CAN SMALLHOLDER EXTENSION TRANSFORM AFRICAN AGRICULTURE?

Joshua W. Deutschmann
Maya Duru
Kim Siegal
Emilia Tjernström

Can Smallholder Extension Transform African Agriculture?
Joshua W. Deutschmann, Maya Duru, Kim Siegal, and Emilia Tjernström
NBER Working Paper No. 26054
July 2019
JEL No. O12,O13,Q12

ABSTRACT

Agricultural productivity in Sub-Saharan Africa (SSA) lags far behind all other regions of the world. A long list of policy experiments has yielded more evidence on what fails than on what works. We analyze a randomized control trial of a rare scaled-up success story: One Acre Fund's small farmer program. Much like anti-poverty "graduation" interventions, the program aims to relax multiple constraints to productivity simultaneously. We show that participation causes statistically and economically significant increases in output, yields, and profits. In our preferred specification, maize production increases by 24% and profits by 16%. We find little evidence of heterogeneous treatment effects on yields, but observe some attenuation of impacts on total output and profits at the top end of the distribution.

Joshua W. Deutschmann
Agricultural & Applied Economics
University of Wisconsin-Madison
427 Lorch St.
Madison, WI 53706
jdeutschmann@wisc.edu

Maya Duru
The Abdul Latif Jameel Poverty Action Lab
400 Main Street
E19-201
Cambridge, MA 02142
mduru@povertyactionlab.org

Kim Siegal
One Acre Fund
80 Broad St. Suite 2500
New York, NY 10004
kim.siegal@oneacrefund.org

Emilia Tjernström
Robert M. La Follette School of Public Affairs
University of Wisconsin–Madison
1225 Observatory Drive
Madison, WI 53706
tjernstroem@wisc.edu

1   Introduction

A central question in development economics concerns the relative importance of agriculture in the process of economic growth. Are dramatic increases in agricultural productivity pre-requisites for industrial-sector development and take-off, as Rostow (1990) argued? Or do we observe aggregate correlations between agricultural productivity and structural change merely because different growth indicators trend together as countries grow? While agriculture's broader role in driving growth is theoretically ambiguous (Matsuyama, 1992), recent empirical evidence suggests a causal link between increased staple yields and national welfare (Bravo-Ortega and Lederman, 2005; Ligon and Sadoulet, 2011; Ravallion and Chen, 2007), especially GDP growth, through the increased use of improved agricultural technologies (Gollin et al., 2018). Further, McArthur and McCord (2017) show that agricultural productivity growth plays an important role in driving structural change, over time reducing the labor share in agriculture. Additionally, beyond its potential influence on industrialization and growth, agricultural productivity directly influences the food security and well-being of billions of people whose livelihoods depend on agriculture (Byerlee et al., 2009).

Agricultural productivity in Sub-Saharan Africa (SSA) lags far behind all other regions of the world (Block, 2014; World Bank, 2008) despite substantial government expenditures in agriculture (Akroyd and Smith, 2007). Some scholars believe that low agricultural productivity has contributed to the region's failure to embark on a path of sustained economic growth. The slow growth of agricultural output, modern input use, and commercialization serve as a kind of litmus test for researchers and policy-makers alike: to some, the existence of constraints to technology adoption and commercialization suggests that relaxing those constraints with well-designed investments could stimulate dramatic increases in African agricultural productivity (Byerlee et al., 2009; Magruder, 2018). To others, widening productivity gaps indicate that policy-makers should instead allocate scarce public resources towards directly stimulating industrialization by, for example, investing in manufacturing.

Even among those who believe that agriculture is crucial to the region's growth, there is recognition that a long list of agricultural policy experiments—whether implemented by governments, NGOs, or researchers—have yielded substantial evidence on what fails and a limited catalogue of successes. Evidence on the effectiveness of agricultural extension services is mixed at best, plagued by issues of measurement, selection, and comparability across programs (Aker, 2011; Anderson and Feder, 2007). Extension services often train lead or contact farmers in villages, who are then tasked with diffusing information to other farmers. Even when these contact farmers adopt new technologies, diffusion to other nearby farmers tends to be limited (Kondylis et al., 2017). A number of studies find that the adoption of new practices or increased input use are often not enough to induce measurable increases in yields (Udry et al., 2019) or profits (Beaman et al., 2013).

If large investments in government extension programs, input subsidies, and insurance products have not transformed agricultural productivity, is this a sign that African agriculture

is a lost cause? An alternative explanation, argued forcefully by Collier and Dercon (2014), is that the predominant focus on smallholder agriculture is to blame. Or perhaps existing public and private providers of extension services have thus far failed to successfully address the constraints facing the large populations of poor staple crop producers, focusing instead on regions with high agricultural potential and those that can grow cash crops (Muyanga and Jayne, 2006).

Our study provides experimental evidence on the impacts of a rare success story that explicitly targets small staple crop producers, and which is operating at scale. We evaluate One Acre Fund's (1AF) small-farmer program. 1AF is an NGO that currently works with over 700,000 farm households in 7 countries in Eastern and Southern Africa, up from only 200,000 in 2014 (and 38 farmers in their 2006 pilot program). The components of 1AF's programming will not surprise readers familiar with the 1960s and 1970s World Bank-promoted agricultural extension investments (Birkhaeuser et al., 1991): participating farmers receive training on improved farming practices, input loans, and crop insurance. One feature that distinguishes the 1AF program from its antecedents is the tight bundling of these component parts into a single program offering (Tinsley and Agapitova, 2018); this is typically not the case for government-led extension, nor for more recent programs targeting individual constraints to productivity.

We analyze a pre-registered cluster-randomized control trial of 1AF's main program in Kenya and find that program participation causes statistically and economically significant increases in yields and profits. The effects are large (especially compared to most agriculture-focused programs): in our preferred specification, total maize output increases by 24% and profits increase by 16%. This effect is robust to a variety of specifications and sample definitions.

Given the recent proliferation of NGOs and social enterprises offering financial and extension services to farmers, this result has implications beyond 1AF. If past programs have in fact underestimated the potential for well-designed programs to increase agricultural productivity in SSA, then basing investment decisions on past evidence could lead to socially sub-optimal choices. Unlike many government-run extension and input subsidy programs, 1AF is largely farmer-funded, covering about 75% of operating costs from farmer participation fees and overhead on input sales (Tinsley and Agapitova, 2018). In this paper, we find that a well-designed, efficiently-run program can not only provide economically significant returns to farmers, but that this is possible with limited subsidies by donors.

We additionally explore several approaches to heterogeneity analysis, to better understand whether impacts vary markedly across individuals. We find little evidence of heterogeneous treatment effects on maize yields. We observe some evidence of heterogeneity in the program's effects on total maize output and profits, with lower treatment effects at the top end of the distribution.

2 Context, data and experimental design

This study analyzes the main operating model of an established agricultural NGO, the One Acre Fund. The organization was funded in 2006, and has grown rapidly in the last several years. In 2014, the NGO reported working with around 200,000 farmers, compared to their current enrollment numbers of over 700,000 smallholder households across six different countries in Eastern and Southern Africa (One Acre Fund, 2019). The NGO's core "market bundle" provides farmer groups with group-liability loans for improved seeds and high-quality fertilizer, regular trainings on modern agricultural techniques, crop and funeral index-based insurance, and market facilitation support to help farmers sell their products for higher prices (Tinsley and Agapitova, 2018). Farmer groups are organized by geographical areas that can be served by a single 1AF officer, and they typically range in size from 8-12 farmers.

The 1AF program's focus on credit, insurance, inputs, and information will feel familiar to many readers. The program components have many parallels to the agricultural extension programs popular in the 1960s and 1970s (Birkhaeuser et al., 1991). The motivation behind these components has support in the economics literature, which has accumulated substantial evidence that failures in these domains hinder farmers' ability or willingness to adopt improved agricultural technologies. Feder et al. (1985) emphasize three of these constraints in their review of the technology adoption literature at the time (credit, risk, and information). Magruder (2018) focuses on the same three constraints as key barriers to adoption in a more recent review of the experimental evidence on technology adoption.[1]

Farmers choose the amount of land to enroll, and 1AF provides the agricultural input loans as a function of the amount of land enrolled. The group liability loans are given to self-selected farmer groups. The loan terms are flexible, allowing farmers to repay in any amount at any time during the growing season. Groups must complete repayment in full by the end of the harvest, but they have a 2-week grace period to ensure repayment by all members. Historically, repayment rates have been over ninety-seven percent (Tinsley and Agapitova, 2018). Field officers conduct specific in-field training in targeted areas throughout the season, and provide educational handouts on fertilizer impact and proper use. Additionally, 1AF carries a weather-index insurance and passes on the benefit to farmers by forgiving input loans in case of crop failure, thus helping them mitigate risk (Tinsley and Agapitova, 2018).

This RCT was conducted in western Kenya, where 1AF has operated for more than ten years and reached over 200,000 enrollees in 2016. Kenya's agriculture sector contributes 51 percent to the country's GDP (25% indirectly) and is dominated by small-scale producers (Government of Kenya, 2010). Despite the importance of the agricultural sector to the economy, most smallholder farmers are not running successful micro-enterprises. Households

---

[1]The focus on providing high-quality inputs also has empirical support: a growing literature establishes that fertilizer available on local markets in the region often falls short of quality standards (Bold et al., 2017; Tjernström et al., 2018).

in Kenya typically derive their income from the production of a variety of crops, often combined with a range of off-farm activities (Sheahan et al., 2013). These statistics are reminiscent of Collier and Dercon's (2014) description of African agriculture's "reluctant micro-entrepreneurs," whom they describe as a "recipe for continued divergence of the [agricultural] sector from global agricultural performance, limiting growth and unlikely to help large scale poverty reduction."

The main crop in Kenya is maize, a staple crop that is important both to the economy and to food security. Accordingly, a large share of 1AF's efforts are devoted to the crop. Seventy percent of Kenya's maize is produced by smallholders who farm between 0.2 and 3 hectares (Government of Kenya, 2010, pg. 11-12). While Kenyan farmers use improved maize varieties and inorganic fertilizer at higher rates than other countries in the region, yields remain low and the country is a net importer of maize, despite policy goals to the contrary. Increasing yields and profits is crucial if we want the agricultural sector to act not simply as a means of subsistence, but as a pathway out of poverty.

2.1 Experimental design

The cluster-randomized experimental design was carried out in the Teso region of Kenya. Recruitment, enrollment, and the intervention itself followed the NGO's standard protocol. Participants self-selected into farmer groups of 8-12 farmers. The randomization took place at the level of a cluster of 2-4 of these joint-liability farmer groups. Typically, once a farmer indicates interest in signing up and satisfies the basic criteria, she pays a small program participation deposit of approximately $5 USD.[2]

Once farmers had paid the fee and signed the contract, they were informed that half of them would be randomly assigned to treatment, while the other half would receive an alternative compensation package consisting of household goods and a discount for program participation the following season—this amounted to roughly 20% of the typical program cost. Randomization took place by public lottery.

While the ideal areas for a study like this would perhaps focus on a new area, where the organization had never operated, this proved difficult in Kenya. Since the study took place in a country where 1AF has already operated for several years, the second-best was an area in which 1AF had already begun operations but where marketing had not reached all of the target areas.[3] The specific villages selected for study inclusion had never been offered the 1AF program, but neighboring villages had previously been offered the program.

---

[2]In addition to the participation fee, farmers who wish to join 1AF must additionally have a phone number and national identification, and agree to repay their loan. For the study population, 1AF added the requirement that farmers give consent to be part of the study, that they had to cultivate maize, and that they be able to cultivate at least a quarter of an acre of maize. Shortly after contract signing, 1AF informed farmers about the study, that their participation would be voluntary, and provided them with informed consent documents.

