig. 3) may reveal muted effects that would not accurately represent the long-run program impacts. The remainder of this section describe both our binary impact estimators as well as the more general continuous treatment model designed to capture an unknown impact pathway.

2.1. Binary treatment model

In the binary analysis, we use ANCOVA estimation for the basic treatment estimates. McKenzie (2012) demonstrates that ANCOVA estimation can result in substantial improvements in power compared to the more common difference-in-difference specifications. The power gains are especially large when the data have low autocorrelation, as is the case for many outcomes in rural development settings like ours.

We begin by defining two indicator variables:

- B_i indicates treatment assignment for household i , equaling 1 for eligible farmers who were assigned to the early treatment group, and 0 for those assigned to the control or late treatment group.
- D_i indicates whether or not a farmer actually participated in the program when invited, so that $D_i = 1$ for treated invited farmers and $D_i = 0$ for non-compliers, who refused the program.

The local average treatment effect (LATE) can then be estimated by the coefficient δ in the instrumental variables ANCOVA regression:

$$y_{i,2} = \alpha + \delta \hat{D}_i + \theta y_{i,1} + \beta' X_{i,1} + \varepsilon_i \quad (1)$$

where $y_{i,2}$ is the outcome variable in the second (post-treatment) period, \hat{D}_i is D_i instrumented by B_i (the assignment to early treatment), $y_{i,1}$ is the baseline, pre-intervention value of the outcome variable for household i , and $X_{i,1}$ is a vector of baseline variables for which we want to control. Since the intervention was randomly assigned, the use of B_i

as an instrument for D_i allows us to obtain consistent estimates of δ . We will present our results both with and without covariates, since the intervention was randomized.

Looking ahead to the continuous treatment model, where we only observe duration of time in treatment for the compliers (households with $D_i = 1$), we also employ a two-sided complier estimator, which instead of instrumenting for program take-up restricts the sample to the complier sample, i.e. farmers in both early and later groups who joined the RBD program. We are able to do this thanks to our third round of data, in which we observe the take-up decisions of the late treatment group, i.e. those farmers who serve as controls in the midline survey. In this case, the vast majority of program costs were spent on participating farmers such that the estimated impacts on this subpopulation (i.e. the treatment on the treated) are likely the most relevant to policymakers.

The estimating equation for the 2SC estimator is the same as in Eq. (1), except that instead of instrumenting for D_i using treatment assignment we use the information from the third survey round to identify the compliers among the late treatment group. The validity of this 2SC estimator relies on the idea that the decision to enroll in the early and late treatment groups was structurally the same, so that we are in fact comparing like with like in using this estimator. This assumption is in addition to the usual no-interference assumption, i.e. that farmers who do not enroll in the program experience no effect from the treatment or the randomization. Section 3 examines the legitimacy of the similarity of the compliance decision in the early and late groups, and report all binary results using both standard LATE and 2SC estimators.

2.2. Continuous treatment model

As discussed in the beginning of this section, there are a number of possible reasons why the impact of the RBD program may have evolved over time. In addition to a possible initial dip in living standards when households first join the program and focus their resources on building up the targeted activity, there are at least three other reasons why the impact of the small-farm intervention may have changed over time. First, program beneficiaries may have experienced a learning effect, with their technical and entrepreneurial efficiency improving over time. Second, the asset program may have created a crowding-in effect if the program incentivized beneficiaries to further invest in their farms. As Keswell and Carter (2014) discuss, these second-round multiplier effects are what distinguish business development and asset transfer programs from cash transfer programs and other common anti-poverty policy instruments. Third, and less positively, if program impacts are short-lived (e.g., if treated farmers drop the improved practices as soon as the 24-month period of intense RBD involvement with their groups end), then impacts may dissipate over time.

One goal of this study is to estimate the impact dynamics and duration response function, and thus recover both the medium-to long-run impacts of the intervention and their time path. Both are of particular relevance from a policy perspective. Indeed, it is the prospect that a skill-building program like the RBD program will facilitate and crowd-in additional asset building that makes them especially interesting as an anti-poverty program. Note that as this roll-out was not built into the experimental design, but rather occurred by circumstance, there remains a possibility that the variation in roll-out is endogenous. If the implementers chose to roll out the treatment to the best farmers first, the continuous impact estimates might be sloped upward due to this selection, rather than the existence of impact dynamics. That said, the research team worked closely with the implementation team for the duration of the program, and feel quite confident that we would have found out if the roll-out timing had been strategic.

In addition, what we need for these results to be valid is that the duration of treatment is randomly determined, conditional on covariates. In other words, even if we believe that this secondary variation in

roll-out was purposeful, the key question is whether we believe that it remains correlated with the error term once we control for covariates. In our case, since we control for household fixed effects, the implementers would have had to have in mind quite sophisticated models to predict treatment impacts for this to be true. Our results are also robust to instrumenting for treatment duration with treatment assignment and survey month (which, as noted in Section 1, explains 90% of the variation in treatment duration).¹⁴ Furthermore, the results in Section 4.4 suggest that the steepest impact curves actually occurred for households that were relatively disadvantaged at baseline, and therefore unlikely to be chosen by a strategic implementer who wants to roll out to the “best” farmers first.

We begin our continuous analysis with a generalization of the binary response function to the continuous treatment case¹⁵:

$$E[y_{it} | d_{it}] = \alpha_i + \tau_2^d t_2 + \tau_3^d t_3 + f(d_{it}), \tag{2}$$

where d_{it} is the number of months since farm i was actively enrolled in the treatment at survey time t , t_2 and t_3 are round dummies, and $f(d_{it})$ is a flexible function that can capture the sorts of non-linear impacts illustrated in Fig. 3 above. These durations run from 0 to 50 months.¹⁶

Based on the semi-parametric estimates of 2 reported in Tjernström et al. (2013), we choose a cubic parametric form to represent the duration impact function, $f(d_{it})$. The household-specific fixed effect term, α_i , controls for all observed and unobserved time-invariant characteristics, including farming skill, soil quality, farmer education, etc. Importantly, the fixed effect estimator controls for any systematic or spurious correlation between time invariant household characteristics and duration of treatment.

While there are several computationally equivalent ways to consistently estimate a fixed effect model like Eq. (2), we build on the correlated effects model of Mundlak (1978) and Chamberlain (1982, 1984) in anticipation of later quantile regression analysis where such models are less easily estimated. We therefore write the individual fixed effects as a linear projection onto the observables plus a disturbance:

$$\alpha_i = \lambda_0 + X'_{i1} \lambda_1 + X'_{i2} \lambda_2 + X'_{i3} \lambda_3 + v_i,$$

where X_{it} denotes a vector of observables, which includes the time dummies and the duration variables. In our case, we have little reason to believe that the way in which the time-varying observables affect the individual effects differ between survey rounds, so we use the average of the time-varying covariates and write the fixed effect as

$$\alpha_i = \lambda_0 + \bar{X}'_i \bar{\lambda} + v_i.$$

Substituting this expression into (2) gives:

$$y_{it}(d_{it}) = \tau_2^d t_2 + \tau_3^d t_3 + f(d_{it}) + \lambda_0 + \bar{X}'_i \bar{\lambda} + [v_i + \epsilon_{it}] \tag{3}$$

where ϵ_{it} is the error associated with the original regression function, Eq. (2). Replacing $f(d_{it})$ with the cubic functional form suggested by the semi-parametric analysis yields:

$$y_{it}(d_{it}) = \tau_2^d t_2 + \tau_3^d t_3 + \zeta_1 d_{it} + \zeta_2 d_{it}^2 + \zeta_3 d_{it}^3 + \lambda_0 + \bar{X}'_i \bar{\lambda} + [v_i + \epsilon_{it}]. \tag{4}$$

¹⁴ The instrumental variables results are available from the authors upon request.

¹⁵ We could alternatively follow the generalization of propensity score matching to the continuous treatment case found in Hirano and Imbens (2004). The Hirano and Imbens estimator only exploits observations with strictly positive amounts of treatment. In our case, this would imply dropping the baseline data for all RBD participants as well as the mid-line data for the late treatment group. For development applications that employ this estimator, see Keswell and Carter (2014) and Agüero et al. (2010).

¹⁶ In a few cases, RBD activities began a few months prior to the baseline survey. For these cases, we have considered households in these clusters as treated at baseline, but their values for d_{it} can exceed the number of months between the first and third rounds of data collection.

OLS estimation of (3) allows us to consistently recover the fixed effect estimators of the impact response function parameters of interest.¹⁷

3. Average impact estimates

Using the binary and continuous treatment models developed above, this section presents estimated average RBD impacts for each of our three primary outcome variables: gross income in the targeted business activity, productive investment, and household living standards as measured using typical living-standards measurement survey consumption expenditure modules. Section 3.1 presents binary results using both the full sample and the 2SC estimator that restricts the sample to complier households. The 2SC complier estimates are strikingly similar to the IV estimates (but are more precisely estimated), suggesting that the compliers in the late treatment group are similar to those in the early treatment group and confirming that the program was carried out in a similar fashion for the two groups. As mentioned above, the compliance rate was around 65% for both early and late treatment groups.

Further, Fig. 4 shows the results from a probit regression of the program take-up decision on early/late treatment. The full table of results is shown in Appendix Table 2. The first model includes only the treatment assignment dummy, the second model adds in baseline characteristics, and the third interacts all the covariates with the treatment dummy. The figure reports the average marginal effects of each variable on the probability of program take-up with other variables held at their sample means, with the associated 90% confidence intervals. We interpret these results as providing little evidence of systematic selection into the program since very few of the variables are significant. Furthermore, the overall effect of being assigned to the early treatment group does not change across the models as we include covariates and interaction terms. Additionally, the partial of the response with respect to the treatment dummy is not statistically significantly different from zero (p -value = 0.703 in the fully interacted model). Taken together, we find no evidence that the take-up decision differed between the early and late groups. Section 3.2 presents results for the continuous treatment model, which uses only the complier sample.

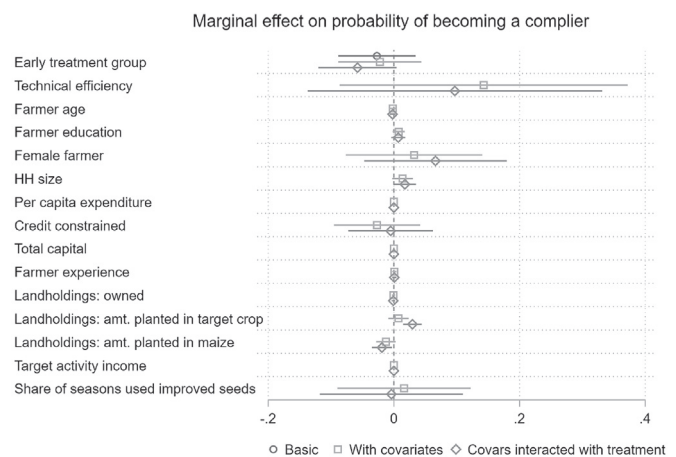


Fig. 4. Decision to take up program, by early and late treatment group status.

¹⁷ An important identifying assumption for the continuous treatment analysis is that the duration impact function, f , does not change. Some potential violations of this assumption include important differences in the climactic conditions during the first months of the program for the early treatment group compared to the late treatment group. We believe that climate should not have a major impact, since the overlap in the support of duration is fairly broad, and the clusters are geographically quite spread out. However, we acknowledge that this is unverifiable.