[3]Choosing completely untouched regions in Kenya would have forced the evaluation team to study regions that are quite unrepresentative of the typical program.
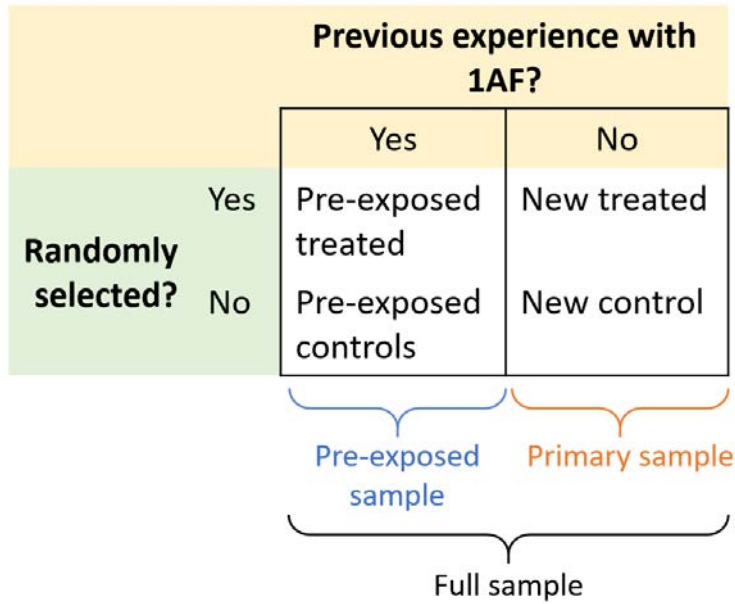
Figure 1: Sample definitions

Thus, a substantial portion of farmers who expressed interest in participating had previously participated in the 1AF program (by "commuting" to the neighboring villages to participate).

Because of the contamination of some of the sample, the experimental design stratified randomization on previous exposure to the 1AF program. Throughout the paper, we estimate treatment effects for two samples: the primary sample, which consists of treated and control farmers who had never previously participated, and the full sample, which includes the pre-exposed farmers. Figure 1 illustrates the difference between the different samples.

Both of the samples present some challenges. On the one hand, if the pre-exposed control farmers continue to benefit from their prior program involvement even after quitting the program (by, for example, continuing to use the new practices that they've learned), then we would expect treatment effects estimated with the full sample to result in a downward-biased impact estimate. On the other hand, farmers who had never been exposed may be different than those who self-selected into the program in earlier years. If prior selection into the program is positively correlated with potential returns to the program (i.e., more entrepreneurial or higher-return farmers opt in earlier), we would expect the impacts on our primary sample of "hold-out" farmers to be a lower bound of the true impacts; if on the other hand the hold-out farmers resemble the never-adopters in Suri (2011) and have surprisingly high returns but perhaps high costs of participation, then it is possible that focusing solely on the primary sample might overestimate impacts.

6

2.1.1 Data

The data collection was directly managed by 1AF. To reduce potential concerns about the independence of the research design, the research team took three main steps. First, it worked with the International Initiative for Impact Evaluation (3ie) to help design and review all parts of the trial—including the experimental design, the field protocols, sampling and randomization, as well as the data collection instruments (Dubey and Yegbemey, 2017). 3ie concluded that the randomization was conducted successfully, and noted that 1AF staff showed high levels of professionalism in conducting the randomization.[4] Second, the data collection applied best-practice protocols to ensure data quality, such as back-checks and in-field supervision. Third, 1AF contracted an independent survey firm, Intermedia Development Consultants (iDC), to carry out a three-step audit of the data collection. The overall summary evaluation concludes that "...the data collection is well planned and executed. The possibilities for improving performance are quite limited. The survey team members are well selected and committed to the process. The recording and transmission of data is well done, with minimal errors." Fourth, some readers may worry about social desirability bias in contexts where participants or enumerators want to "help" by showing the program in a good light. In our case, the main dependent variable was physically weighed, so should help alleviate such concerns. Finally, two of the authors on this paper were brought in as independent evaluators; this part of the team reviewed the pre-analysis plan, cleaned the data, constructed variables, and conducted the analysis according to the pre-analysis plan. Appendix D contains more details on variable construction,

Baseline data collection occurred in November and December of 2016—after program enrollment but before treatment assignment. The public lottery, which assigned clusters of farmer groups to treatment, took place in January 2017. Enumerator teams rolled out input use surveys after the planting of the main season in 2017, from April through June. The data from these surveys provide detailed information on the extent of farmer compliance with behaviors that the program encourages, including whether farmers spaced their plants correctly and applied the correct fertilizer dosage at the recommended time during the season. We use these data as indicators of the extent to which farmers are actually learning and changing their behaviors as a result of the 1AF training.

2.2 Outcome variables

We pre-registered three main outcome variables: program maize yields, total maize output, and profits. The relevant measures of harvests and land sizes are observed, rather than based on farmer self-reports. Program maize yields are so termed because they compare yields on treatment farmers' enrolled plot to control farmers' overall per-acre yields. Total maize output measures the overall output on farmers' maize land, and profits are computed

---

[4]A letter from 3ie attesting to their review and approval of the pre-analysis plan can be found in Appendix A.

as the projected value of the output less farmers' costs. The data for the physical harvests and yields were collected during the main harvest period in 2017. Enumerators collected fresh and dry harvest weights from two randomly placed 8 x 5 meter boxes for each selected plot.[5] Land sizes were measured by GPS readings. Based on recent research documenting substantial non-classical measurement error in both variables (see for example Abay et al. 2019; Carletto et al. 2013; Desiere and Jolliffe 2018; Gourlay et al. 2017), we expect these methods to produce substantially better yield estimates than the self-reports commonly reported in economics studies.

The two physical output measures have different benefits. The program maize yields measure the direct impacts on the land that farmers chose to enroll. The total output measure better measures overall welfare impacts and should account for any spillover effects between enrolled and non-enrolled plots. For example, if farmers decide to reallocate labor and other complimentary inputs to the enrolled plot, sacrificing production on the non-enrolled plot, then yields on the enrolled plot would overestimate the true impacts on production and welfare. Alternatively, if there are learning spillovers or if farmers reallocate some inputs to the non-enrolled plots, then effects on program land only could understate the effects.[6]

2.2.1 Balance

Table 1 presents full-sample baseline means of pre-specified control variables, as well as balance tests across treatment and control samples. Appendix B displays distributions of the non-binary variables. The summary statistics depict a population of mostly married household heads with low levels of secondary school completion. The majority of households earned more than half of their income from farm labor in the last year and planted roughly one acre of maize, harvesting roughly half a ton (500 kgs) of maize per acre. Three-quarters of the sample use improved agricultural technologies at baseline, but a closer look at the intensity reveals that the actual rate of input use is quite low.

About half the sample answer that they have knowledge of 1AF planting practices, which is primarily driven by the pre-exposed farmers.[7] By chance, household size and 2016 maize yields differ significantly across treatment and control. An *F*-test of joint orthogonality of the variables in Table 1 is not significant in the full sample, but it is significant at the 10%

---

[5]For control farmers, per-acre yields are computed by scaling the harvest box measurements to one acre. For treatment farmers, harvests are scaled separately for the enrolled and non-enrolled plots. For the total output measure, these are scaled to the total size of the plot and then aggregated. Harvest box measurements are considered a more reliable way to measure physical yields than relying on self-reports, which are subject to recall bias and have been shown to suffer from several types of non-classical measurement error.

[6]Input reallocation could lead to positive yield effects overall if the marginal returns to inputs on the non-enrolled plot exceeded those on the enrolled land, which seems plausible if input levels are constrained in the absence of the program.

[7]The levels are lower in the primary sample: six percent in the control group and fourteen percent in the treatment group; Appendix B provides more detail, including separate balance tables for the primary sample and the pre-exposed sample. That said, the difference between treated and control groups is statistically significant in the primary-only sample.

Table 1: Baseline balance across treatment and control groups

	(1)		(2)		
	Control		Treatment		Difference
Variable	N/[Clusters]	Mean/SE	N/[Clusters]	Mean/SE	(1)-(2)
Panel A: Binary variables					
Married (0/1)	1056	0.878	1066	0.880	-0.002
	[60]	(0.012)	[60]	(0.011)	
Household head has secondary school	1056	0.385	1066	0.426	-0.040
	[60]	(0.020)	[60]	(0.023)	
Household income >50% from farm labor	1056	0.777	1066	0.779	-0.001
	[60]	(0.020)	[60]	(0.019)	
Used improved ag technology in 2016	1056	0.784	1066	0.802	-0.018
	[60]	(0.025)	[60]	(0.017)	
Reports knowledge of 1AF practices	1056	0.468	1066	0.526	-0.058
	[60]	(0.033)	[60]	(0.026)	
Intercropped maize and beans in 2016	1056	0.475	1066	0.477	-0.001
	[60]	(0.027)	[60]	(0.031)	
Reports having credit access in 2016	1056	0.709	1066	0.726	-0.017
	[60]	(0.022)	[60]	(0.020)	
Panel B: Continuous variables					
Household size	1056	6.647	1066	6.820	-0.173*
	[60]	(0.095)	[60]	(0.099)	
Acres under maize cultivation in 2016	1056	1.004	1066	1.024	-0.020
	[60]	(0.042)	[60]	(0.034)	
Maize yield (kg/acre) in 2016	1056	534.787	1066	579.901	-45.115*
	[60]	(29.222)	[60]	(28.181)	
F-statistic (test of joint significance)					1.089
Number of observations					2122

Notes: Field office fixed effects are included in all estimation regressions. Standard errors are clustered at farmer group cluster level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

level in the primary sample. Household size and baseline maize yields are by chance different between the two groups, with maize yields almost nine percent larger in the treatment group. We are not very concerned about this difference, since the variables that are slightly imbalanced are not highly correlated with our outcome variables. This may seem surprising for baseline maize yields, but several factors contribute to the low correlations: first, baseline yields are based on self-reported harvests and land sizes, both of which likely suffer from substantial measurement error. Second, the stochastic nature of agricultural production leads to substantial year-to-year variation in productivity.

2.2.2   Pre-exposed vs. new participants

Table B.2 and Figure B.2 compare participants across the pre-exposed and new participant samples. The two samples differ substantially. Farmers who previously self-selected into the program are more likely to have an educated father in the household, more likely to earn at least fifty percent of household income from farm labor, more likely to report using improved seeds or fertilizer at baseline, more likely to report baseline knowledge of 1AF practices, less likely at baseline to intercrop beans with maize, more likely to report having access to credit, reported using more land for maize at baseline, have bigger households, and reported higher maize yields per acre at baseline. Several of these differences (education, land endowments, yields, etc.) suggest that farmers who enrolled earlier may come from a better-off population. That said, several of these variables could also have been affected by program participation, so the slight attenuation of estimated program impacts in the full sample (compared to the primary sample) should be interpreted accordingly.

3   Results

Our econometric approach for the main analyses is straightforward, given the randomization. For each outcome y_{is}, we estimate the following regression:

$$y_{is} = \alpha + \beta T_{is} + \delta X_{is} + \gamma_s + \varepsilon_{is} \tag{1}$$

where T_{is} is the treatment dummy for individual i in field officer area s, X_{is} includes the list of pre-specified controls (when included), γ_s is a field office area fixed effect, and ε_{is} is clustered at the farmer group cluster level, i.e., the level of the treatment assignment. Our pre-analysis plan (PAP) specified three outcomes of interest: program maize yields, total maize output, and profits. The PAP specified that X_{is} in Eq. 1 include controls for marital status, household size, amount of land owned, education, agricultural reliance, credit constraints, use of agricultural technology in the prior season, intercropping, and knowledge of 1AF practices.[8] We follow the PAP closely, but also present results both with and without

---

[8]Where appropriate, regressions should also include a control for whether the farmer was pre-exposed. Additionally, the PAP planned to include a spillover inverse probability weight. Our results are robust

controls, which rarely leads to any changes in the results.

3.1 Intermediate outcomes: behavioral change

We begin by examining some intermediate outcomes that were not described in our PAP but that are useful for thinking about "first-stage" effects: does program participation change the kinds of behaviors that 1AF encourages? The NGO's farmer trainings include several components, including a crop calendar that guides farmers through the different stages of production. Another cornerstone of the program is called the "power of three," which refers to hybrids, fertilizer, and 1AF practices. Field officers teach these practices, and follow up with farmer groups afterwards. Of particular focus are land preparation (emphasizing the timing of planting relative to the onset of rains for the season), input application (focusing on the use of hybrid seeds and fertilizer, both with respect to the type of fertilizer and the timing of different fertilizer applications), proper spacing between rows of plants and of plants within a row. Each practice is tied to farming aspects that the organization has identified as important for crop yields and are documented in a 95-page manual.

Given the many parts of the program, and the wide range of improved practices that 1AF encourages, farmers could respond to the program in a variety of ways and through many different mechanisms. If we observe no changes in farmer practices, we could imagine that a cheaper version of the program—perhaps without the intensive training—could achieve similar results at a lower cost. The bundled nature of the program prevents us from confidently digging deep into the mechanisms through which the program works, but we use detailed data on behavior to examine uptake of practices that are actively encouraged by the NGO.

We construct several metrics related to farming practices: first, we generate indicator variables that equal one if participants' plant spacing was within 5 cm of the recommended spacing, separately for rows and plants.[9] Second, we measure whether farmers applied fertilizer at the appropriate time of the season.[10] Maize requires substantial nitrogen application to produce high output, but requires relatively little up front and greater amounts midway through the season, which is what is referred to as "top dressing." Research suggests that even highly experienced producers can fail to notice crucial features of the production process (Hanna et al., 2014) or less-salient profitability margins (Beaman et al.,

---

to including this weighting, but for simplicity we present all results without weights. The PAP specified additional analysis to explore "enduring impacts" using only the control sample, i.e. exploring whether the pre-exposed farmers in the control group were still better off.

[9]1AF participants receive a stick that corresponds to the recommended spacing between rows, and a string that marks the recommended space between plants in a row. The planting string has colored ties every 25 cm along the length of the to indicate where the next plant should go. The spacing is clearly defined to encourage appropriate plant density. Placing plants too close together could result in plants competing for sunlight, water and nutrients, but planting too sparsely will waste space, which is a binding constraint for most smallholders.

[10]1AF recommends that farmers apply DAP (diammonium phosphate) at the time of planting, and CAN (calcium ammonium nitrate) at top dressing, which takes place several weeks after planting.

2013; Duflo et al., 2008). Given that fertilizer rates are often discussed without specific attention to the within-season timing, it seems plausible that farmers may not be aware of the importance of timing.[11]

Table 2: Take-up of program practices and input use, primary sample

Panel A: Take-up of program practices	(1) Row Spacing	(2) Plant Spacing	(3) Fertilizer Timing	(4) Used Plow
1AF participant	0.22***	0.21***	0.66***	0.05
	(0.040)	(0.030)	(0.030)	(0.030)
Control group mean	0.37	0.09	0.26	0.73
Observations	757	757	757	757
Panel B: Input costs (USD)	Fertilizer	Seeds	Paid Labor	Unpaid Labor
1AF participant	21.74***	2.61***	5.68*	5.67***
	(1.950)	(0.860)	(3.210)	(1.070)
Control group mean	17.56	14.53	30.33	14.08
Observations	691	691	691	691

This table presents results from linear regressions of the outcome variables in each column on the treatment dummy. Panel A shows the effect of 1AF participation on the use of practices recommended by the NGO: the spacing used between plants and fertilizer use at the correct time in the season. Row and plant spacing are indicator variables for whether or not a farmer was within 5cm of the recommended spacing, for the rows in which they planted and the plants within rows, respectively. Fertilizer timing is an indicator variable for whether or not farmers applied DAP at planting and CAN at top dressing (the timing recommended by the NGO). Used plow is an indicator for whether or not the farmer used a plow to prepare their plot. Panel B shows the effect of 1AF participation on the intensive margin of farmer expenses for fertilizer, seeds, paid labor, and unpaid labor. Costs are expressed in USD. For more on how we define labor costs, see Appendix D. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. All regressions include field office (site) fixed effects. This table includes only farmers in the primary sample.

Panel A of Table 2 reports treatment effects from linear probability models of the indicator variables for adopting 1AF practices. Columns (1)-(3) show that program participants are substantially more likely than control farmers to follow spacing recommendations and to apply the different types of fertilizer at the recommended time. Column (4) shows the treatment effect on the use of an ox-plow. While 1AF provides farmers with specific recommendations about when to plough their fields, their manual is agnostic about the use of a plough versus more manual methods. They specifically instruct farmers not to wait for an ox plow to be available, but to instead focus on timing relative to the onset of the rains. Incorrect farming practices certainly have the potential to reduce productivity—both directly by reducing productivity and indirectly through effects on adoption due to lowered returns to said inputs.[12] Still, improving planting practices alone without an accompanying increase in the intensive margin of investments seems unlikely to be able to put the region

---

[11]In the region, typical government recommendations regarding fertilizer use tend to focus primarily on the amount of fertilizer rather than on application timing. While some extension manuals do distinguish between fertilizer application at planting and at top dressing, they typically do not explain the rationale behind these different timings or emphasize the importance of timing. See for example National Farmers Information Services (2019).

[12]The latter effect resembles what we would expect to observe in the presence of low-quality inputs as in Bold et al. 2017).

on a dramatically different growth trajectory.

Panel B shows the causal effect of program participation on the intensive margin of participant farming investments. Enrolled farmers spend 124% more on fertilizer, 18% more on seeds, 19% more on paid labor, and 40% more on unpaid labor.[13] These are substantial expenditure increases, all of which (including unpaid labor) are accounted for in the profit calculations. Treated farmers' fertilizer expenditures are more than double those in the control group, seed expenditures increase by eighteen percent. Labor costs increase by nineteen percent for paid labor and forty percent for unpaid labor.[14]

Table 3 reports the full-sample analog to Table 2, controlling for previous program exposure and its interaction with the treatment. We interpret these results with caution, since we do not know whether differences between the pre-exposed and new farmers are due to their past participation in the program or due to pre-existing differences between the two populations. That said, we believe that it is informative to compare the treatment effects for the two populations. If information constraints constitute the main obstacle to agricultural productivity in this context and one-time exposure to the program suffices to relax those constraints, then it could be socially efficient to focus on training new farmers, rather than providing the same information to participants who have already received it. If behavioral change instead takes place more slowly, re-exposure to the training could be beneficial.

The results in Panel A of Table 3 show that the program has substantially smaller effects on 1AF-promoted practices for the pre-exposed farmers, but the point estimates are still positive and significant. While it is possible that pre-exposed farmers are actively choosing not to adopt these practices when not enrolled, both spacing and fertilizer timing are positively correlated with yields.[15] We therefore think it is more plausible that less-than complete adoption for pre-exposed farmers who were randomized into the control group is due to imperfect recall rather than representative of an active choice to not adopt.

Panel B of Table 3 reports effects on the overall level of investments in maize production. Unlike for behavioral change, the treatment effect on the pre-exposed is statistically indistinguishable from that on new enrollees. Combined with the fact that the pre-exposed farmers use more inputs even when not enrolled, as seen by the coefficient on the pre-exposed dummy, this could suggest that credit constraints limit pre-exposed farmers from applying as much fertilizer and seed as they would like. Pre-exposed farmers do enroll slightly more land in the program than new farmers and plant slightly more land overall, so their higher input use could reflect an increase along the intensive margin of maize production, or that

---

[13]Note that 1AF charges prices for seeds and fertilizer that are comparable to local market prices; their average gross margin on inputs is 32 percent according to Tinsley and Agapitova (2018), which is similar to markups in the agro-dealer sector in this region. The program effects on costs are therefore unlikely driven by prices.

[14]Appendix D contains more details on how we define labor costs, especially for unpaid labor and as it relates to profits.

[15]Production function estimation results showing the relationship between various behaviors and yields available from the authors.

Table 3: Take-up of program practices and input use, full sample

Panel A: *Take-up of program practices*	(1) Row Spacing	(2) Plant Spacing	(3) Fertilizer Timing	(4) Used Plow
1AF participant	0.23***	0.21***	0.66***	0.05
	(0.040)	(0.030)	(0.030)	(0.040)
Pre-exposed	0.07*	0.02	0.15***	0.08***
	(0.030)	(0.020)	(0.030)	(0.030)
1AF participant	-0.10**	-0.07**	-0.18***	0.02
× pre-exposed	(0.050)	(0.040)	(0.040)	(0.040)
Control group mean	0.41	0.12	0.37	0.76
N	2122	2122	2122	2122
Panel B: *Input costs (USD)*	Fertilizer	Seeds	Paid Labor	Unpaid Labor
1AF participant	21.89***	2.76***	6.23*	5.92***
	(1.960)	(0.850)	(3.340)	(1.130)
Pre-exposed	5.84***	1.60**	6.38***	-0.68
	(1.700)	(0.770)	(2.160)	(0.970)
1AF participant	-0.03	0.36	3.11	1.25
× pre-exposed	(2.750)	(1.130)	(4.090)	(1.380)
Control group mean	22.24	15.42	32.85	13.31
N	1919	1919	1919	1919
Panel C: *Input costs/acre*	Fertilizer	Seeds	Paid Labor	Unpaid Labor
1AF participant	22.54***	0.38	-0.99	3.49*
	(2.100)	(1.000)	(3.620)	(1.790)
Pre-exposed	2.85	-0.83	3.68	-3.63**
	(1.980)	(0.980)	(3.180)	(1.470)
1AF participant	-2.33	0.03	3.93	0.69
× pre-exposed	(2.620)	(1.150)	(4.260)	(2.090)
Control group mean	27.45	19.37	39.92	19.69
N	1919	1919	1919	1919

This table presents results from linear regressions of the outcome variables in each column on the treatment dummy. Panel A shows the effect of 1AF participation on the use of practices recommended by the NGO: the spacing used between plants and fertilizer use at the correct time in the season. Row and plant spacing are indicator variables for whether or not a farmer was within 5cm of the recommended spacing, for the rows in which they planted and the plants within rows, respectively. Fertilizer timing is an indicator variable for whether or not farmers applied DAP at planting and CAN at top dressing (the timing recommended by the NGO). Used plow is an indicator for whether or not the farmer used a plow to prepare their plot. Panel B shows the effect of 1AF participation on the intensive margin of farmer expenses for fertilizer, seeds, paid labor, and unpaid labor. Costs are expressed in USD. For more on how we define labor costs, see Appendix D. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. All regressions include field office (site) fixed effects. This table includes both primary and pre-exposed farmers.

they have access to more land or capital to begin with.[16]

Panel C accounts for this by reporting program impacts on per-acre use of these inputs. These results show that the difference in overall input use for pre-exposed farmers seems to be driven by differences in the amount of land planted to maize, with attenuated effects on seed costs and paid labor, but very similar impacts on fertilizer rates and unpaid labor use to those in Panel B. The estimated effects of the program on fertilizer and labor therefore seem to be taking place at least partly through increased intensification and not primarily through extensive margin effects on maize acreage planted.