Table 2
Impact of RBD program on target activity income: ANCOVA estimates.

	ITT		LATE (IV)		LATE (complier sample)	
Early treatment	675.8 (554.0)	675.3 (575.5)	1061.1 (867.2)	1059.3 (896.9)	1169.8* (687.9)	1237.4* (719.4)
Baseline target activity income	0.84*** (0.072)	0.82*** (0.081)	0.84*** (0.071)	0.81*** (0.080)	0.90*** (0.073)	0.88*** (0.082)
Program farmer: education		64.9 (78.6)		59.5 (77.8)		-26.5 (98.2)
Program farmer: years of experience		-6.99 (23.4)		-7.06 (23.1)		-5.71 (31.3)
Household members		118.8 (113.4)		109.8 (114.7)		-55.1 (129.4)
Landholdings: owned		13.7 (8.32)		13.5* (8.19)		13.9** (6.83)
Share of seasons used improved seeds		1042.1* (595.5)		949.2 (621.3)		454.2 (733.3)
Program farmer: gender		-399.4 (509.6)		-411.4 (496.3)		-126.0 (662.1)
Constant	5834.5*** (975.7)	4254.7*** (1039.1)	5865.9*** (953.3)	4376.0*** (1006.6)	5677.4*** (1151.2)	5270.2*** (1459.2)
Observations	1341	1279	1341	1279	864	829
Adjusted R ²	0.577	0.577	0.579	0.579	0.618	0.616

Cluster-robust standard errors in parentheses.

*p < .1, **p < .05, ***p < .01.

All regressions control for crop fixed effects. Share of improved seeds measures the share of seasons at baseline that the beneficiary household used improved seeds/methods.

3.1. Binary impact estimates

Before turning to the continuous treatment estimators that allow us to exploit the full variation in our data, this section presents standard binary impact estimators which identify impacts based on the comparison at midline between early and late treatment groups.

Table 2 shows the RBD program's estimated impact on annual gross farm income from the activities targeted by the program. Income is measured in 2005 purchasing power parity adjusted US dollars. As discussed in Section 2, observed income increases in RBD-targeted crops do not necessarily imply increased overall incomes, as productive inputs could have been reallocated from other activities (e.g., maize or off-farm employment) to the target crops. Of these hypotheses, we can only seriously look at maize production since most farm households produce maize. In results available from the authors, we use the continuous treatment methods detailed in section 2.2 to show that maize area and production decline with duration of time in the program.¹⁸ Gross changes in target crop income may indeed overstate the total income impacts.

Given this observation, another way to examine this question is to look at the combined changes in consumption and investment spending. If households are unable to borrow to finance consumption or capital investment, then the sum of changes in consumption and investment spending can be no more than the true income change. Program implementers perceived that most farmers in our study area were tightly credit-constrained, and the original program plan included a component to boost farmer capital access. Ultimately, the credit program was not implemented and a decision was made to rely on the land titling element of the program to boost credit access. However, as discussed in section 1.1 above, the titling element of the program was never implemented.

Using the constraint elicitation methodology developed by Boucher et al. (2009), we find that 40% of farmers in our sample are non-price rationed in credit markets (see Table 1). At baseline, 49% of all farmers had borrowed money in either the baseline or the preceding year.

¹⁸ After a year in the program, maize area and production are estimated to decline by a marginally statistically significant 10%. After 30 months in the program the decline is estimated to be a statistically significant 30%.

The majority of loans were short-term production credit, and most loans were obtained by sesame and livestock producers whose standing crops and animals are collateralizable. While we cannot entirely rule out borrowing for capital investment and consumption, the evidence indicates that at least a majority of farmers faced constraints to doing so. Subject to these caveats, we include additional results in Appendix Table 7, showing treatment effects on the sum of changes in capital stock between rounds and household expenditure. If we worried that the total income were crowded-out by other activities or offset by increased input costs, we would not expect to see increases in this measure for credit constrained farmers. The treatment effects on this measure are remarkably similar to the estimated treatment coefficients on target activity income.

Table 2 shows the results from ANCOVA regressions on income at midline, with Intention-to-Treat (ITT) estimates in columns (1)–(2), and the full-sample LATE estimates in columns (3)–(4). These are the standard impact estimates under randomized treatment assignment, and make no assumptions about the uptake processes in the late treatment group, who act as control group in the midline. Columns (5)–(6) show the LATE results when we restrict the sample in both early and late treatment group to compliers only.

The ITT and LATE estimates show substantial average impacts of the program. The LATE estimates are roughly \$1,000, but they are not significant. However, the results from the 2SC estimator are very similar to the results in columns (3) and (4), which were obtained using standard instrumental variables regression in the full sample. The main difference is that the 2SC estimates are statistically significant at the 10-percent level; this increase in precision comes from not having to instrument for uptake. Economically, these point estimates imply an average income increase of around 17 percent at the midline. As discussed earlier, these impacts on income from the targeted activity are upper bound estimates of the impacts on net household income.

An important objective of beneficiaries' business plans was the accumulation of farm assets. With the objective of increasing farmers' productivity, the program provided some equipment or supported the construction of new productive installations once the business plan was approved. We follow the same strategy used in the previous section to examine the program's effects on capital investment. The outcome

Table 3
Impact of RBD program on farm investment: ANCOVA estimates.

	ITT		LATE (IV)		LATE (complier sample)	
Early treatment	556.9 (402.7)	604.1** (291.2)	889.5 (634.7)	954.4** (452.8)	525.7 (350.8)	717.7** (325.9)
Baseline investment	0.94*** (0.050)	0.96*** (0.028)	0.94*** (0.050)	0.96*** (0.028)	0.99*** (0.034)	0.98*** (0.036)
Program farmer: education		55.5 (44.9)		49.7 (43.4)		18.0 (48.2)
Program farmer: years of experience		16.3* (8.31)		16.4** (8.20)		11.1 (10.8)
Household members		127.2* (65.5)		117.7* (64.5)		127.2 (83.0)
Landholdings: owned		-13.2 (15.8)		-13.3 (15.5)		3.43 (2.98)
Share of seasons used improved seeds		513.4 (400.3)		425.3 (379.9)		351.3 (380.0)
Program farmer: gender		177.6 (419.1)		172.2 (412.4)		377.5 (559.4)
Constant	955.6*** (319.6)	121.7 (923.2)	959.2*** (319.2)	223.8 (917.6)	880.4** (411.8)	-603.7 (742.2)
Observations	1341	1260	1341	1260	860	817
Adjusted R ²	0.861	0.880	0.861	0.880	0.860	0.889

Cluster-robust standard errors in parentheses.

*p < .1, **p < .05, ***p < .01.

All regressions control for crop fixed effects. Share of improved seeds measures the share of seasons at baseline that the beneficiary household used improved seeds/methods.

variable used is the sum of investments in mobile capital (tools and equipment, excluding livestock) and in fixed capital (buildings, installations, and fences located on the farmer's land).¹⁹ The results are similar if disaggregated by type of capital. Note that in contrast to the income analysis, these measures are cumulative impacts (increments to a stock) over the period of observation.

Table 3 shows estimated program impacts on capital investment. The unadorned binary impact ITT estimates in column (1) are positive (\$557), but not statistically significant. Including covariates increases the precision of the estimates, and column (2) shows estimated impacts of \$604, significant at the 5-percent level.

Columns (3) and (4) of Table 3 report the LATE estimates with and without covariates. The estimated impacts of the program on farm investment are around \$900 and significant at the 5-percent level if covariates are included. The average household in our sample had around \$7000-\$8000 in total farm capital at baseline, so these program impacts correspond to an average increase of 12 percent over the baseline capital stocks. The results from the complier sample are again consistent with the standard LATE estimates, with the expected increase in statistical precision. The results also suggest that the increases in the value of production of the targeted income do translate into true increases in net household income. Indeed, the magnitude of the increases in farm capital suggest that most of the income increase was real and was allocated to productive investment.

The ultimate goal of the RBD was to boost the living standards of small-scale farm families. To investigate impacts that proxy for this dimension, we adopted the household expenditure module utilized in Nicaraguan living standards surveys. As with our money-metric outcome measures, we transformed consumption expenditures into 2005 purchasing-power-parity adjusted \$US. Because the number of household members fluctuates both within and between years, we adjusted the different expenditure components by potentially different household sizes to arrive at per-capita measures. Specifically, food expendi-

tures was converted to a per-capita measure using as a denominator the number of household members who had actually been in residence during the short recall period used to measure food spending. Other expenditure categories with longer recall periods were adjusted using the full roster of household residents (defined as those who habitually reside and sleep in the household).

As can be seen in Table 4, the effect of the program on consumption are small in magnitude, with negative point estimates in the full sample (columns (1)–(4)) and small but positive estimates in the complier sample. None of these estimates are statistically distinguishable from zero, and their absolute magnitudes are at most a few percent of baseline consumption levels. Coupled with the other impacts estimated using the mid-line, binary treatment estimators, these results indicate a program that boosted income and saw most, if not all, of that increased income devoted to capital accumulation.

3.2. Continuous treatment estimates

As discussed in Sections 1 and 2, there are multiple reasons to believe that the impacts of this type of program might evolve over time. To capture the potentially non-linear duration response functions, whereby impacts depend on how much time has passed since the producer enrolled in the RBD program, we exploit the fact that treatment was rolled out in a staggered fashion within the early and late treatment groups. This created variation in the duration of treatment (as shown in Fig. 2). The coefficient estimates of ζ_1 , ζ_2 , and ζ_3 , from estimating Eq. (4), our preferred cubic specification, are shown in Table 5. The rest of this section will discuss the graphical representations of these results graphically since the temporal path is somewhat hard to infer from the coefficients alone.

Turning first to impacts on program income, we see that the duration of time in program has a statistically significant impact on gross income in the treated activity. Drawing out the implications of the estimates shown in Table 5, Fig. 5 graphs the estimated cubic relationship with the associated 90-percent confidence intervals. Predicted target activity income at the start of the program is on average roughly \$9500. Income increases over the first two years in the program, flattening out at an average predicted target activity income of \$11,600, for an impact estimate of \$2100. This longer-term impact is twice the level

¹⁹ Some elements of fixed capital were difficult to value as they were often constructed by the farmer rather than purchased on the market. RBD program staff assisted with the evaluation, but a few items (in particular, erosion barriers and certain types of fencing) are not included in our measure of fixed capital.

Table 4
Impact of RBD program on household consumption: ANCOVA estimates.

	ITT		LATE (IV)		LATE (complier sample)	
Early treatment	-7.46 (137.0)	-12.7 (138.9)	-11.9 (216.2)	-20.0 (216.8)	159.3 (184.2)	105.0 (199.4)
Baseline expenditures	0.43*** (0.053)	0.35*** (0.063)	0.43*** (0.053)	0.35*** (0.062)	0.38*** (0.060)	0.29*** (0.061)
Program farmer: education		96.2*** (24.2)		96.4*** (24.6)		99.4*** (33.7)
Program farmer: years of experience		13.5** (6.36)		13.5** (6.27)		21.0** (9.56)
Household members		-229.6*** (25.9)		-229.4*** (25.8)		-277.2*** (34.0)
Landholdings: owned		2.30 (1.79)		2.31 (1.76)		3.74* (2.16)
Share of seasons used improved seeds		-9.17 (155.5)		-7.24 (161.0)		-188.7 (192.5)
Program farmer: gender		103.1 (230.1)		103.3 (226.4)		233.4 (334.3)
Constant	2030.1*** (260.9)	2716.4*** (348.9)	2030.2*** (259.1)	2714.7*** (340.6)	2288.3*** (312.0)	3078.2*** (385.3)
Observations	1378	1292	1378	1292	884	838
Adjusted R ²	0.384	0.456	0.384	0.455	0.299	0.379

Cluster-robust standard errors in parentheses.