3.2 Main outcomes

Figure 2 presents the distributions of total production, maize acreage, and overall maize yields by treatment status. For treatment farmers, total maize production and acreage are the sum of harvests and acreage on their enrolled and non-enrolled plots, and the yields represent the harvests divided by each plot's respective size. In the top figure, we can see that treatment farmers are less likely to attain very low production levels. Further, the treatment group distribution lies to the right of the control group distribution for much of the support. However, the second panel shows that treatment farmers also plant slightly more land to maize (by 0.1 acres on average, significant at the 5% level). The bottom panel tries to adjust for this and shows the overall per-acre yields that farmers get. Similar to the level effects, treatment group farmers are less likely to experience very low yields, and the distribution of yields is clearly shifted to the right. Note that the top figure here corresponds to what we call the total output measure, since it accounts both for yields on the enrolled plot and the non-enrolled plot, but the bottom panel is different than the program maize yields as it counts yields on both program and non-program land for treated farmers.

Table 4 presents regression estimates of the average treatment effects (ATEs), following Eq. 1, for our three outcomes of interest in both the primary and full samples. Panel A reports the results from the primary sample and Panel B shows results in the full sample (Tables C.1 and C.2 report the full set of coefficients for the regressions with control variables). Participation in the 1AF program has an economically and statistically significant impact on maize yields, total output, and profit. The impact on program yields ranges between 25-28% across the different samples and specifications, total output is 17-24% greater in the treatment group, and profit impacts range between 8% in the full sample with covariates to 16% in the primary sample. The point estimates are systematically attenuated in the full sample, but program participation still has a substantial and positive impact, even in this sample where a marked portion of participants had previously been exposed to the program. These results are robust to various different specifications, sample definitions, and attempts

---

[16]Pre-exposed treatment farmers enroll about 0.1 acre more than new enrollees; t-statistic for test of the difference between pre-exposed and new farmer enrolled maize acreage: -3.8771. Pre-exposed farmers also plant more land to maize overall than new farmers, regardless of treatment status, the difference is around 0.05 acres; t-statistic: -2.5036.
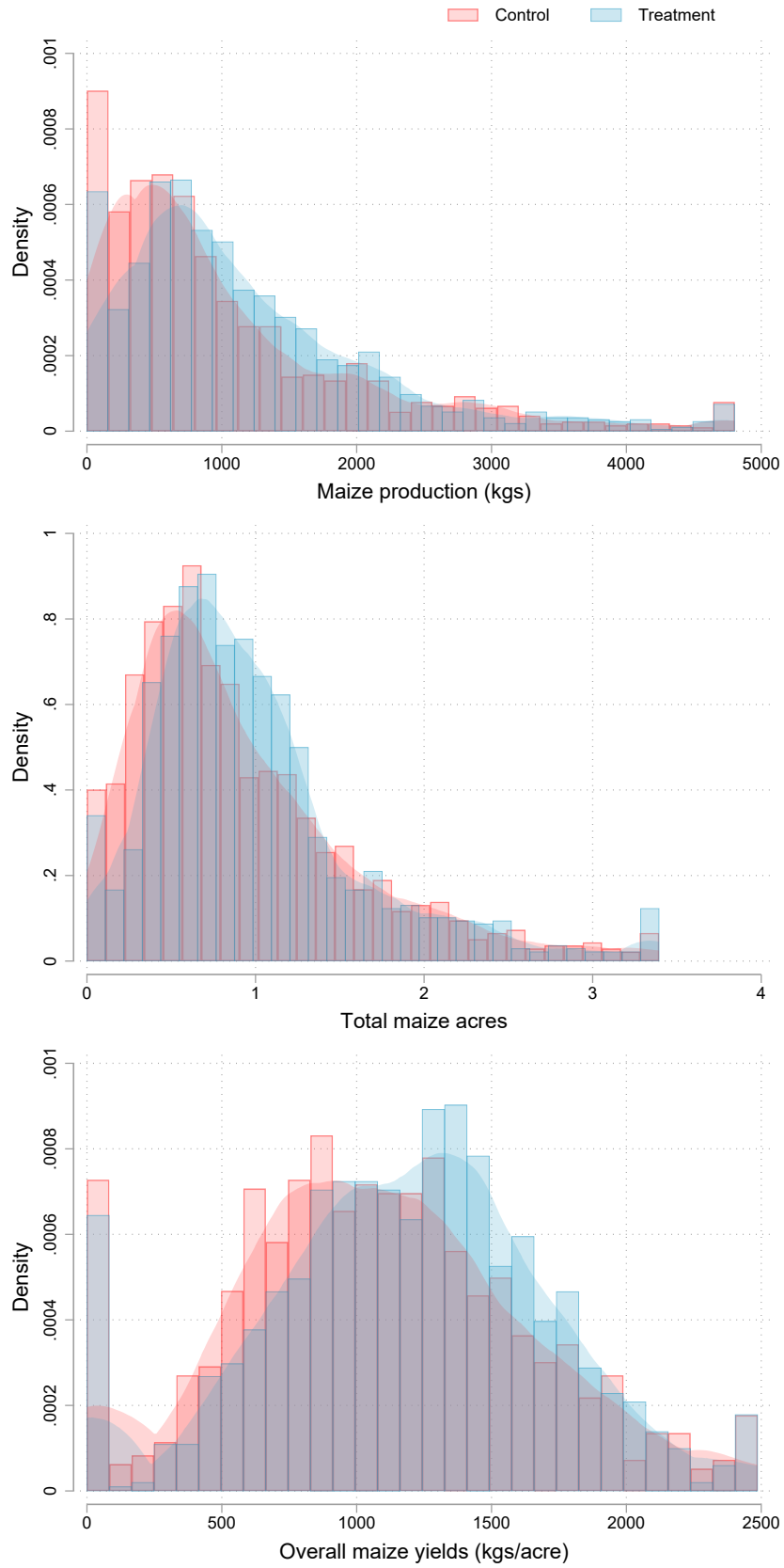
15

Figure 2: Distributions of production, acreage and yields

Table 4: Main results

	Program Maize Yields		Total Maize Output		Profit	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Primary sample						
1AF participant	304.76***	318.90***	250.52***	264.92***	49.50	54.80**
	(35.370)	(36.050)	(85.530)	(75.750)	(30.030)	(27.220)
Baseline controls	N	Y	N	Y	N	Y
Control group mean	1123.58		1082.93		339.73	
Observations	757	757	691	691	691	691
Panel B: Full sample						
1AF participant	295.95***	290.61***	258.57***	200.22***	46.40**	28.48*
	(23.590)	(23.350)	(49.400)	(44.760)	(17.760)	(16.740)
Baseline controls	N	Y	N	Y	N	Y
Control group mean	1144.56		1160.03		366.40	
Observations	2122	2122	1919	1919	1919	1919

This table presents results from linear regressions of the three outcomes of interest on the treatment dummy. Program maize yields are measured in kgs per acre, total maize output is measured in kgs, and profits are measured in USD. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. The primary sample includes only farmers who had never previously participated in the 1AF program. The full sample additionally includes a sample of farmers who had previously enrolled in the 1AF program. All regressions include field office (site) fixed effects. Columns 2, 4, and 6 include the full set of pre-specified controls, omitted here. Binary baseline controls are married, at least secondary education by male household head, whether the household receives half of its income from farm labor, whether the household used improved seeds at baseline, whether the household reported knowledge of 1AF practices at baseline, whether the farmer intercropped beans and maize at baseline, and whether the household reported having access to credit. Continuous baseline controls are household size, a measure of Fall Army Worm presence collected during the 2017 season, baseline maize acres, and baseline maize yields per acre. Regressions on the full sample in columns 2, 4, and 6 additionally control for whether households participated in 1AF in the past. See Appendix E for details on the reduced sample size in columns 3-6.

to deal with attrition (see Appendix E for more details on the latter).

As mentioned above, treatment farmers often only enroll a portion of their land in the program, and almost twenty-five percent of participants enroll all of their maize land.[17] For those who have both program and non-program land, there are many reasons to believe that the program could result in different types of spillover effects. For example, if farmers substitute labor and other complimentary investments towards the enrolled plot, leaving fewer resources for non-program land, then we could see high yields on treated land but reduced productivity on non-program land. If instead there are learning spillovers, whereby farmers for example apply the improved practices that they have learned to their non-program land, or reallocate some inputs to the non-program land, then we might expect yields on both plots to be higher than yields in the control group. Figure 3 suggests that the yield effects are largely driven by productivity increases on enrolled land; treatment group yields on non-enrolled land are statistically indistinguishable from control farmers yields. We will investigate distributional aspects of the treatment effects in the next section. Note that the enrolled plot was not randomly selected, so it is also possible that plot selection affects these results. For example, if participants enroll their best plot, the non-enrolled plot could produce less than the high-quality plot even with positive spillovers.

Figure 4 shows the distribution of treatment plot yields separately for farmers who enroll all of their land and those who enroll a fraction or their land assets. If farmers reallocate a substantial portion of the program-supplied inputs to non-enrolled plots, we would expect to see that yields on enrolled plots are higher for those who do not have alternative land towards which to substitute away from the enrolled plot. We observe the opposite here: participants who enroll all of their land in the program obtain lower yields than those who enroll only part of their land (Kolmogorov-Smirnov test of equality of distributions p-value: 0.01). Of course, this could reflect other differences between those with high- and low amounts of land ownership, or with different perceptions of the benefits of the program. Still, the sign of the difference is suggestive that reallocation towards the non-enrolled plot is not a major factor.