*p < .1, **p < .05, ***p < .01.

All regressions control for crop fixed effects. Share of improved seeds measures the share of seasons at baseline that the beneficiary household used improved seeds/methods.

Table 5
Impact of RBD program on target activity income, investment and household consumption: Fixed effects, continuous treatment estimates.

	Farm Income	Investment	Expenditures
Months treated	254.0* (146.5)	299.1* (165.9)	-29.1 (46.0)
Months treated ²	-9.84 (8.53)	-11.4 (12.7)	2.14 (2.25)
Months treated ³	0.12 (0.14)	0.16 (0.22)	-0.039 (0.037)
Program farmer: education	361.2*** (119.4)	638.9*** (117.9)	243.8*** (41.1)
Program farmer: years of experience	64.0** (25.7)	61.8** (30.4)	19.3** (7.61)
Household members	-168.6 (145.9)	66.2 (147.5)	-466.2*** (39.9)
Landholdings: owned	44.1*** (13.3)	43.1*** (11.9)	7.22** (2.84)
Share of seasons used improved seeds	2537.5*** (725.9)	2379.9*** (710.4)	31.2 (175.5)
Program farmer: gender	-2651.3*** (686.1)	-1792.0 (1094.8)	-32.7 (249.2)
Constant	-3085.2 (2304.4)	17477.6 (15760.3)	12738.4*** (4309.7)
Observations	2459	2478	2518
Adjusted R ²	0.329	0.216	0.250

Cluster-robust standard errors in parentheses.

*p < .1, **p < .05, ***p < .01.

Not shown in table: Time and crop dummies, Mundlak instruments for fixed effects.

of the mid-line impact estimate reported in Table 3. From the graph, it appears as though most of the benefits of the program occurred during the 24 months during which farmers were actively enrolled in the program, then flattening out. That said, incomes remain at the higher level, suggesting that a temporary intervention that offers subsidies sticks and has lasting impacts as in the Carter et al. (2013) study of Mozambique.

Fig. 6 plots the estimated impacts on capital stock, together with a 90% confidence interval. As can be seen, the estimated impact of the program on beneficiaries' total capital stock increases significantly over the duration of the project, continuing to rise even after the end

of active programming (24 months). The predicted capital stock at the start of treatment is around \$8500, and by month 42 this has risen to \$12,500—implying an investment impact of than \$4000. This is again well in excess of the midline binary LATE estimates, which suggested impacts just under \$1000. While this 3.5 year impact on capital stock is large, it is broadly consistent with a stream of three estimated annual income increases on the order of \$1000 - \$2500.

As noted earlier, a substantial fraction of participant farmers are reported to be credit-constrained in the sense of having unmet demand for loans they would like to take. This suggests that many farm households would have had to self-finance investment out of their current

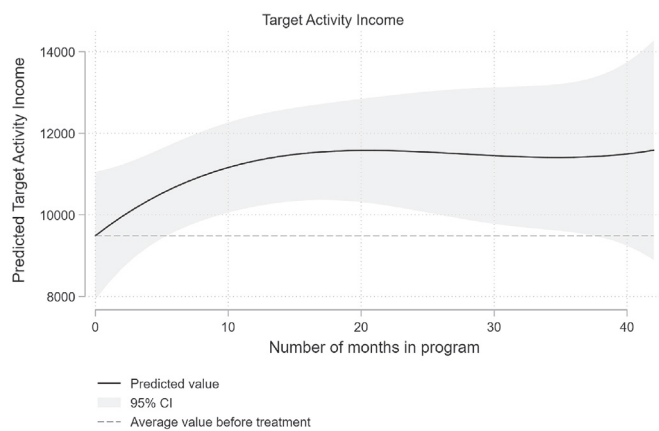


Fig. 5. Predicted target activity income by months of treatment.

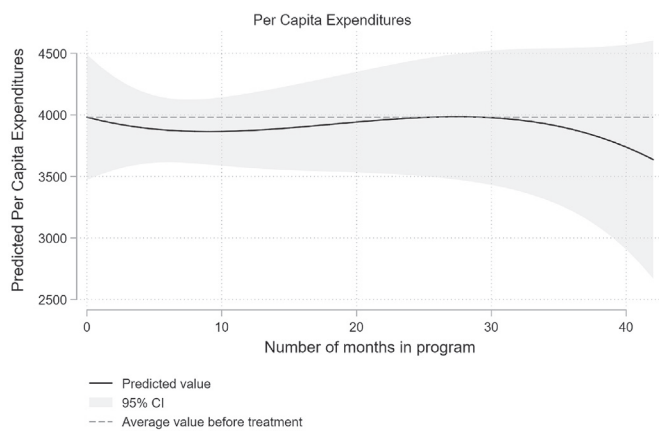


Fig. 7. Predicted household consumption by months of treatment.

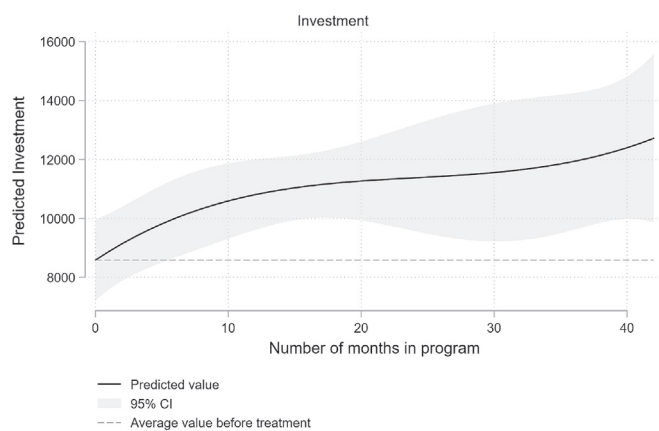


Fig. 6. Predicted capital stock by months of treatment.

income. Given that the estimated total increase in capital stock is similar to the accumulated annual income increases over the program, we are not entirely surprised that the longer-term pattern mimics the binary results, which showed small and statistically insignificant program impacts on household living standards.

Column 3 of Table 5 shows the estimated cubic function of treatment duration on household consumption. The individual coefficients are small in magnitude and not statistically significantly different from zero. The key question whether the overall impact duration relationship is statistically significant. Fig. 7 displays the cubic relationship as well as the 95% interval estimate of the duration response implied by the cubic estimates. As can be seen, the point estimates show no signs of consumption growth over the time of the program, and the interval estimator always includes zero.

In summary, we see evidence of a program that on average boosted incomes and that most, if not all, of that increased income was plowed into the accumulation of productive capital. While these average impacts are important in their own right, they do not reveal whether there is substantial heterogeneity in the impacts in terms of levels or in terms of how households allocate income increases between consumption and investment. The next section therefore looks more carefully at program impacts at different parts of the distribution.

4. Impact heterogeneity

In their study of a asset transfer and business development program in Bangladesh, Bandiera et al. (2017) find that program impacts on consumption and asset are four to ten times larger for upper quantile

households compared to lower quantile households. Banerjee et al. (2015) detect a similar pattern in their study, although the smaller impacts for lower quantiles are more uniformly significant. Reflecting on these findings, there are multiple reasons why programs like the RBD may have heterogeneous impacts, including:

1. *Heterogeneous access to the financial capital needed to make the most of an RBD intervention;*
2. *Complementarity between an RBD intervention and unobservable assets that are not equally distributed across the population, such as farming skills, learning capacity and business acumen; and,*
3. *Differential luck, with some succeeding and others failing for stochastic reasons.*

Earlier analysis conducted with only the mid-line data from this study revealed substantial evidence of impact heterogeneity. The program impacts only weakly influenced the well-being of the poorest-performing 50% of the population when compared against the poorest-performing segment of the untreated households. This initial analysis also suggested quite high returns to the best-performing segment of the treated group as compared with top performers in the then untreated control group (Toledo and Carter, 2010). In this section, we use all three rounds of data and our continuous treatment model to further explore impact heterogeneity in an effort to distinguish between mechanisms like item 3 versus mechanisms like 1 and 2, which would imply that the program does not work well (or as well) for certain types of households.

4.1. Econometric approach

Conventional regression methods (such as those just employed above in Section 3) estimate average or mean relationships. They assume that the vector of covariates affects only the location of the conditional distribution of y , not other aspects of y 's conditional distribution. Conditional quantile regression methods allow us to see whether the statistically average relationship is in fact a good description of the relationship in all parts of the distribution. Specifically, quantile regression allows us to recover the regression parameters that best describe the impacts on observations in different portions of the error distribution for our regression model.

Observations in the higher quantiles are those that “do better” than would be predicted by the observation’s level of treatment and other regression variables (e.g., are in the upper tail of the conditional distribution of the outcome variable). For simplicity, we refer to observations in the higher quantiles as “high performers,” but for now this should be interpreted to mean high-performing observations—not necessarily high-performing household types. Conversely, observations in the lower quantiles are those are in the lower tail of the conditional distribution of the outcome variable. Quantile regression allows us to

see if the marginal impact of RBD program participation at various parts of the conditional distribution of the outcome variables differs from the impacts at the mean—i.e. the average relationship estimated in Section 3.

Note that if the average regression model explains the data well, the impact estimates should be the same for all quantiles. However, if there is unobserved heterogeneity in the impacts, then the impact slopes across quantiles may be different. As mentioned above, there are conceptual reasons to suspect that the RBD program might have heterogeneous impacts. Reason 3 for heterogeneity above would imply “high-performing observations,” whereas reasons 1 and 2 would imply the existence of “high-performing households.”²⁰

To recover conditional quantile estimates, we employ the method developed by Abrevaya and Dahl (2008) that extends a correlated random-effects framework (like regression equation (3) above) to apply to conditional quantile models. While quantile models have been widely used in empirical studies since their development by Koenker and Bassett (1978), they are not often applied to panel data, likely because of the difficulty of differencing in the context of conditional quantiles. This problem arises because quantiles are not linear operators, so that the conditional quantile of a difference is not simply a difference of the conditional quantiles. Importantly, this methodology based on correlated random-effects preserves the fixed effects characteristics of the results, inoculating them against systematic or spurious correlation between the duration of treatment and initial and time-invariant conditions. Note also that the conditional errors are estimates of $v_i + \varepsilon_{it}$ from Eq. (3). That is, the error contains the time-invariant, random effect component.

4.2. Generalized quantile estimates

This section explores the heterogeneity of the impact or duration response function by estimating the conditional quantile functions for our preferred (cubic) parametric continuous treatment models. Parameter estimates for the Abrevaya and Dahl (2008) estimator can be obtained with any quantile regression package. Standard errors are obtained through bootstrapping (we use 500 replications), drawing households and clusters with replacement from the sample and estimating the variance-covariance matrix from the resulting empirical variance matrix. We present the results graphically, showing the predicted values of the outcome variables as a function of the length of time in the program, for the 25th, median and 75th quantiles, with bootstrapped 95% confidence intervals displayed as shaded areas around the point estimates. The full regression tables can be found in Appendix Table 3.

Fig. 8 (a) displays the results from the quantile analysis of income in the targeted activity. As can be seen, these estimates corroborate the hypothesis that program impacts are heterogeneous across the participant population. The impacts of the program are greatest at the high end of the distribution, with the 25th percentile impacts smaller in magnitude but still significantly different from zero. The high performers in the 75th quantile also experience a steeper impact response function than the lower quantiles. Indeed, towards the end of the program duration target activity incomes at the 75th conditional quantile are more than \$4500 greater than at the program beginning, more than three times the long-term impact level for the producer at the median or 25th quantile of the conditional income distribution. The lower quantile estimates are between \$600–\$1200. The statistical significance of these impact paths can be approximated by comparing the confidence

²⁰ If heterogeneity is driven by capital constraints (reason 1), then low performance could be attenuated by augmenting business services with a credit program. However, if low performers instead lack other types of characteristics such as human capital, it is less obvious how to ameliorate low performance.

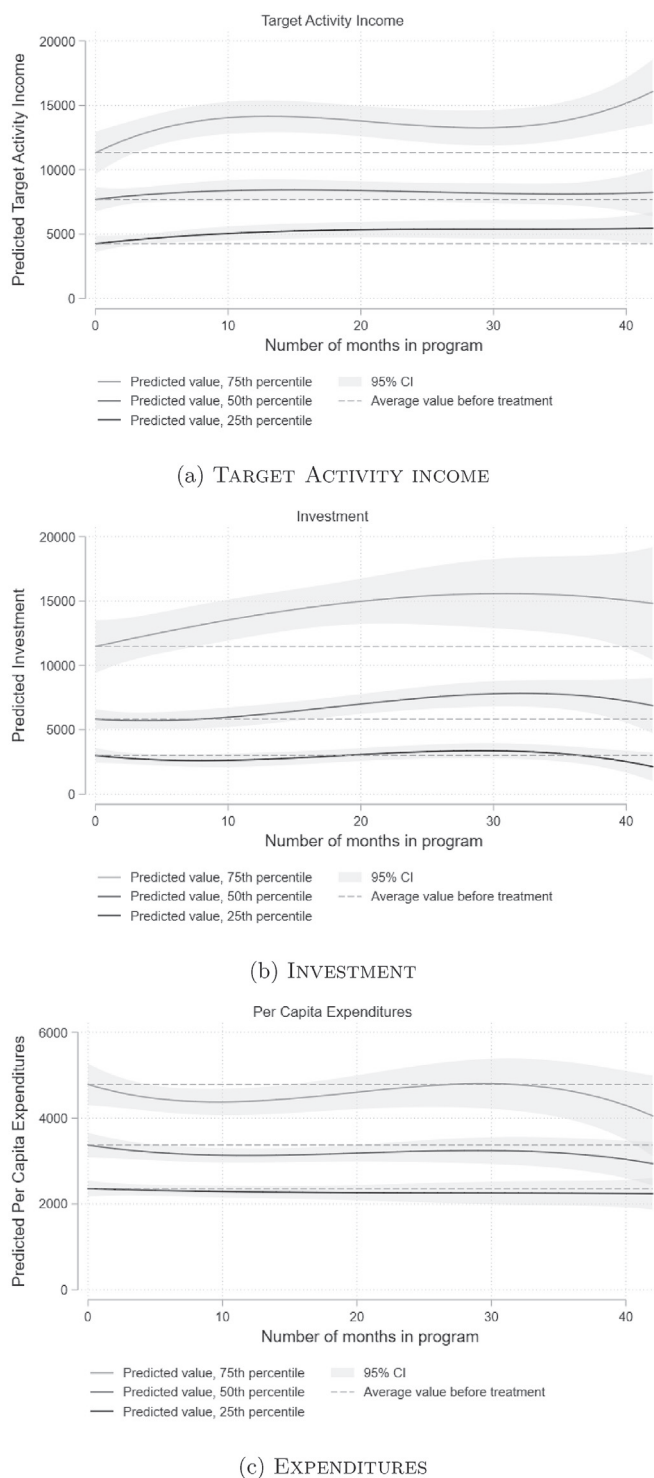


Fig. 8. Generalized quantile impact results.

interval to the dashed lines, which denotes the income level at zero months of treatment.

The estimated program impacts on capital investment also vary substantially across conditional quantiles (Fig. 8 (b)). The level and shape of the temporal impact path on investment increases as we move upwards in the conditional distribution of capital. For households in the 75th conditional quantile, investment increases by roughly \$3300 over the course of the program, a magnitude only slightly smaller than the target activity income increases seen in the panel above. At the median,

households also increase their investment by a substantial amount: roughly \$1900. The lowest quantile if anything displays negative or zero impacts on capital investment, and have no more capital stock at the end of the program than at baseline.

For per capita consumption, we see no positive impacts for any quantile, and for the median and 75th conditional quantile, the program impacts on consumption are negative for at least part of the treatment duration. The next section explores whether there exist household types who benefit from the program, or whether households move between conditional quantiles over time due to external factors like weather realizations or luck.

4.3. Are there household types who benefit from the RBD intervention?

It is tempting to interpret this impact heterogeneity as signaling that the RBD program did not work for everyone. However, as discussed above, it is possible that the lower quantiles are comprised of observations in which output was diminished by a negative shock. For example, a program like the RBD would be unlikely to have any impacts in the face of a localized drought since improved varieties, marketing channels, etc. would be useless if production dropped to zero due to weather events.

One way to gain purchase on this problem and to garner some insight on the source of this heterogeneity is to ask whether the same households consistently occupy the same quantile position in the conditional error distribution. If they do—meaning there are consistently upper quantile households and consistently lower quantile household—then we have evidence that program impacts vary systematically by (unobserved) household type.

To explore this idea, we recovered the residual for each observation in each round from a median regression. Denote by q_{it} the error quantile which contains household i 's residual round t . Using a standard analysis of variance decomposition, we can decompose the total variation in q_{it} as follows:

$$\sum_{i=1}^N \sum_{t=1}^3 (q_{it} - \bar{q})^2 = \sum_{i=1}^N \sum_{t=1}^3 (q_{it} - \bar{q}_i)^2 + 3 \sum_{i=1}^N (\bar{q}_i - \bar{q})^2,$$

where N is the number of households in the dataset, \bar{q} is the overall mean in the dataset, while \bar{q}_i is the mean quantile for household i over the 3 rounds of the data. The first term on the right hand side is the within sum of squares (WSS), while the second term is the between sum of squares (BSS). If no household changed position in the error distribution from year to year, then the WSS would be zero. Conversely, if a household's error quantile varied randomly from year to year (sometimes high, sometimes low and sometime in between), then $\bar{q}_i \approx \bar{q} \forall i$ and the BSS would be a small fraction of the total variation in q_{it} .

Table 6 shows that the fraction of total variation that is between households ranges from 67% to 84% for our three primary outcome variables. While a modest fraction of the variation comes from households moving between error quantiles over time, the bulk of the overall variation is coming from time-invariant differences between

households. In other words, there is evidence that particular households tend to occupy upper quantiles, and others tend to occupy lower quantiles. Given that the latter seem to enjoy little benefit from the RBD program across all three indicators, this finding suggests that the RBD program is a blunt instrument for improving the economic welfare of an important subset of household, but quite effective for those who have the right complementary market access and, or skills.²¹ The stability of households' position in the conditional distribution also suggests that it is worthwhile digging deeper into impact differences by exploring whether baseline variables can explain the treatment heterogeneity.

4.4. Identifying the characteristics of households that benefit more from the intervention

The heterogeneous impact results reported above are broadly similar to those reported in the Bandiera et al. (2017) study of an asset transfer and business development program in Bangladesh. Using binary quantile treatment effect analysis, the authors find that the impacts on living standards and asset accumulation are small and marginally significant for some 35–45% of the population despite being large and robust on average. Similar to our quantile regression results, these findings are provocative but not completely satisfying since they say nothing about the factors that distinguish those who benefited from the intervention from those who did not. If we could identify observable characteristics that predict which households are likely to benefit from entrepreneurial interventions, we could leverage such results by targeting interventions better,²² or by identifying other limiting constraints that a modified intervention might relax, allowing more broadly shared impacts.

In an effort to better understand the sources of impact heterogeneity, this section proceeds in three steps. First, we employ the generalized random forest (GRF) method of Athey et al. (2018)—a non-parametric machine learning algorithm—to identify which characteristics are associated with greater RBD program impacts. Although our sample size pushes the limits of what is feasible using GRFs, we find some evidence that households who are initially disadvantaged along multiple dimensions benefit the most from the intervention. Second, we incorporate an indicator of “initial disadvantage” into our continuous treatment regression model to see if this new measure explains differential impacts and/or reduces the impact heterogeneity across conditional quantiles. Third, we repeat the approach developed in section 4.3 to ask if the residual quantile variation is largely comprised of variation within or between households. In other words, once we control for the observable factors that drive heterogeneity, do other unobserved household characteristics such as entrepreneurial zeal explain impact heterogeneity?

²¹ Efforts reported in Toledo (2011) to unpack the reasons behind the impact heterogeneity reported in Toledo and Carter (2010) are only partially satisfying. That analysis focused on explanation (1) above, categorizing households based on their credit-rationing status. While credit market status is of course endogenous, that analysis revealed no simple relationship between performance and contemporaneous credit rationing status. Indeed, the only factor uncovered was past credit history. RBD impacts on farms with prior credit history appeared quite large and significant. Unfortunately, the interpretation of prior credit history as a factor explaining heterogeneous program impacts is ambiguous. It seems most likely that those with past credit histories are actually those with higher levels of farming and business acumen (pointing toward explanation 2 above). It may also be that those acumen levels were themselves endogenously produced by prior access to credit and business opportunities.

²² For interventions that are relatively costly, including both the Nicaragua RBD product as well as graduation programs such as that studied by Bandiera et al. (2017), the returns to more effective targeting are potentially large.

Table 6

Within vs. Between variation in error percentiles.

Panel A: Core Continuous Treatment Model		
	Within variation (%)	Between variation (%)
Income	33	67.0
Total Investment	16.3	83.7
Consumption	27.1	72.9
Panel B: GRF-informed Continuous Treatment Model		
Income	33.1	66.9
Total Investment	16.5	83.5
Consumption	27.2	72.8

4.4.1. Generalized random forest analysis of impact heterogeneity

As detailed in Appendix E, we apply GRF methods to our midline data to see how a number of covariates shape the impact of the RBD program on income. In particular, we include an estimate of a farmer’s technical efficiency,²³ program farmer years of education, owned land assets, program farmer experience with the target crop, and initial farm capital. Fig. 9 summarizes the results of this analysis. Each row in the figure looks in detail at one of these covariates, with the graphs showing predicted treatment effects for the level of the baseline covariate indicated on the horizontal axis. Each column in the figure holds the other covariates at different percentiles: column 1 holds them at the 25th percentile, column 2 holds everything at the median in the sample, and the third column holds covariates at the 75th percentile.