---

[17]See Figure C.1 for the distribution of enrolled land shares and its relationship with overall maize acres
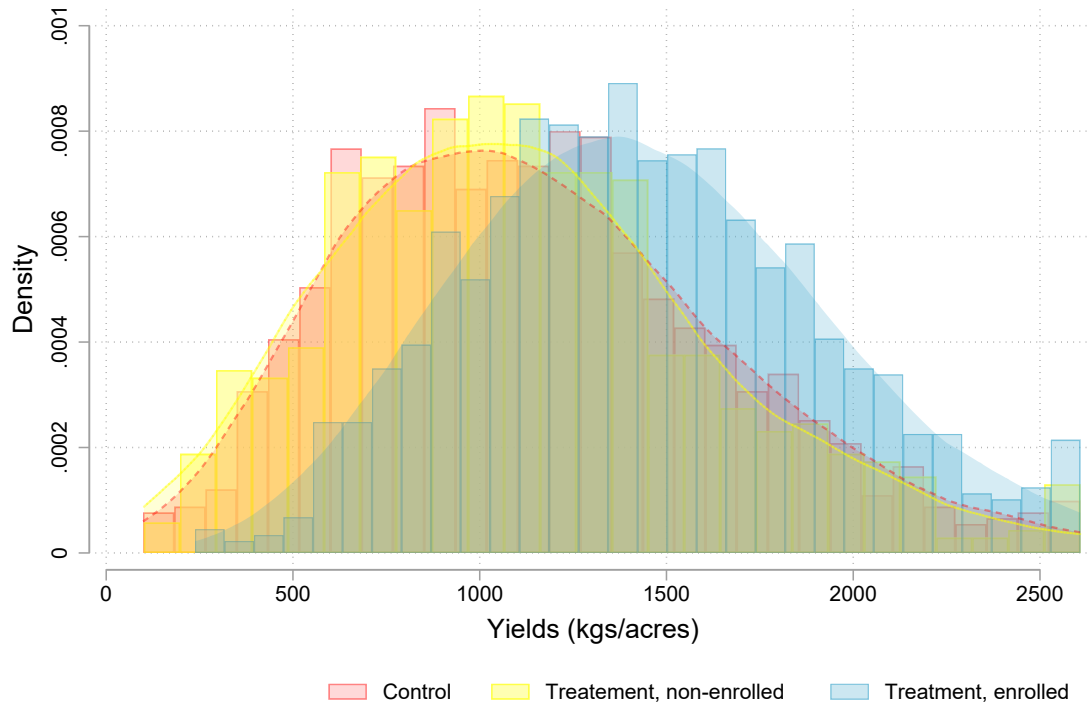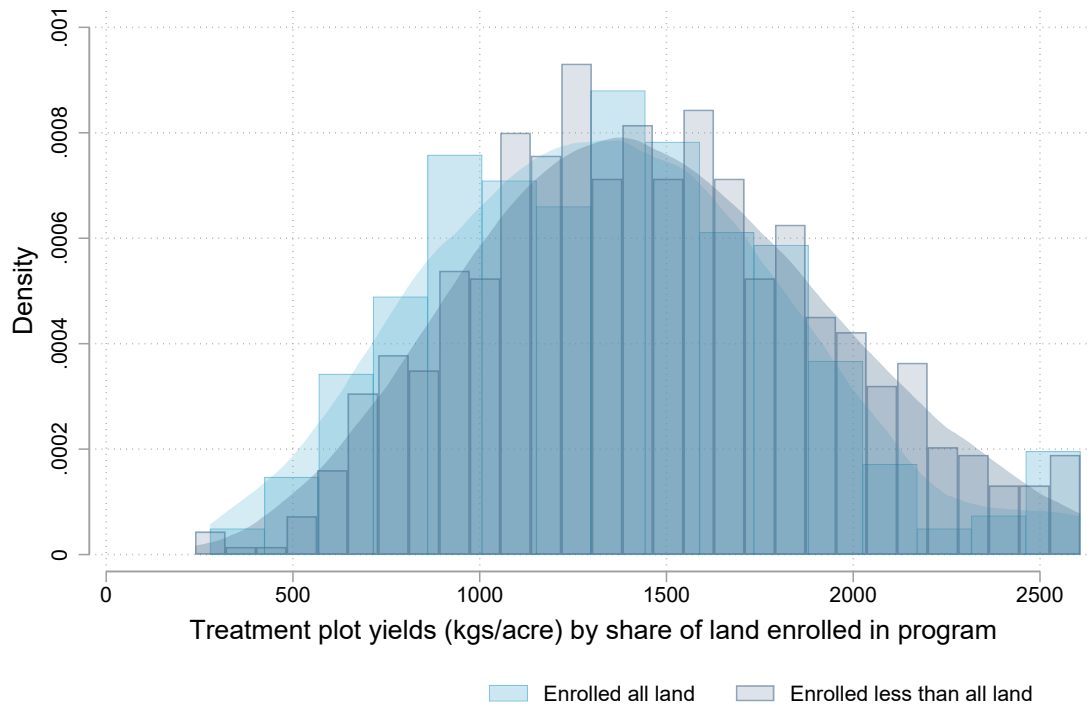
Figure 3: Maize yields by plot type

Figure 4: Treatment yields by fraction of land enrollment

4   Heterogeneity

Although we show that the ATE of the 1AF program is statistically and economically significant for program yields, total maize output, and profits, we may still wonder if this average masks underlying heterogeneity in program impacts. In particular, given that the effect on profits is slightly attenuated when we account for missing data (see Appendix E), it is useful to know whether effects are larger for specific sub-populations.

Figure 5 shows estimates of the treatment effect of program participation on the unconditional quantiles of the outcome distributions. We use the Frölich and Melly (2010) implementation of Firpo (2007). These estimates help us understand whether program participation leads to consistent changes in outcomes along different quantiles of the unconditional outcome distributions.[18] We can see that the treatment effect is remarkably consistent across the distribution of program maize yields. For total maize production and profits, treatment effects for farmers at the top end of the output distribution are substantially lower; at the very top effects become imprecisely estimated and statistically indistinguishable from zero. For these two outcomes, we cannot detect significant treatment effects for the top two deciles of the distribution.

Another way to look at this kind of effect is through distribution regressions, discussed in Chernozhukov et al. (2013). These are closely related to unconditional quantile partial effects but can be easier to interpret since the x-axis is values rather than percentiles of a distribution. Figure 6 shows coefficients from a series of regressions; each coefficient is an estimate of β_x from the following regression:

$$\mathbb{1}(Y > x)_{ij} = \alpha + \beta_x T_i + \delta P_i + \gamma_j + \epsilon_i,$$

where Y is the outcome of interest, x varies along the x-axis of each figure along the support of the outcome variable, T_i is the treatment dummy, and γ_j is a field office area (site) fixed effect. These results are from estimations with a linear probability model, but the results are robust to using a logit model to estimate the threshold probabilities. An appealing aspect of

---

[18]Standard quantile regressions, by contrast, which are computed by taking the horizontal distance between the treatment and control CDFs, assume rank preservation, i.e., require individuals' potential outcomes under treatment and control to preserve their rank order in the distribution. We additionally conduct a Monte Carlo simulation exercise suggested in Appendix E of Heckman et al. (1997). To simulate the distribution of impact standard deviations under the null hypothesis of no heterogeneity, we repeatedly sample the control group to generate synthetic treatment and control groups. This gives us a distribution of the standard deviation of percentile effect differences under the null, which we then compare to the impact standard deviation seen in the data. For all outcomes, we fail to reject the null of no heterogeneity. This suggests that we are unable to detect treatment effect heterogeneity under the assumption of perfect positive dependence between treatment and control outcome percentiles (also called the location shift assumption). Since this is a strong assumption, we implement the rank preservation test proposed in Bitler et al. (2005). This tests for rank reversal in baseline characteristics between quartiles of the treatment and control distribution. For each sample definition and outcome variable, we fail to reject the null in the test for joint-orthogonality. Results of both exercises are available from the authors upon request. Further, the unconditional quantile treatment effects reported here barely differ from standard quantile regressions, which only capture quantile treatment effects if the location shift assumption holds.
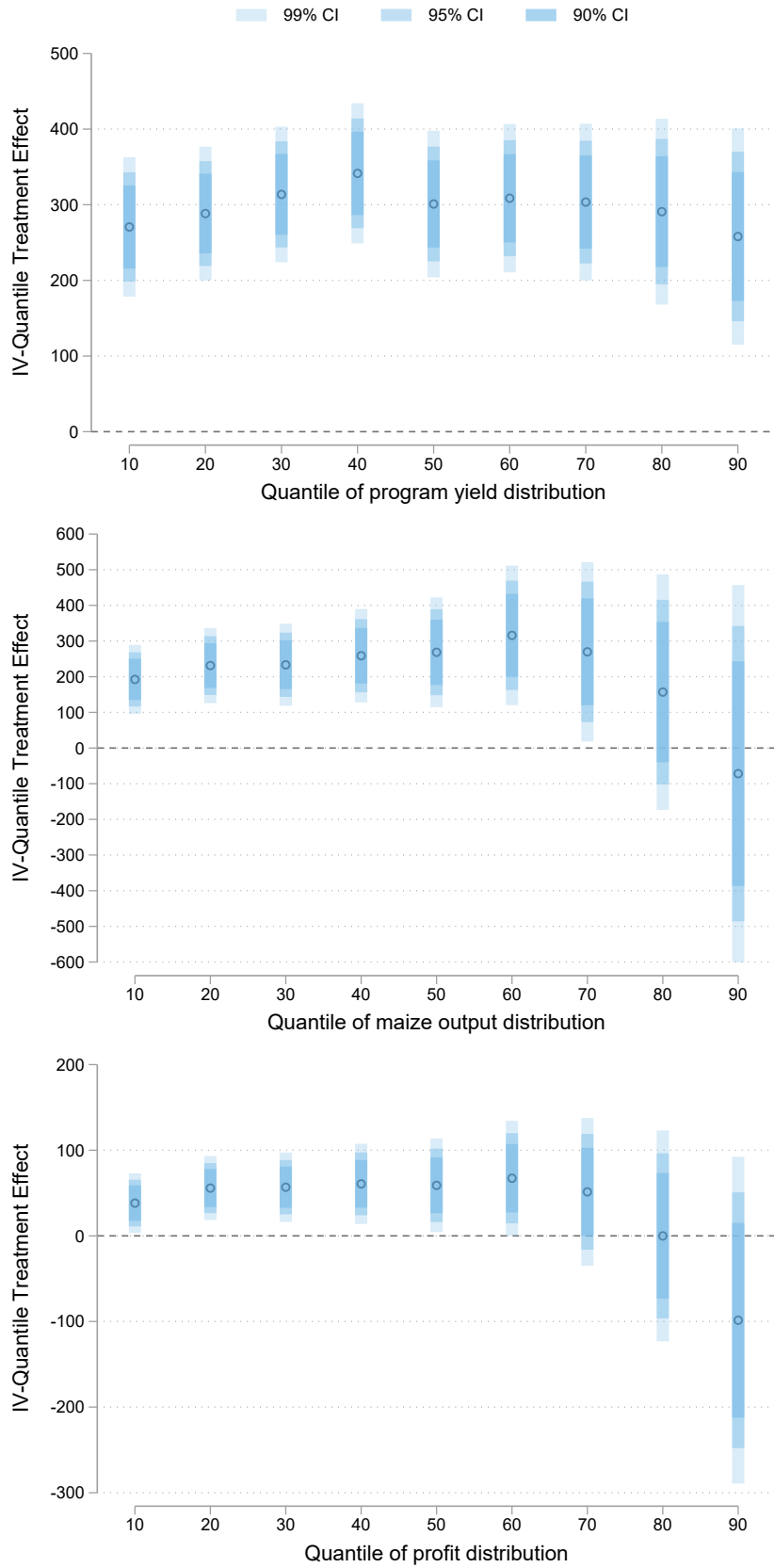
Figure 5: Program effects on unconditional quantiles of the outcome distribution

distribution regressions is that we can pinpoint where along the distribution effects occur.

The top panel shows the effects along different values of maize. We can see that the effect of the treatment on the proportion of participants above a certain yield is substantial, especially between thresholds of 750 kg/acre and 1750 kg/acre. These correspond to the bulk of the control group yield distribution, confirming the take-away from the quantile effect results: the effects are substantial and consistent over most of the yield distribution. The results for maize output and profits suggest that the effects are largest around the lower end of the distribution, but unlike yields, there is no binding upper limit on the total output. While total maize output can at least in theory be expanded by bringing more land under cultivation, maize yields likely have some physiological upper bound, beyond which decreasing marginal returns to inputs start to make additional intensification unprofitable.

Finally, we implement methods proposed by Chernozhukov et al. (2018) (henceforth CDDF) to estimate key features of heterogeneous effects on our outcomes of interest. A key difference between this approach and the previous two is that it focuses on understanding whether or not specific covariates can predict what participants will be most and least affected. While we do not detect much heterogeneity along the distribution of the outcome variable, this does not automatically rule out the existence of subgroups for whom the treatment is more or less effective. As we did not pre-specify sub-group heterogeneity analyses, we wanted to avoid specification searching and decided to use the CDDF method.[19] A key challenge with machine learning tools in high-dimensional settings is that they typically require strong assumptions to produce consistent estimators of conditional average treatment effects (CATE). The new method developed in Chernozhukov et al. (2018) sacrifices some generalizability, but in return the authors are able to rely on fewer assumptions.

In particular, instead of trying to make inference on the full CATE function, the method focuses on making inference on key *features* of the CATE. These features are (1) the Best Linear Predictor (BLP) of the CATE function, (2) Sorted Group Average Treatment Effects (GATES), reporting predicted treatment effects at different deciles of the predicted treatment effect distribution, and (3) Classification Analysis (CLAN), showing how covariates of interest differ between the units that we predict will be the most and least affected, and these most-affected and least-affected groups are also defined by the highest and lowest deciles of the predicted treatment effect distribution. Below, we provide some more intuition for the BLP, as it is perhaps the least obvious of the three (BLP results available from the authors).