A somewhat consistent pattern emerges: the first column shows more precisely estimated and generally larger impacts. Furthermore, impacts are also systematically larger whenever the covariate being varied along the x-axis is at or below its median value. For example, the first graph in row 3 shows that predicted treatment effects for households in the lowest 25% of the distribution of experience, education, initial capital and technical efficiency is downward sloping in the baseline amount of land owned. The treatment effect then become insignificant at the 55th percentile of the land ownership distribution. While the multiple splitting of the data used by GRF strains our data, they do seem to indicate that the program has larger and more consistent—i.e., less variable—impacts on income for farm households that suffer multiple disadvantages. In other words, farmers who have lower initial land assets, lower levels of technical efficiency, and lower levels of education and experience may benefit the most from the program.

This finding is particularly interesting given that the RBD program established various asset ownership minimums, below which a farm household was not eligible for the program (see Appendix Table1). We turn now to see the extent to which initial disadvantage explains the overall impact heterogeneity identified earlier.

4.4.2. The GRF-informed continuous treatment model

Based on the results from the supervised learning algorithm above, we define a new variable called “disadvantaged.” We classify households as disadvantaged if they owned below the median amount of land at baseline and had less than the median years of experience in the target crop.²⁴ We then incorporate this new “disadvantaged” indicator variable into the continuous treatment model 4. The regression results are displayed in Appendix Table 6.

The disadvantaged dummy on its own is negative and statistically significant for investment and household expenditures; its interactions with the treatment duration variables are not individually statistically significant. Graphing these regression results for disadvantage and non-disadvantaged households (in the spirit of Figs. 5–7; graphs available upon request) shows that the disadvantaged experienced more rapid income gains than the non-disadvantaged in the first year of treatment but that the income impacts appear to flatten out somewhat once they are no longer actively enrolled in the program.

The most striking difference between the disadvantaged and their more-advantaged counterparts are the impacts on investment. Drawing out the implications over the months, we observe that the investment impacts of the RBD program are entirely concentrated among those farmers above the median experience and land ownership at baseline.

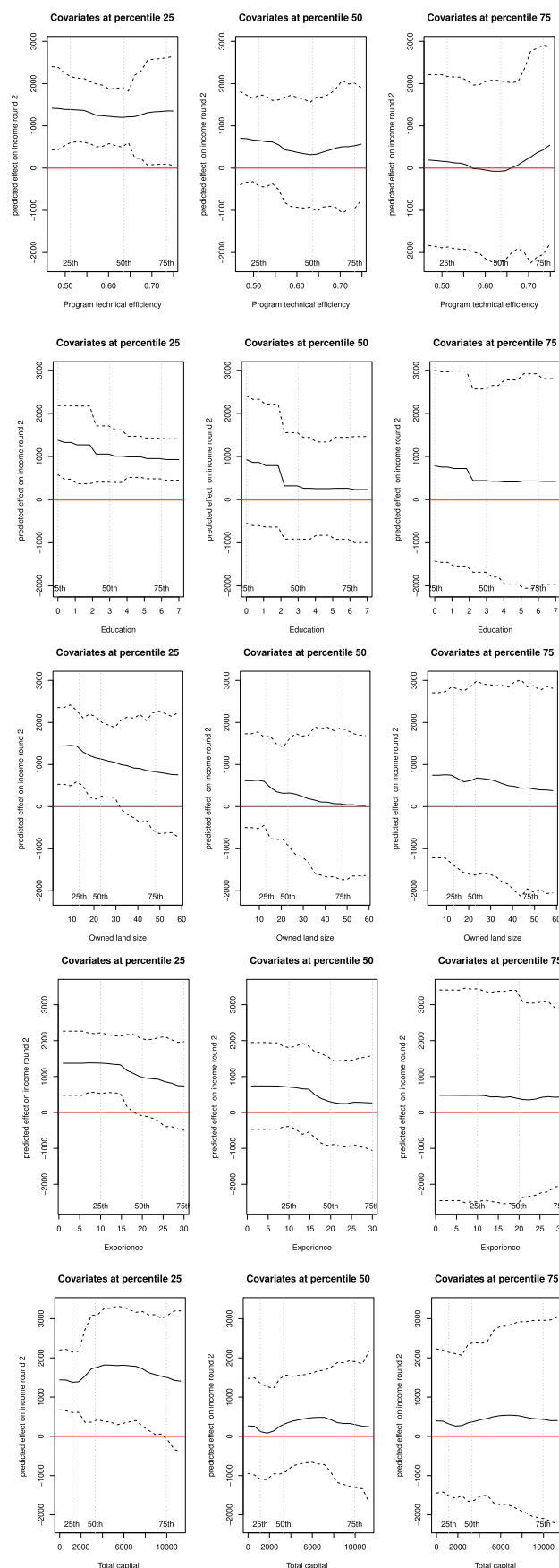


Fig. 9. GRF conditional treatment effects.

²³ Technical efficiency measures calculated using stochastic frontier methods were kindly provided by Malacarne et al. (2017). Full details on the estimation procedure available from the authors.

²⁴ We chose the cutoffs defining these indicator variables based on the causal tree analysis.

While differences in estimated consumption impacts are not statistically significant, the investment results hint at a pattern in which the initially disadvantaged may have allocated their income increases more toward consumption and less towards investment than do their initially-advantaged neighbors.

Finally, we repeat the quantile heterogeneity analysis for the GRF-augmented continuous regression model. Despite the fact that the disadvantaged variable explains some of the variation in outcome, substantial heterogeneity remains. Moreover, as seen in Panel B of Table 6, most of that heterogeneity remains between households. In other words, while the GRF analysis located those observable factors that best explain differences in impacts, there remains some as yet unidentified (and unmeasured) characteristics of high-performing households that complements and boosts the impacts of the RBD program. These may be technical skills, business acumen, entrepreneurial ability, or something else, but in the end we cannot identify them beyond saying that there are time-invariant, household specific attributes.

5. Conclusion

Nicaragua's Rural Business Development Program (RBD) was a 24-month intervention designed to boost the productivity and incomes of a largely poor and rural farming population by enhancing their business knowledge and improving their access to markets and technologies. The program also offered temporary input and investment subsidies. Gauging the effectiveness of a program like the RBD that addresses the rural poor as incipient entrepreneurs rather than as passive recipients of transfers, faces several challenges. The first challenge is to gauge the long-term impacts of the RBD intervention. Beneficiaries may continue to learn, invest and realize further benefit from the intervention long after the 24 month period. It is also possible that any immediate impacts are not sustained if beneficiaries return to their prior status after the period of direct intervention ends. The second challenge is to understand the heterogeneity of impact across the target population, both in terms of participation in a somewhat complex program, and in terms of the business success of those who choose to participate.

To address these challenges, we employed a 5-year roll-out design that randomized beneficiaries' exposure to the RBD program. In addition to standard treatment effect estimates, the design allowed us to identify a two-sided complier sample and to focus the analysis only on those who (eventually) enrolled in the program. Using 3 rounds of data from this design, we find that on average has substantial impacts on income in the targeted activities (\$1200 to \$2100 annually)²⁵ and on agricultural investment (\$1000 to \$3700). Somewhat surprisingly, there are virtually no impacts on household consumption expenditures, and in fact some weak evidence that the program *reduced* expenditures, as might be expected if beneficiaries were investing more in their farm but facing liquidity constraints.

The RBD program did not include a direct credit market intervention. The overall MCC program in Nicaragua operated in part on the theory that improved property registration would indirectly improve smallholder access to capital by increasing their collateral and credit-worthiness to the extant banking sector. Whether or not that strategy would have worked remains an open question, as the property registration component of the program was eliminated in early 2009 (see footnote 6 above).

At a direct program cost of \$2500 per-farmer enrolled in the program, these average estimates indicate that the RBD was a cost-effective instrument for boosting the average income and assets of its beneficiary farmers. However, its effectiveness as an instrument to

address rural poverty depends on the distribution of impacts across the program's overall target population. Looking at the full distribution of impacts is especially important for efforts like the RBD program that target beneficiaries' income-generating and entrepreneurial capacities.

Drawing out the implications of our results, we first note that just over one-third of the target population declined to participate in the program. The one-third who did not participate had modestly lower living standards at baseline.²⁶ Second, our fixed-effects analogue conditional quantile regression methods explore the degree to which the average pattern of impact faithfully reflects the full distribution. We find evidence of significant heterogeneity in impacts on program income and investment, with generally smaller program impacts in the lower quantiles of the conditional error distribution. Observations in the 25th quantile show long-term income and investment impacts that are one-third to one-half the size of the estimated average impacts, but still significant and positive.²⁷ We further provide evidence that households' positions in the error distribution are relatively constant over time, implying that lower-quantile observations are comprised of "low-performer" type households for whom RBD program impacts are less effective.

Efforts to identify the source of this heterogeneity using Generalized Random Forests was only partially successful. The GRF analysis does point toward larger impacts residing with farmers who were initially disadvantaged in the sense of having less land, lower amounts of capital, less experience with the target crop, and lower technical efficiency in agriculture. At a minimum, these results imply that programs like the RBD program should exercise caution when excluding farmers whose resources are thought to be too modest, as it may be exactly these households that have the most to gain from such interventions.²⁸ But even after controlling for the component of heterogeneity that relates to initial disadvantage, we find that program impacts remain highly heterogeneous and that there is some unobserved household characteristic (entrepreneurial zeal?) that separates high from low performing households.

In the end, the existence of these two groups that did not benefit from the RBD program (those that did not participate, and those that experienced modest impacts even when participating) serves as a useful reminder that not all small farms can upgrade and succeed. If the goal is to eliminate rural poverty, then this limitation needs to be kept in mind as other interventions may be needed to improve prospects for this sub-population and their children. Looking forward, it may be that next-generation RBD programs can reduce the size of this minority, perhaps by incorporating elements of the psychological asset building found in graduation programs studied by Banerjee et al. (2015) and Bandiera et al. (2017).

²⁶ Recent work by Macours and Vakis (2008) on poverty and aspirations suggest that there may be some individuals who could benefit from interventions such as the RBD, but that they need smaller, confidence, and aspiration-building steps before they are willing to jump into a more forward-looking and entrepreneurial profile.

²⁷ Importantly, there is no evidence that households with fewer assets—i.e. those closest to the asset minima that defined program eligibility—benefitted less from the program. To the contrary, our analysis suggests that households that may be near the asset minima might be the ones who benefit the most. While the asset floors and ceilings used to establish RBD eligibility were based on best-practice intuition, it is clear from a targeting perspective that more work needs to be done to see if there is such a thing as a farm that is too small to benefit from this kind of intervention.

²⁸ The tendency to exclude less well-resourced farmers from this class of development programs appears to come from the belief that such farmers are not "commercial" and therefore will not purchase expensive inputs or invest in their farm operations.

²⁵ As discussed above, these estimates are upper bound estimates on the impacts on total family income.

Acknowledgments

We thank Anne Rothbaum, Lola Hermosillo, Jack Molyneaux, Juan Sebastian Chamorro, Carmen Salgado, Claudia Panagua, Sonia Agurto and the staff at FIDEG, Conner Mullally, and seminar participants at NEUDC, University of the Andes, University of Wisconsin - Madison, University of California, Davis, and the University of California, Berke-

ley for helpful comments. We also thank two anonymous referees for suggestions that helped strengthen the paper. We gratefully acknowledge funding from the Millennium Challenge Corporation, as well as financial support from the US Agency for International Development Cooperative Agreement No. EDH-A-00-06-0003-00 through the BASIS Assets and Market Access CRSP.