Briefly, the method splits the data into an auxiliary subset, separate from the main data (the data is split into main and auxiliary many times, as is standard with ML techniques). Letting Y^0 and Y^1 denote potential outcomes under control and treatment, respectively, we can write out two key functions: $b_0(Z) := E[Y^0|Z]$, which is the baseline conditional average, and $s_0(Z) := E[Y^1|Z] - E[Y^0|Z]$. Given a randomly assigned treatment variable

---

[19]Our discussion of the method below draws heavily on the discussion in Chernozhukov et al. (2018). Section 6.2 in their paper describes the implementation algorithm in detail.
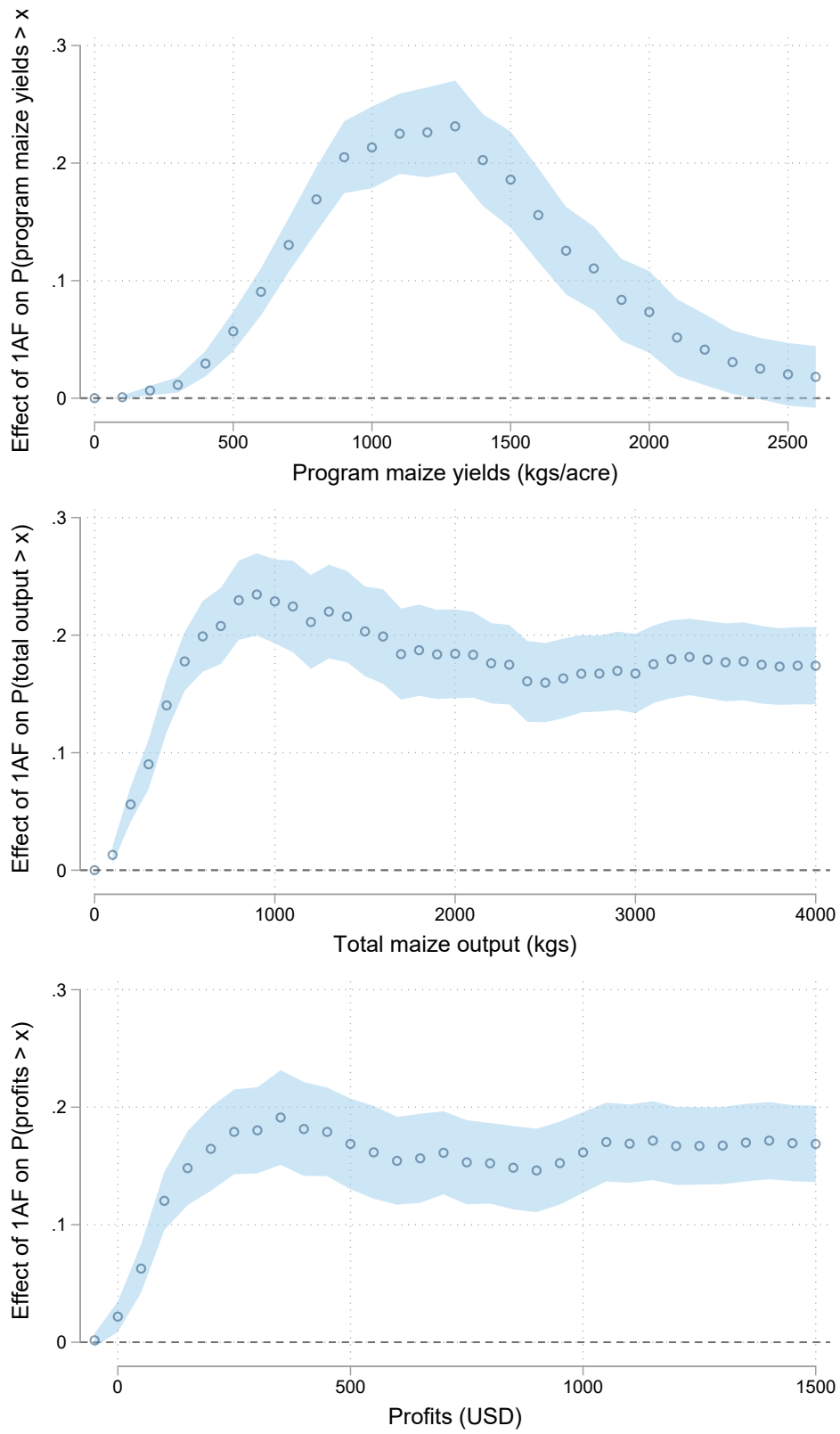
Figure 6: Distribution regressions for main outcome variables

23

D, a known propensity score $p(Z)$, and a few more assumptions on the propensity score, the observed outcome can be written as a regression function (here, conditional on D, Z): $Y = b_0(Z) + Ds_0(Z) + U$, where $E[U|Z, D] = 0$.

We then proceed by using each auxiliary sample to train an ML estimator and obtain ML estimates of the baseline and treatment effects, called proxy scores. We will refer to the estimated proxies of $b_0(Z)$ and $s_0(Z)$ as $B(Z)$ and $S(Z)$, respectively. Note that we can then use these predicted proxies in the main sample to estimate the BLP of the conditional average treatment effect. Essentially, we regress the observed outcome on the treatment variable minus the propensity score (to estimate the average treatment effect, or ATE), and on the treatment variable minus the propensity score *interacted with* deviations of the $S(Z)$ that we estimated in the auxiliary data from the expected value of $S(Z)$ in the main sample. The coefficient on this second interaction term is what provides information about treatment effect heterogeneity. More specifically, we obtain the BLP parameters by estimating the following relationship in the main sample, using weighted OLS:

$$Y_i = \hat{\alpha}' X_{1,i} + \hat{\beta}_1 (D_i - p(Z_i)) + \hat{\beta}_2 (D_i - p(Z_i))(S_i - \mathbb{E}_N S_i) + \hat{\varepsilon}_i \qquad (2)$$

where $S(Z)$ is written as S for simplicity, $\mathbb{E}_N[w(Z_i)\hat{\varepsilon}_i X_i] = 0$ with $w(Z_i) = \{p(Z_i)(1 - p(Z_i))\}^{-1}$. Further, $X_{i,1}$ is constructed as $X_i = [X'_{1,i}, (D_i - p(Z_i)), (D_i - p(Z_i)(S_i - \mathbb{E}_N S_i)]'$, and $X_{1,i}$ includes a constant, $B(Z_i)$, and $S(Z_i)$. In the above regression, the estimated β_1 is the ATE, and β_2 is best linear predictor of the existing heterogeneity. If what we estimate in the auxiliary sample $(S(Z))$ is a perfect proxy for the true heterogeneity, $s_0(Z)$, then $\beta_2 = 1$. If there is no heterogeneity, and the estimates from the ML are pure noise, then $\beta_2 = 0$.

We include as explanatory variables a number of baseline controls that could plausibly be correlated with outcomes: pre-exposure to 1AF program, household size, baseline fertilizer use, asset score, father above secondary education, self-reported credit access), as well as several additional variables: rainfall and temperature by growing season phase (pre-planting, immediate post-planting, and post-top-dressing). The results are largely unaffected by the inclusion of additional variables. We run the method using six ML methods: Neural Nets, Lasso, Ridge, Elastic Net, Boosting, and Random Forest. In our case, our vector of covariates has very low ability to predict heterogeneity; the confidence intervals on $\beta_2 = 0$ are centered at zero, but are also imprecisely estimated. We do not report the results here, but they are available from the authors. This does not tell us with certainty that there exists no heterogeneity in treatment effects, but it does inform us that the vector of covariates included has no power to predict treatment effect heterogeneity.

Figure 7 presents the Sorted Group Average Treatment Effects (GATES) estimated using Neural Nets.[20] We can see that for the lowest-ranked groups, the GATES estimate is not distinguishable from zero other than for program maize yields. That said, the groups are not

---

[20]Results from other ML methods, available on request, are qualitatively similar, but neural nets performed the best in the simulations.

(a) Program maize yields
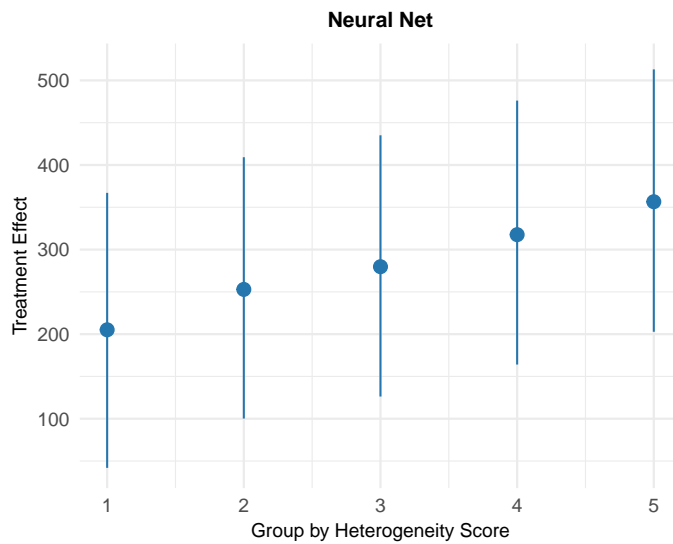
(b) Total maize output
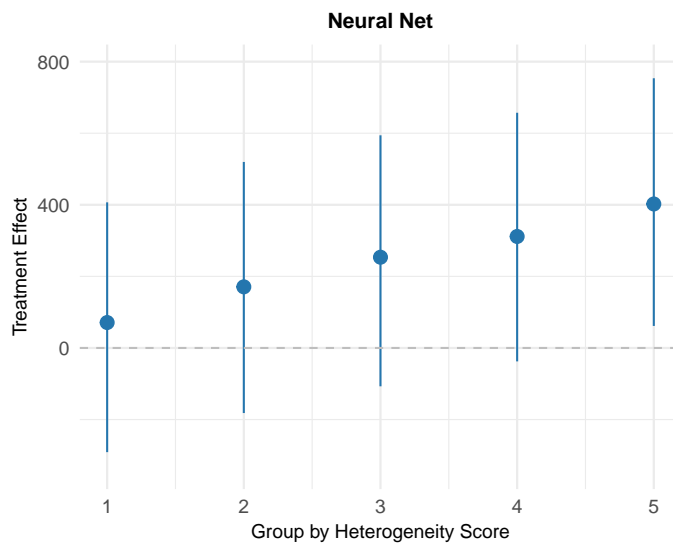
(c) Profits
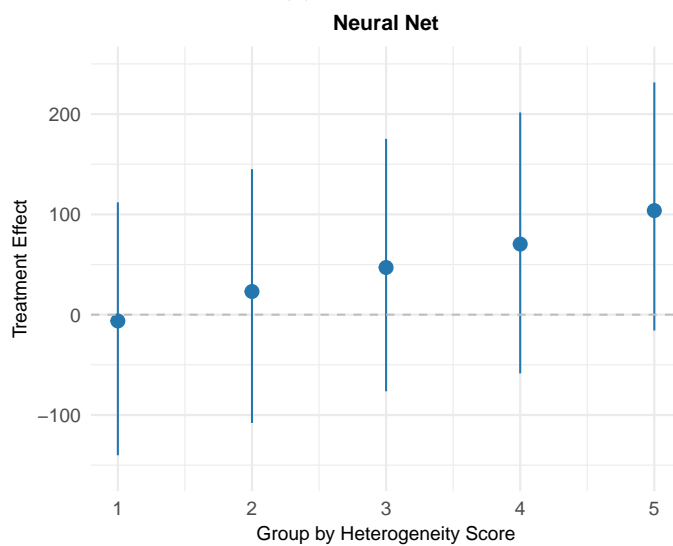
Figure 7: Sorted Group Average Treatment Effects (GATES)

25

starkly different from each other, and if we examine the means of baseline variables across the least-affected and most-affected groups (CLAN), they do not differ significantly in a systematic way (results omitted). It may be that the method requires a larger sample size to have enough power to conclude that the GATES are significantly different from each other. Since it relies on covariates to predict heterogeneity, it could also be that we are simply not including the correct covariates. It is also possible that the program simply works well for the majority of the population that self-select into enrolling.