Appendix A. Eligibility criteria by productive activity

Appendix Table 1

Eligibility criteria used to identify farmers in target activities.

	Sesame	Beans	Vegetables	Cassava	Livestock
Asset Floor ^a	7 ha	3.5 ha	1.4 ha	3.5 ha	10 mature cows
Asset Ceiling	35.2 ha	35.2 ha	14.1 ha	70.4 ha	100 mature cows
Prior Experience	1.4 ha in sesame	0.7 ha in beans	Some vegetable production	1.4 ha in cassava	Developed livestock activity
Water	–	–	On-farm water source	–	On-farm water source
Legal Status	Farmer has land title or is in possession of land				
Age	Farmer must be at least 20 years old				
Environment	Land located outside of national protected areas				

^a Minimum farm size reduced when farm is irrigated.

Appendix B. Probability of compliance

Appendix Table 2

Decision to take up program, by treatment group status: probit regression.

	(1)	(2)	(3)
Early treatment	–0.072 (0.10)	–0.059 (0.11)	–0.0033 (0.95)
Technical efficiency		0.38 (0.37)	–0.26 (0.48)
Program farmer: age (years)		–0.0049 (0.0072)	–0.0066 (0.0089)
Program farmer: education (years)		0.020 (0.015)	0.0014 (0.032)
Program farmer: gender (=1 for female)		0.086 (0.18)	0.24 (0.24)
Household members		0.036 (0.027)	0.046 (0.031)
Per capita expenditures (PPP\$)		–0.000050 (0.000043)	–0.000025 (0.000082)
Credit constrained (=1 if constrained)		–0.071 (0.11)	0.18 (0.19)
Total capital		0.0000064 (0.0000091)	–0.0000049 (0.000022)
Program farmer: experience (years)		0.0018 (0.0057)	0.0030 (0.0081)
Landholdings: owned (manzanas)		–0.0015 (0.0042)	–0.0083 (0.0067)
Landholdings: amt. planted, target crop		0.020 (0.026)	0.17*** (0.048)
Landholdings: amt. planted in maize		–0.033 (0.024)	–0.076* (0.046)
Target activity income (PPP\$)		0.0000077 (0.000027)	–0.000070* (0.000041)
Share of seasons used improved seeds		0.043 (0.17)	0.052 (0.29)
Early x technical efficiency			1.00 (0.64)
Early x Age			0.0019 (0.014)
Early x Education			0.034 (0.041)
Early x Female			–0.10 (0.36)
Early x Household size			0.0029 (0.052)
Early x Expenditures			–0.000025 (0.000099)

(continued on next page)

Appendix Table 2 (continued)

	(1)	(2)	(3)
Early x Credit constrained			-0.36 (0.24)
Early x Total capital			0.000020 (0.000026)
Early x Experience			-0.0022 (0.011)
Early x Land owned			0.011 (0.0080)
Early x Land in target crop			-0.18*** (0.050)
Early x Land in corn			0.047 (0.055)
Early x Target activity income			0.000092* (0.000050)
Early x Improved seed			-0.12 (0.34)
Crop fixed effects?	NO	YES	YES
N	1396	564	564

Probit marginal effects; cluster-robust standard errors in parentheses.

Appendix C. Quantile regression tables

Appendix Table 3
Quantile regression coefficients.

Quantile:	Income			Investment			Consumption		
	25 th	50 th	75 th	25 th	50 th	75 th	25 th	50 th	75 th
Months treated	113.5 (86.8)	117.1 (134.8)	511.1*** (163.8)	-113.3 (74.8)	-68.1 (101.5)	228.0 (208.3)	-10.0 (21.8)	-47.1* (27.2)	-95.4* (50.5)
Months treated ²	-3.79 (5.22)	-5.62 (8.04)	-28.4*** (9.73)	9.12* (4.66)	10.0 (6.62)	-1.78 (13.7)	0.34 (1.34)	2.80* (1.61)	6.55** (2.93)
Months treated ³	0.042 (0.087)	0.075 (0.13)	0.45*** (0.16)	-0.16** (0.078)	-0.19 (0.12)	-0.042 (0.23)	-0.0039 (0.022)	-0.046* (0.027)	-0.11** (0.049)
Program farmer: education	56.5** (22.1)	117.1*** (36.8)	222.5*** (71.3)	82.6** (38.8)	348.0*** (63.2)	621.8*** (120.7)	107.2*** (13.6)	141.0*** (14.8)	197.3*** (23.2)
Program farmer: experience	15.1*** (5.59)	10.8 (7.83)	16.3 (15.6)	6.55 (6.20)	28.0* (15.6)	56.6 (36.3)	8.56*** (3.16)	13.9*** (3.68)	19.0*** (6.60)
Household members	-30.2 (27.9)	-32.7 (43.1)	-74.2 (71.5)	8.65 (32.9)	3.12 (56.3)	-13.8 (137.6)	-183.3*** (20.2)	-276.4*** (20.7)	-364.1*** (24.7)
Landholdings: owned	20.2** (8.47)	43.2*** (9.00)	88.9*** (23.3)	34.8*** (7.38)	53.7*** (10.6)	99.9*** (24.4)	2.06 (1.94)	6.27** (2.44)	13.3*** (4.66)
Share seasons used improved seeds	486.5*** (163.0)	645.6*** (250.6)	1248.3*** (471.6)	412.3** (177.8)	1042.5*** (308.6)	1795.1* (975.2)	166.5** (72.7)	250.6*** (88.6)	192.1 (143.7)
Program farmer: gender	-493.2** (212.8)	-757.4*** (257.4)	-1184.0*** (409.2)	-334.8 (246.2)	-327.8 (524.7)	-1516.8 (937.3)	-4.39 (102.2)	-67.4 (147.2)	171.4 (252.6)
Constant	5243.4** (2546.1)	2223.8 (2968.7)	-562.4 (5524.1)	-349.2 (12933.7)	15438.0 (20757.7)	18662.8 (38793.2)	3062.7 (3889.5)	9073.0 (7114.6)	23869.3*** (8670.5)
N	2459	2459	2459	2478	2478	2478	2518	2518	2518

Bootstrapped standard errors in parentheses, from 500 reps drawn with replacement over households.

The regressions also include time fixed effects, activity fixed effects and averages of the time-varying variables.

*p < .10, **p < .05, ***p < .01.

Appendix D. Hypothesis testing

To test whether the differences that we observe between the different quantiles are statistically significant, we employ the minimum-distance framework in Abrevaya and Dahl (extended from Buchinsky’s (1998) framework to the panel data context) to test the equality of the parametric duration response variables’ effects across quantiles. Since both *months*, *months*², and *months*³ enter into our preferred cubic model, the relevant test is a joint test of equality. In other words, the null hypothesis is

$$H_0 : \zeta_{1,\tau_1} = \zeta_{1,\tau_2} = \zeta_{1,\tau_3} \wedge \zeta_{2,\tau_1} = \zeta_{2,\tau_2} = \zeta_{2,\tau_3} \wedge \zeta_{3,\tau_1} = \zeta_{3,\tau_2} = \zeta_{3,\tau_3},$$

where ζ_1, ζ_2 and ζ_3 are the estimated coefficients on *months*, *months*², and *months*³, respectively, and τ_1, τ_2 and τ_3 are the different estimated quantiles (25th, 50th and 75th).

In following the Abrevaya and Dahl (2008) testing framework, the only changes we make are to allow for an additional round of data and the fact that we include averages of the time-varying regressors, instead of their value in each round. The minimum-distance test statistic has a limiting chi-square distribution, with degrees of freedom equal to the number of restrictions (in our case 6). These test statistics and their associated p-values are shown in Table 4. As the table shows, the estimated effect of months in the program on capital and consumption both vary significantly across the quantiles, but the evidence for income is less precise.

Appendix Table 4
Tests of marginal-effect equality across quantiles.

Outcome variable	χ^2_6 -statistic	p-value
Income	8.41	0.21
Total capital	15.69	0.016
Per-capita consumption	11.62	0.071

For each outcome variable, the p-values reported are for the null hypothesis of joint equality of the marginal effects of the variables *months*, *months*², and *months*³ for the quantiles .25, .50 and .75. Results are based on 300 bootstrap replications, block-bootstrapped over household ID and cluster variable.

Appendix E. Causal random forest analysis of heterogeneity

This appendix lays out the random forest methodology used to motivate the heterogeneity analysis in Section 4.4. This discussion draws both upon Athey et al. (2018) and Davis and Heller (2017). Broadly speaking, a regression random forest is a collection of classification trees that are built using data and a statistical learning algorithm. Each classification tree is created by randomly drawing a subset of observations (y_i, x_i) , called the *training sample*, of size $s < n$. Denote this training sample by J . The training sample is drawn without replacement from the full sample of size n , and forms the basis for a tree. The training sample is recursively split into subgroups along values of some covariates, and we can then predict an individual’s outcome, y_i , using the mean \bar{y} of observations that share similar covariates, x . In turn, the average of these predictions across many trees is an estimate of the conditional mean of y . The algorithm determines which splits that are implemented based on which minimize some in-sample goodness-of-fit criterion.

The Generalized Random Forest (GRF) method developed by Athey et al. (2018) improves upon random forest methods by proving that forest estimators are consistent and have an asymptotic normal distribution with a variance that can be estimated. In the context of heterogeneous treatment effects, the GRF algorithm can be used to build a causal random forest (CRF) to estimate the following conditional average treatment effect (CATE):

$$\tau(X) = E[Y(W = 1) - Y(W = 0) | X = x] \tag{5}$$

where Y is the outcome variable, and $W = 1$ indicates treatment, which is assumed to be independent of unobservable variables conditional on the observable covariates, X . The set of covariates X are farmer and farm characteristics in our case, but more generally can be any covariates that might produce heterogeneous treatment effects.

Each training sample, J , is further divided into two new subsets, J_1 and J_2 . The sub sample J_1 is used to build the tree whereas J_2 is used with the ensemble of trees (i.e., the *forest*) to calculate weights. These weights then get used to minimize a weighted local function, in order to estimate the conditional treatment effect. These weights indicate how important each observation is in estimating the conditional treatment effect. Below, we give an example of how the method allows us to estimate a conditional treatment effect by describing the process in terms of these two training sub-samples.²⁹

We use the GRF algorithm to build a CRF as a pre-regression analysis. Given our relatively limited sample size, and that random forest methods seem to work best with large samples (Davis and Heller, 2017), we use this method to identify which of our covariates are associated with heterogeneous treatment effects on farmers’ income. We focus on income as it was the primary focus of the program; any effects on consumption or investment ought to follow from increases in income, and we therefore assume that impact heterogeneity in investment or consumption ought to follow similar patterns to treatment heterogeneity in income.

Using the training sample J_1 to build an ensemble of causal trees

The sub-sample J_1 is used to build the *nodes* and the *leaves* of a tree. To create a node, m covariates are randomly selected by the algorithm, in order to determine which of them is the best splitting variable. Two main steps determine the “best” splitting covariate:

STEP 1

The observations that belong to potential parent node P are used to calculate $\hat{\tau}_P(X)$, an estimate of the conditional treatment effect $\tau(X)$ described in Eq. (5). $\hat{\tau}_P(X)$ is calculated by solving the following moment condition, which is equivalent to the least-square estimator condition:

$$\sum_{i \in P} \tilde{w}_i (\tilde{y}_i - \tau_P(x) \tilde{w}_i) = 0 \tag{6}$$

where \tilde{w}_i and \tilde{y}_i represents the values of Y and W in deviations with respect to their corresponding sample mean in the parent node P .