5   What happens after the treatment year?

An additional means by which we can assess the value of the 1AF program is in how participants themselves value the program. If their experiences are positive, we might expect this to be reflected in whether they choose to re-enroll in 1AF in subsequent years and how much land they enroll. To answer this question, we match farmers to 1AF administrative program data, which includes information about farmers' enrollment history and loan repayment.

Of particular interest is how farmers make the decision to re-enroll in the program. We do not have strong *a priori* expectations about the extensive margin effect. If farmers value the program because of the resources that it provides, farmers may be more likely to re-enroll if they participated in the previous season. However, if the program is effective at helping farmers "graduate" by nudging them across a poverty trap, then we might expect some successful treated farmers to drop out of the program. Table 5 shows that being randomly allocated to participation in 2017 did not significantly increase the probability that farmers enrolled in 2018. This is somewhat surprising, but the farmers who were held out in the control group received a discount on their participation fee in 2018 to compensate for having to wait a year due to the study. This could have off-set the effect of being exposed to the program.

In addition to the decision at the extensive margin, farmers also face an additional choice: how much land to enroll. If farmers find it useful to enroll more land to access more credit and/or larger quantities of high-quality inputs, then they may increase enrollment year-on-year after their first year of participation. However, if the program is more useful to farmers for its information effects, then we may not expect any change in land enrolled. We know from Table 1 that farmers in treatment and control groups were not significantly different in baseline maize acres cultivated. We might therefore expect that treated and control farmers on average would choose to enroll the same amount of land in the program in 2018.

We first consider the primary sample, who had no previous exposure to the 1AF program. Table 6 shows that being randomly assigned to participate in the program in 2017 significantly increased land enrolled in 2018. Column (1) includes only the sample of farmers who did

Table 5: Enrolled in 2018, primary and full samples

	(1) Enrolled 2018 (primary sample)	(2) Enrolled 2018 (full sample)
1AF participant	0.01	0.01
	(0.03)	(0.03)
Pre-exposed		0.07**
		(0.03)
1AF participant		-0.02
× pre-exposed		(0.03)
Observations	757	2122
R^2	0.343	0.446
Control group mean	0.66	0.67

Results in this table are from linear regressions of the outcome variables on the treatment dummy. The outcome variable is a binary indicator of whether the farmer appeared in the 2018 administrative data as enrolled in the program. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. All regressions include field office (site) fixed effects.

Table 6: Acres enrolled in 2018, primary sample

	(1) 2018 Acres	(2) 2018 Acres, Missings=0
1AF participant	0.13***	0.09**
	(0.04)	(0.04)
Observations	487	757
R^2	0.131	0.232
Control group mean	0.62	0.41

Results in this table are from linear regressions of the outcome variables on the treatment dummy. Missing enrolled acres in (2) are coded as zero, so this table includes farmers who chose not to re-enroll. Standard errors (in parentheses) clustered at treatment assignment (farmer group cluster) level. All regressions include field office area (site) fixed effects.

choose to re-enroll, whereas column (2) adds farmers who did not re-enroll by treating their enrolled acres as zero. This second column should be a very conservative estimate of the effect. This lends further support to the idea that farmers find the access to credit and quality inputs useful. We might also interpret these results as suggesting that program participation in one year may be easing constraints, allowing farmers to increase enrollment in the next year. Alternatively, it could suggest a learning mechanism, whereby farmers are more interested in enrolling more of their land having experimented with the program in the past year.

We try to go a bit further to distinguish between eased constraints and farmer learning as explanations for increased land enrollment by treatment farmers. In column (1) of Table 7, we show enrollment decisions in 2018 by farmers who joined 1AF in 2016. Although initial enrollment was non-random, control farmers were randomly held out from the program in 2017. Thus, exposure to one vs two years of the program is random. The effect of 2017 participation on 2018 acres enrolled is positive and significant. This sample was already aware of how the program worked, having participated in 2016, so we take this as suggestive evidence that the mechanism that leads farmers to increase enrollment in the next year runs through a relaxation of constraints (either by increasing yields and profits, or by continuing to allow farmers to update their beliefs about the returns to the program).

Table 7: Acres enrolled in 2018, from past new enrollees

	(1) 2018 Acres (Joined 2016)	(2) 2018 Acres (Joined 2015)
1AF participant	0.23***	0.08**
	(0.04)	(0.04)
Observations	394	293
R^2	0.180	0.136
Control group mean	0.58	0.62

Results in this table are from linear regressions of the outcome variables on the treatment dummy. The outcome variable is the enrolled land size recorded in the 2018 administrative data. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. All regressions include field office (site) fixed effects.

6   Discussion

We uncover large ATEs from participation in 1AF's core program, which seem to vary relatively little across the population, across a variety of heterogeneity tests. A large theoretical literature on poverty traps, and a growing empirical body of evidence shows that poor households may need bundled interventions in order to move out of poverty (Bandiera et al., 2017; Banerjee et al., 2015). If we believe that farmers face multiple simultaneous constraints, offering an intervention that only relaxes credit constraints, or only provides

information, may not be sufficient to raise yields and profits in a significant or economically meaningful way.

A reader could of course question whether this bundled approach is the most cost-effective one. Could a simpler program work as well at lower cost? The most common criticism against bundled programs tends to be aimed at donor-funded and government programs. The 1AF model, in contrast, is largely farmer-funded. By revealed preference, given the rates of re-enrollment in the program and the high rates of loan repayments, it seems that farmers experience a substantial return on their investments.

More generally, it is plausible that bundled programs that address multiple constraints may be effective across a greater number of contexts if the nature and severity of market failures varies enough across space. In specific contexts where implementers have identified the most binding constraint, it is likely more cost-effective to focus on a single, highly effective component. However, a standardized program that addresses multiple constraints may enable an organization to scale interventions without having to tweak program specifics extensively in each new context. While opponents may contend that more complex programs are costlier to implement than more targeted interventions, this argument ignores the potential cost of the market research that would be required to tailor targeted programs to suit many diverse new contexts. Determining what the binding constraints are in many small, local markets could in many cases be prohibitively costly.

Returning to the broader question of the role of the agricultural sector in Africa's future, and whether or not smallholder farmers can help drive growth, we of course recognize the limits of a single study in a particular region in a specific year. A deeper discussion of external validity (even as it applies to extrapolating to other 1AF locations) would be valuable, but is beyond the scope of the current paper. That said, if we review the 1AF approach—and that of the growing number of similar programs—in light of the critiques outlined in Collier and Dercon (2014), several aspects of the program can be interpreted as trying to address their concerns about the (lack of) promise of smallholder farming systems.

Collier and Dercon (2014) identify three areas of potential economies of scale in SSA agriculture: first, they discuss skills and technology. They argue that in agriculture, as in most sectors, larger producers are more likely to successfully handle new information, process it, and manage adoption risks. Large organizations may also be better able to internalize learning costs. Large non-profit organizations like 1AF could plausibly absorb some of these costs, leveraging their scale and human capital to distill information and channel the relevant information to smallholders when and where they need it.[21] A second reason for economies of scale is finance costs, since scale influences both the costs of obtaining capital, logistics,

---

[21]Tinsley and Agapitova (2018) report that 1AF launched an initiative called "Tubura University," described as a set of in-house development courses to provide its staff with training in English, computing, leadership and management skills. The organization has also put together a "scale innovations team" with the goal of exploring ways to increase client density, run research projects to build organizational knowledge, and to propose changes to the model as the competitive environment develops.

and bargaining power. Large-scale NGOs and other organizations can and do leverage grant funding for working capital from both agro-input suppliers and banks (Tinsley and Agapitova, 2018). Most financial institutions shy away from input loans, preferring to provide cash credit over shorter time periods; by extending their credit to smallholders, organizations like 1AF can effectively reduce the transaction costs and asymmetric information facing small farmers. Third, economies of scale matter for trading, marketing, and storage. By integrating several parts of the value chain (input importing and quality checks, as well as storage and market facilities), it would seem like 1AF and other large organizations can share the benefits of scale with their clients.

Perhaps this new evidence will nudge a few cynics into reconsidering the future for (smallholder) agriculture in sub-Saharan Africa. For those already optimistic about the sector, we hope that it provides compelling input into a discussion about optimal instruments for boosting productivity. Thinking about large-scale NGOs as enabling creative vertically-integrated opportunities for farmers may hold a clue to the types of investments needed to transform African agriculture into a more dynamic sector.

References

ABAY, K. A., L. E. M. BEVIS, AND C. B. BARRETT (2019): "Measurement Error Mechanisms Matter: Agricultural Intensification with Farmer Misperceptions and Misreporting," Working Paper.

AKER, J. C. (2011): "Dial "A" for Agriculture: A Review of Information and Communication Technologies for Agricultural Extension in Developing Countries," *Agricultural Economics*, 42, 631–647.

AKROYD, S. AND L. SMITH (2007): "Review of Public Spending to Agriculture," Report, DFID/World Bank.

ANDERSON, J. R. AND G. FEDER (2007): "Chapter 44 Agricultural Extension," in *Handbook of Agricultural Economics*, ed. by R. Evenson and P. Pingali, Elsevier, vol. 3 of *Agricultural Development: Farmers, Farm Production and Farm Markets*, 2343–2378.

BANDIERA, O., R. BURGESS, N. DAS, S. GULESCI, I. RASUL, AND M. SULAIMAN (2017): "Labor Markets and Poverty in Village Economies," *The Quarterly Journal of Economics*, 132, 811–870.

BANERJEE, A., E. DUFLO, N. GOLDBERG, D. KARLAN, R. OSEI, W. PARIENTE, J. SHAPIRO, B. THUYSBAERT, AND C. UDRY (2015): "A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries," *Science*, 348, 1260799–1260799.

BEAMAN, L., D. KARLAN, B. THUYSBAERT, AND C. UDRY (2013): "Profitability of Fertilizer: Experimental Evidence from Female Rice Farmers in Mali," *American Economic Review: Papers and Proceedings*, 103, 381–86.

BIRKHAEUSER, D., R. E. EVENSON, AND G. FEDER (1991): "The Economic Impact of Agricultural Extension: A Review," *Economic Development and Cultural Change*, 39, 607–650.

BITLER, M. P., J. B. GELBACH, AND H. W. HOYNES (2005): "Distributional Impacts of the Self-Sufficiency Project," Working Paper 11626, National Bureau of Economic Research.

BLOCK, S. (2014): "The Decline and Rise of Agricultural Productivity in Sub-Saharan Africa since 1961," in *African Successes*, University of Chicago Press, for the National Bureau of Economic Research, 13–67.

BOLD, T., K. C. KAIZZI, J. SVENSSON, AND D. YANAGIZAWA-DROTT (2017): "Lemon Technologies and Adoption: Measurement, Theory and Evidence from Agricultural Markets in Uganda," *The Quarterly Journal of Economics*, 132, 1055–1100.

BRAVO-ORTEGA, C. AND D. LEDERMAN (2005): "Agriculture and National Welfare around the World: Causality and International Heterogeneity since 1960," Working paper 3499, The World Bank.

BYERLEE, D., A. DE JANVRY, AND E. SADOULET (2009): "Agriculture for Development: Toward a New Paradigm," *Annual Review of Resource Economics*, 1, 15–31.

CARLETTO, C., S. SAVASTANO, AND A. ZEZZA (2013): "Fact or Artifact: The Impact of Measurement Errors on the Farm Size–Productivity Relationship," *Journal of Development Economics*, 103, 254–261.