The estimator $\hat{\tau}_P(X)$ is used to calculate pseudo-outcomes for each observation in the node P ,

$$\rho_i = \tilde{w}_i (\tilde{y}_i - \hat{\tau}_P(x) \tilde{w}_i) \left(\frac{\sum_{i \in P} \tilde{w}_i^2}{\#\{i \in P\}} \right)^{-1} \tag{7}$$

Large values of ρ (in absolute value) for a subset of observations would suggest that a partition of the sample could identify different values for $\hat{\tau}$.

STEP 2

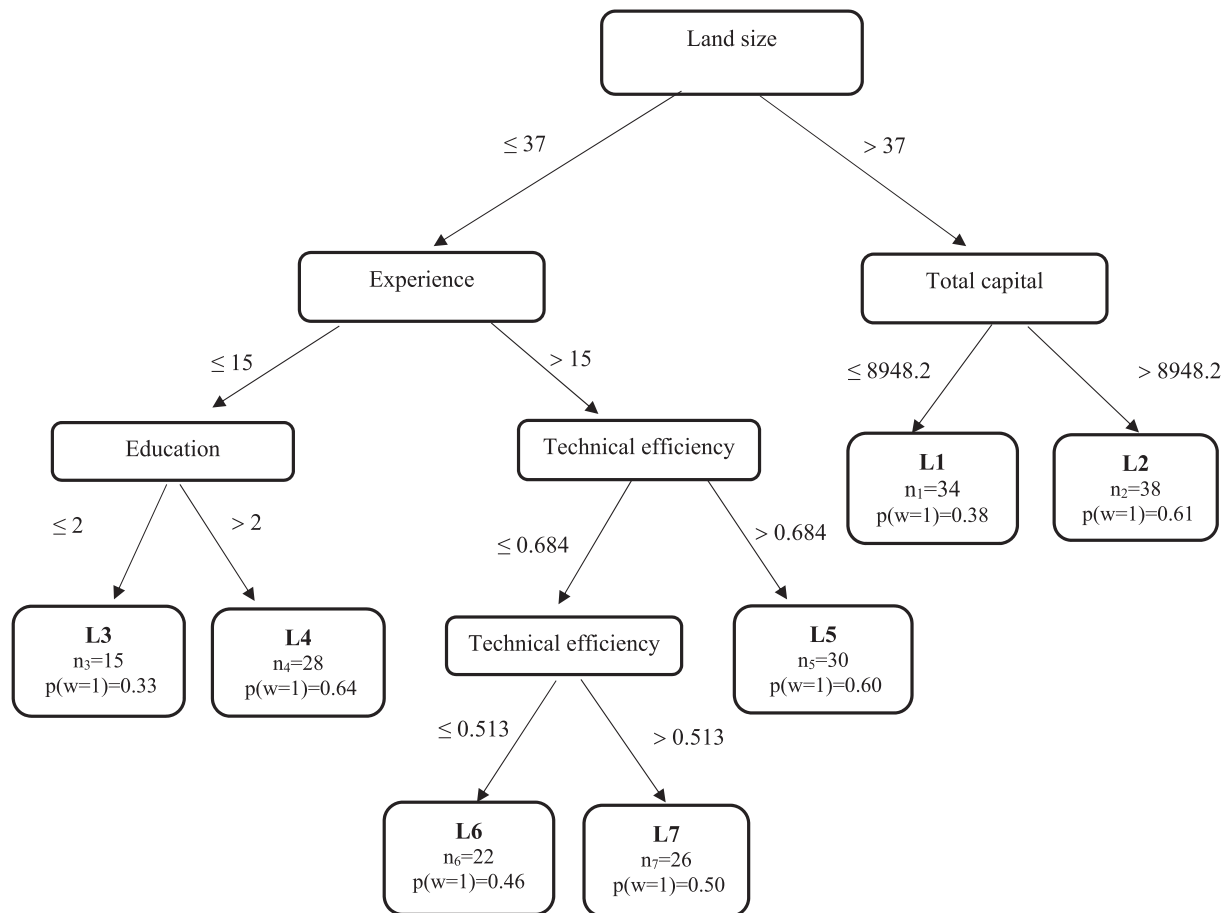
These pseudo-outcomes are then used in the second stage where the nodes are divided based on the values of a given covariate. Denote the divided nodes C_1 and C_2 . The “best” splitting covariate is chosen based on maximizing the following criterion:

²⁹ This description is specific and less technical than that described in Athey et al. (2018); we discuss causal treatment effects without confounding. See the referenced paper for a more technical explanation of the algorithm and for other applications.

$$\tilde{\Delta}(C_1, C_2) = \sum_{j=1}^2 \frac{1}{\#\{i \in C_j\}} \left(\sum_{i \in C_j} \rho_i \right)^2 \tag{8}$$

This splitting algorithm is repeated at each child to create new partitions, when it is possible.

Fig. 1 shows an example of one the 5000 built trees of our CRF. We use the R function *causalforest* to build a CRF for round-2 (midline) target activity income in the complier sample ($n = 772$). We use six covariates that encapsulate key baseline characteristics: technical efficiency in the target crop, land ownership, farmer experience in the target crop, education, credit constraint status (0/1), and total capital. We use the cross-validation option to tune the following parameters of the forest: the number of covariates tried in each split (5), the maximum allowed imbalance in a split (0.051),³⁰ and a penalty factor for imbalanced splits (0.796).³¹ We randomly draw 50 percent of the data to build each tree. Given the small sample size of the subsamples ($n_{J_1} = n_{J_2} = 193$), we set the minimum number of observations per leaf equal to 10.



Appendix Fig. 1. Example of a tree using the GRF algorithm.

Following Athey et al. (2018) we transform the outcome variable (target activity income) and the treatment variables to improve the performance of the forest predictions in a context of confounding and heterogeneity.³² The example tree in Fig. 1 has seven leaves (indexed by l) meaning that the training sample J_1 has been partitioned in seven subsamples based on the values of five covariates. The n_l values at the end of each leaf represents the number of observations from the training sample J_1 ($n_{J_1} = 193$) that falls in that leaf.³³ The first split divides the sample into those observations with more than 37 *manzanas* land, and those with less. The figures in the square brackets beneath show the unweighted treatment impact of the program on observations in each leaf. Partitions are chosen to maximize the difference between these two values.

Note that trees do not have the same number of leaves and the same number of covariates, but 91% of the trees have 6 or 7 leaves. Table 5 shows the percentage of “importance” of each covariate in the forest in terms of the frequency with which a variable is used as a splitting variable in the

³⁰ Imbalance in splits is not only with respect to the number of observations but also with respect to the treatment variable.

³¹ The cross-validation was done using 500 trees and 50 forests.

³² The transformation subtracts the marginal effect of X on Y and the marginal effect of X on W from Y and W , respectively. To do that, a generalized random forest for Y and X and another forest for W and X are estimated. The forests are used to make predictions for each observation using the subsample that was not selected for training (out-of-bag training sample). The objective of this transformation is to look for heterogeneity without the effect of the covariates on Y and W . We also tuned the parameters of these forests.

³³ For reference we also include the proportion of treated observations ($p(w = 1)$). As we explain below, these specific numbers are not directly involved in the calculation of the conditional treatment effect.

forest.³⁴ Technical efficiency, land size and total capital are the most important variables, while credit-constraint status, experience, and education are less important.

Appendix Table 5
Frequency of covariates as splitting variables in the GRF.

Technical Efficiency	Education	Land Size	Experience	Credit-constrained	Total Capital
22%	12%	24%	12%	6.5%	23%

Using the training sample J_2 and the tree ensemble to predict conditional treatment effects

Having generated an ensemble of trees, we use the forest to create treatment effect predictions conditional on specific values of the covariates. For example, if we are conditioning on x_p , then in the b^{th} tree we can calculate a tree weight $\alpha_{bi}(x_p)$ for the i^{th} observation of the training sample:

$$\alpha_{bi}(x_p) = \frac{\mathbf{1}(\{x_i \in \text{leaf}_b(x_p)\})}{(\#x_i \in \text{leaf}_b(x_p))}$$

We use the example tree (ET) of Fig. 1 to illustrate how to calculate the weights for this tree. Denote by “median farmer” a fictional observation that has the median values for each of the variables used in the forest. We can describe this median farmer by a vector: $x_m = (0.636, 3, 23, 20, 0, 3318.7)$ describes a farmer with median baseline values of (i) technical efficiency, (ii) education, (iii) land size, (iv) experience, (v) credit constraint status (with 0 denoting that the farmer is not credit constrained), and (vi) total capital. This *median farmer* falls in leaf seven (L7 in Fig. 1). In total, 27 observations from subsample J_2 (i.e., not the training sample) fall in the same leaf as x_m . Therefore, the weight for each of these observations is $\alpha_{ET,i}(x_m) = \frac{1}{26}$, while the weight is zero for the observations that do not fall in the leaf of x_m .

We then repeat the tree weight calculation for each of the B trees of a forest. Thus, the final forest weight for the i^{th} observation, $\alpha_i(x_p)$, is the average of all its tree weights. That is,

$$\alpha_i(x_p) = \frac{\sum_{b=1}^B \alpha_{bi}(x_p)}{B}$$

Observations with larger weights are more similar to x_p because they often fall in the same leaf as x_p . These weights are then used to estimate $\tau(x_p)$, the treatment effect conditional on x_p , by solving the following weighted moment equation:

$$\sum_i^n \alpha_i(x_p) \psi_{\tau(x_p)}(Y, W) = \sum_i^n \alpha_i(x_p) [w_i - \bar{w}_\alpha] [(y_i - \bar{y}_\alpha) - \tau(x_p)(w_i - \bar{w}_\alpha)] = 0 \tag{9}$$

where $\bar{w}_\alpha = \sum_{i=1}^n \alpha_i(x_p) w_i$ and $\bar{y}_\alpha = \sum_{i=1}^n \alpha_i(x_p) y_i$. It can be seen that the solution equation for $\hat{\tau}(x_p)$ is similar to a weighted least squares estimator. Athey et al. (2018) show that the asymptotic variance of $\hat{\tau}(x_p)$ can be approximated by estimating the following expression³⁵:

$$\text{Var}(\hat{\tau}(x_p)) = \frac{H_n(x_p)}{E(W^2 | X = x_p) - E(W | X = x_p)^2}$$

where $H_n(x_p) = \text{Var} \left[\sum_{i=1}^n \alpha_i(x_p) \psi_{\tau(x_p)}(Y, W) \right]$. The expected values of the denominator are estimated by GRFs. To estimate the numerator, we draw 2500 bootstrapped samples of size $\frac{n}{2}$. The estimator of $H_n(x_p)$ is calculated as a between-groups variance where the trees of each group are constructed using the same bootstrapped sample. We build two trees per group.

To continue with the example of the median farmer, there are 772 forest weights for x_m that can be used to estimate $\hat{\tau}(x_m)$. As a result, the predicted treatment effect on income for this median farmer is \$318.3 with a 90% confidence interval of [-916.2, 1552.9]. Similarly, predictions of the treatment effect are presented in Fig. 9 for different sets of covariates values. In the first column of each graph, we fix the other variables at value of the 25th percentile, making the predictor vector $x_{p25} = (0.512, 0, 13.3, 10, 1, 1146.8)$. The median farmer used in the example is in the second column, and the third column presents results with other variables fixed at the 75th percentile. We estimate treatment effects at different values of the covariate displayed at the horizontal axis, holding fixed other covariate values at the percentile shown at the top of the graph. Vertical lines show the first three quartiles for the x-axis variable; we also display 90% confidence intervals. The results suggest a positive and significant treatment effect for disadvantaged farmers in terms of their baseline features.

³⁴ This frequency is adjusted for the level of the node where the variable was used to split. Deeper nodes in the tree have less weight.

³⁵ See Athey et al. (2018) for details about the proof of asymptotic normality of $\hat{\tau}(x_p)$ and its variance equation; they define a pseudo-forest estimator in order to complete the proof. To simplify the explanation, we informally replace the notation of this pseudo-forest estimator with $\hat{\tau}(x_p)$

Appendix F. Heterogeneity analysis of the GRF-informed continuous treatment model

Appendix Table 6

Disadvantage regression coefficients.

	Target Activity Income	Investment	Expenditures
Months treated	239.1 (160.2)	371.4** (174.2)	-30.9 (46.2)
Months treated ²	-7.58 (10.2)	-15.3 (13.8)	2.21 (2.35)
Months treated ³	0.077 (0.18)	0.22 (0.24)	-0.041 (0.039)
Disadvantaged = 1	284.2 (1403.2)	-2098.0* (1094.4)	-591.4** (236.2)
(Disadvantaged = 1) × (Months treated)	98.8 (275.3)	-429.0 (397.0)	7.33 (62.6)
(Disadvantaged = 1) × (Months treated ²)	-13.0 (21.6)	19.9 (30.6)	0.021 (4.44)
(Disadvantaged = 1) × (Months treated ³)	0.24 (0.38)	-0.30 (0.53)	0.00071 (0.075)
Program farmer: education	361.5*** (118.2)	634.0*** (116.4)	242.2*** (40.9)
Program farmer: years of experience	61.8* (34.3)	2.09 (35.9)	12.3 (8.17)
Household members	-167.8 (144.5)	44.7 (148.0)	-470.4*** (40.7)
Landholdings: owned	43.9*** (13.5)	39.9*** (11.3)	6.83** (2.75)
Share of seasons used improved seeds	2522.9*** (726.5)	2280.3*** (707.5)	33.9 (177.8)
Program farmer: gender	-2651.1*** (683.7)	-1890.0* (1081.3)	-42.2 (255.7)
Constant	-3113.0 (2386.8)	19697.9 (15815.8)	13120.5*** (4325.9)
Observations	2459	2478	2518
Adjusted R ²	0.328	0.226	0.251

Cluster-robust standard errors in parentheses.

*p < .1, **p < .05, ***p < .01.

Not shown in table: Time and crop dummies, Mundlak instruments for fixed effects.

Appendix G. Treatment effects on capital flows and expenditure

Appendix Table 7

Impact of RBD program on the sum of investment and expenditure.

	ITT		LATE (IV)		LATE (complier sample)	
Early treatment	531.0 (443.2)	630.3* (333.4)	848.7 (698.0)	998.5* (517.7)	608.4 (400.0)	809.1** (381.3)
Baseline investment	-0.074 (0.062)	-0.058 (0.038)	-0.075 (0.062)	-0.059 (0.038)	-0.016 (0.039)	-0.024 (0.041)
Baseline consumption	0.64*** (0.10)	0.59*** (0.11)	0.64*** (0.10)	0.59*** (0.11)	0.52*** (0.087)	0.44*** (0.088)
Program farmer: education		86.1* (45.5)		79.0* (44.8)		77.4 (55.0)
Program farmer: years of experience		21.4** (9.29)		21.5** (9.16)		22.9* (13.1)
Household members		13.8 (63.5)		5.52 (61.5)		-61.3 (90.6)
Landholdings: owned		-13.2 (16.2)		-13.4 (15.9)		4.71 (3.15)
Share of seasons used improved seeds		324.5 (388.0)		229.1 (368.0)		74.8 (399.4)
Program farmer: gender		95.2 (491.2)		89.1 (484.7)		330.4 (649.9)
Constant	2091.6*** (465.4)	1796.0** (808.3)	2089.0*** (464.0)	1884.2** (793.8)	2561.1*** (503.3)	1849.3* (922.9)
Observations	1341	1260	1341	1260	860	817
Adjusted R ²	0.156	0.190	0.160	0.194	0.137	0.167

Cluster-robust standard errors in parentheses.

*p < .1, **p < .05, ***p < .01.

All regressions control for crop fixed effects. Share of improved seeds measures the share of seasons at baseline that the beneficiary household used improved seeds/methods.

Appendix H. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jdeveco.2018.11.006>.

References

- Abrevaya, J., Dahl, C.M., 2008. The effects of birth inputs on birthweight. *J. Bus. Econ. Stat.* 26, 379–397, <https://doi.org/10.1198/073500107000000269>.
- Aguero, J., Carter, M.R., Woolard, I., 2010. The Impact of Unconditional Cash Transfers on Nutrition: the South African Child Support Grant. (Working Paper).
- Anderson, J., Feder, G., 2003. Rural Extension Services. World Bank Policy Research Working Paper No. 2976.
- Ashraf, N., Giné, X., Karlan, D., 2009. Finding missing markets (and a disturbing epilogue): evidence from an export crop adoption and marketing intervention in Kenya. *Am. J. Agric. Econ.* 91, 973–990, <https://doi.org/10.1111/j.1467-8276.2009.01319.x>.
- Athey, S., Tibshirani, J., Wager, S., 2018. Generalized Random Forests. arXiv:1610.01271 <http://arxiv.org/abs/1610.01271>.
- Bandiera, O., Burgess, R., Das, N., Gulesci, S., Rasul, I., Sulaiman, M., 2017. Labor markets and poverty in village economies. *Q. J. Econ.* 132, 811–870.
- Banerjee, A., Duflo, E., Chattopadhyay, R., Shapiro, J., 2011. Targeting the Hard-core Poor: an Impact Assessment. Working Paper. Institute for Financial Management and Research Centre for Micro Finance.
- Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Pariente, W., Shapiro, J., Thuysbaert, B., Udry, C., 2015. A multifaceted program causes lasting progress for the very poor: evidence from six countries. *Science* 348.
- Bardhan, P., Mookherjee, D., 2011. Subsidized farm input programs and agricultural performance: a farm-level analysis of West Bengal's green revolution, 1982 - 1995. *Am. Econ. J. Appl. Econ.* 3, 186–214, <https://doi.org/10.1257/app.3.4.186>.
- Boucher, S.R., Carter, M.R., Guirkinger, C., 2008. Risk rationing and wealth effects in credit markets: theory and implications for agricultural development. *Am. J. Agric. Econ.* 90, 409–423.
- Boucher, S.R., Guirkinger, C., Trivelli, C., 2009. Direct elicitation of credit constraints: conceptual and practical issues with an application to peruvian agriculture. *Econ. Dev. Cult. Change* 57, 609–640.
- Buchinsky, M., 1998. Recent advances in quantile regression models: a practical guideline for empirical research. *J. Hum. Resour.* 88–126.
- Carter, M.R., Laajaj, R., Yang, D., 2013. The impact of voucher coupons on the uptake of fertilizer and improved seeds: evidence from a randomized trial in Mozambique. *Am. J. Agric. Econ.* 95, <https://doi.org/10.1093/ajae/aat040>.
- Carter, M.R., Laajaj, R., Yang, D., 2014. Subsidies and the Persistence of Technology Adoption: Field Experimental Evidence from Mozambique. Working Paper 20465. National Bureau of Economic Research.
- Chamberlain, G., 1982. Multivariate regression models for panel data. *J. Econom.* 18, 5–46.
- Chamberlain, G., 1984. Panel data. In: Griliches, Intriligator, M. (Eds.), *Handbook of Econometrics*.
- Cole, S.A., Fernando, A.N., 2016. Mobile'izing Agricultural Advice: Technology Adoption, Diffusion and Sustainability. Harvard Business School Finance Working Paper 13-047. Social Science Research Network, Rochester, NY.
- Davis, J.M.V., Heller, S.B., 2017. Using causal forests to predict treatment heterogeneity: an application to summer jobs. *Am. Econ. Rev.* 107, 546–550, <https://doi.org/10.1257/aer.p20171000>, <https://www.aeaweb.org/articles?id=10.1257/aer.p20171000>.
- Evenson, R.E., 2001. Chapter 11 - economic impacts of agricultural research and extension. In: *Agricultural Production*, vol. 1. Elsevier, pp. 573–628. Part A.
- Feder, G., Slade, R.H., Lau, L.J., 1987. Does agricultural extension pay? The training and visit system in northwest India. *Am. J. Agric. Econ.* 69, 677–686, <https://doi.org/10.2307/1241702>.
- Hirano, K., Imbens, G.W., 2004. The propensity score with continuous treatments. In: Gelman, A., Meng, X.L. (Eds.), *Applied Bayesian Modeling and Causal Inference from Incomplete-data Perspectives*. John Wiley & Sons, Ltd, pp. 73–84.
- Hulme, D., Hanlon, J., Barrientos, A., 2012. Just Give Money to the Poor: The Development Revolution from the Global South. Kumarian Press.
- Keswell, M., Carter, M.R., 2014. Poverty and land redistribution. *J. Dev. Econ.* 110, 250–261 Elsevier.
- King, E.M., Behrman, J.R., 2009. Timing and duration of exposure in evaluations of social programs. *World Bank Res. Obs.* 24, 55–82, <https://doi.org/10.1093/wbro/lkn009>.
- Koenker, R., Bassett, G., 1978. Regression quantiles. *Econometrica: J. Econ. Soc.* 33–50.
- Macours, K., Vakis, R., 2008. Changing Households' Investments and Aspirations through Social Interactions: Evidence from a Randomized Transfer Program in a Low-income Country. (Working Paper).
- Malacarne, J., Boucher, S.R., Carter, M.R., 2017. Goalpost Effects: Understanding Impact Heterogeneity in Nicaragua's Rural Business Development Program.
- McKenzie, D., 2012. Beyond baseline and follow-up: the case for more T in experiments. *J. Dev. Econ.* 99, 210–221, <https://doi.org/10.1016/j.jdeveco.2012.01.002>.
- Mundlak, Y., 1978. On the pooling of time series and cross section data. *Econometrica: J. Econ. Soc.* 69–85.
- Tjernström, E., Carter, M.R., Toledo, P., 2013. Identifying the impact dynamics of a small farmer development scheme in Nicaragua. *Am. J. Agric. Econ., Pap. Proc.* 95, 1359–1365.
- Toledo, P., Carter, M.R., 2010. Impact of business services on the economic well-being of small farmers in Nicaragua. BASIS Brief 2010-01, 1–6.
- Toledo, P., 2011. Impact Evaluation of a Rural Business Program Using Field Experiment Data. (PhD Dissertation).
- World Bank, 2008. Nicaragua - Poverty Assessment (Vol I - Main Report). Report 39736. Washington, D.C.