CARTER, M. R., E. TJERNSTRÖM, AND P. TOLEDO (2019): "Heterogeneous Impact Dynamics of a Rural Business Development Program in Nicaragua," *Journal of Development Economics*, 138, 77–98.

CHERNOZHUKOV, V., M. DEMIRER, E. DUFLO, AND I. FERNANDEZ-VAL (2018): "Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments," Working Paper 24678, National Bureau of Economic Research.

CHERNOZHUKOV, V., I. FERNÁNDEZ-VAL, AND B. MELLY (2013): "Inference on Counterfactual Distributions," *Econometrica*, 81, 2205–2268.

COLLIER, P. AND S. DERCON (2014): "African Agriculture in 50 Years: Smallholders in a Rapidly Changing World?" *World Development*, 63, 92–101.

DESIERE, S. AND D. JOLLIFFE (2018): "Land Productivity and Plot Size: Is Measurement Error Driving the Inverse Relationship?" *Journal of Development Economics*, 130, 84–98.

DUBEY, P. AND R. N. YEGBEMEY (2017): "Technical Support to the Impact Evaluation of the Core One Acre Fund Program on Yields and Profits of Maize Farmers in Teso, Kenya," Field Report, International Initiative for Impact Evaluation.

DUFLO, E., M. KREMER, AND J. ROBINSON (2008): "How High Are Rates of Return to Fertilizer? Evidence from Field Experiments in Kenya," *American Economic Review: Papers and Proceedings*, 98, 482–488.

FEDER, G., R. E. JUST, AND D. ZILBERMAN (1985): "Adoption of Agricultural Innovations in Developing Countries: A Survey," *Economic Development and Cultural Change*, 33, 255–298.

FIRPO, S. (2007): "Efficient Semiparametric Estimation of Quantile Treatment Effects," *Econometrica*, 75, 259–276.

FRÖLICH, M. AND B. MELLY (2010): "Estimation of Quantile Treatment Effects with Stata," *The Stata Journal*, 10, 423–457.

GOLLIN, D., C. W. HANSEN, AND A. WINGENDER (2018): "Two Blades of Grass: The Impact of the Green Revolution," Working Paper 24744, National Bureau of Economic Research.

GOURLAY, S., T. KILIC, AND D. LOBELL (2017): "Could the Debate Be Over? Errors in Farmer-Reported Production and Their Implications for the Inverse Scale-Productivity Relationship in Uganda," Working paper 8192, The World Bank.

GOVERNMENT OF KENYA (2010): "Agricultural Sector Development Strategy 2010-2020," Technical Report.

HANNA, R., S. MULLAINATHAN, AND J. SCHWARTZSTEIN (2014): "Learning Through Noticing: Theory and Evidence from a Field Experiment," *The Quarterly Journal of Economics*, 129, 1311–1353.

HECKMAN, J. J., J. SMITH, AND N. CLEMENTS (1997): "Making The Most Out Of Programme Evaluations and Social Experiments: Accounting For Heterogeneity in Programme Impacts," *The Review of Economic Studies*, 64, 487–535.

KONDYLIS, F., V. MUELLER, AND J. ZHU (2017): "Seeing Is Believing? Evidence from an Extension Network Experiment," *Journal of Development Economics*, 125, 1–20.

LIGON, E. A. AND E. SADOULET (2011): "Estimating the Effects of Aggregate Agricultural Growth on the Distribution of Expenditures," Working paper 1115, Department of Agricultural & Resource Economics, UC Berkeley.

MAGRUDER, J. R. (2018): "An Assessment of Experimental Evidence on Agricultural Technology Adoption in Developing Countries," *Annual Review of Resource Economics*.

MATSUYAMA, K. (1992): "Agricultural Productivity, Comparative Advantage, and Economic Growth," *Journal of Economic Theory*, 58, 317–334.

MCARTHUR, J. W. AND G. C. MCCORD (2017): "Fertilizing Growth: Agricultural Inputs and Their Effects in Economic Development," *Journal of Development Economics*, 127, 133–152.

MUYANGA, M. AND T. S. JAYNE (2006): "Agricultural Extension in Kenya: Practice and Policy Lessons," Working Paper 202617, Egerton University, Tegemeo Institute of Agricultural Policy and Development.

NATIONAL FARMERS INFORMATION SERVICES (2019): "Field Management – NAFIS," Report, http://www.nafis.go.ke/agriculture/maize/field-management-practices/.

ONE ACRE FUND (2019): "How We Grow," Report, https://oneacrefund.org/what-we-do/how-we-grow/.

Ravallion, M. and S. Chen (2007): "China's (Uneven) Progress against Poverty," *Journal of Development Economics*, 82, 1–42.

Rostow, W. W. (1990): *The Stages of Economic Growth: A Non-Communist Manifesto*, Cambridge University Press.

Sheahan, M., R. Black, and T. S. Jayne (2013): "Are Kenyan Farmers Under-Utilizing Fertilizer? Implications for Input Intensification Strategies and Research," *Food Policy*, 41, 39–52.

Suri, T. (2011): "Selection and Comparative Advantage in Technology Adoption," *Econometrica*, 79, 159–209.

Tinsley, E. and N. Agapitova (2018): "Private Sector Solutions to Helping Smallholders Succeed: Social Entreprise Business Models in the Agriculture Sector," Report, The World Bank.

Tjernström, E., M. R. Carter, and T. Lybbert (2018): "The Dirt on Dirt: Soil Characteristics and Variable Fertilizer Returns in Kenyan Maize Systems," Working Paper.

Udry, C., F. di Battista, M. Fosu, M. Goldstein, A. Gurbuz, D. Karlan, and S. Kolavalli (2019): "Information, Market Access and Risk: Addressing Constraints to Agricultural Transformation in Northern Ghana," Draft Report.

World Bank (2008): "World Development Report 2008: Agriculture for Development," Report, World Bank.

Appendix

A Pre-Analysis Plan Review Letter

3ie International Initiative for Impact Evaluation

New Delhi, 23rd July 2018

To whom it may concern

Reg.: Confirmation of the review of the Pre-Analysis Plan (PAP)

Dear Sir/Madam

This letter is to confirm that 3ie reviewed the PAP of the Impact Evaluation of the One Acre Fund program on yields and profits of maize and beans farmers in Teso, Kenya.

The PAP was submitted to 3ie by Maya Duru and Kim Siegal. The PAP review process was led by Rosaine N. Yegbemey between July and November, 2016.

The PAP went through three main rounds of review with several iterations of comments and a couple of Skype calls. Considering the context of the evaluation and the 1AF team's responses to the comments, the revised PAP was found to be appropriate to the goals of the study and of sufficient level or rigor.

For any questions on the PAP review process, please contact Rosaine N. Yegbemey at ryegbemey@3ieimpact.org.

Marie Gaarder
Director of Evaluation Office and Global Director for Innovation and Country Engagement
International Initiative for Impact Evaluation (3ie)

B   Additional baseline results
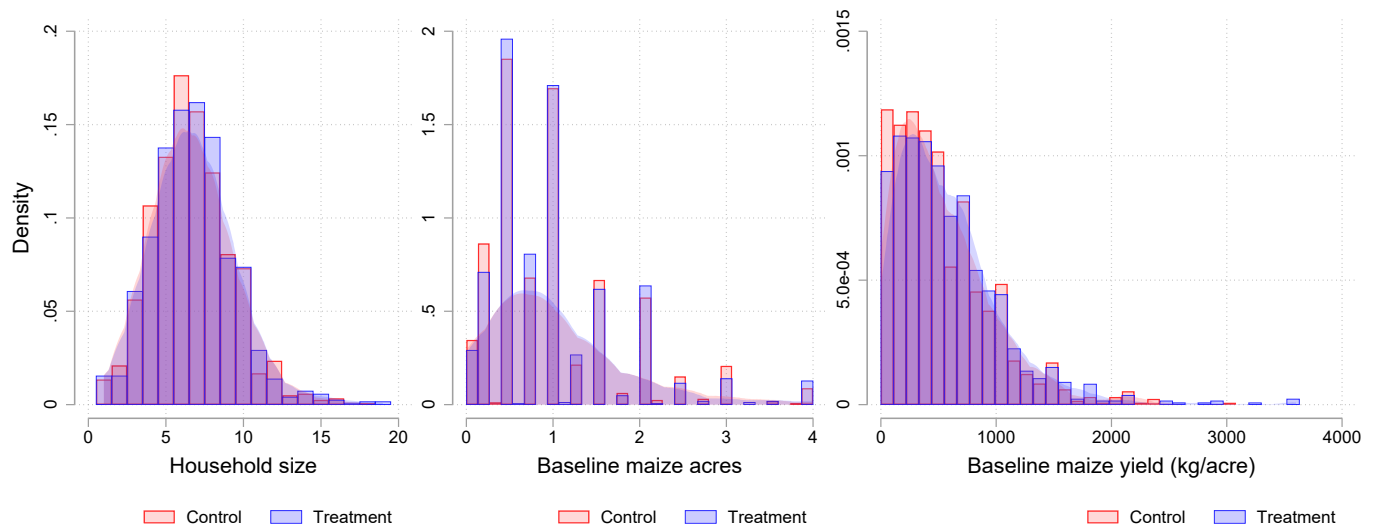
Figure B.1: Distributions of non-binary baseline characteristics, by treatment status
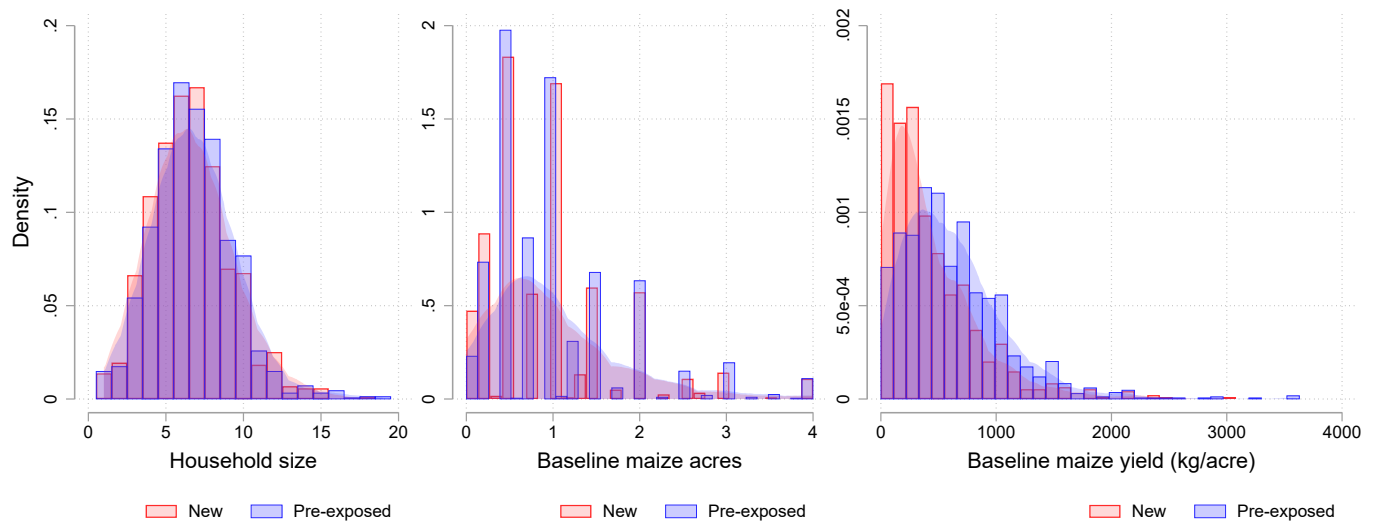
Figure B.2: Distributions of non-binary baseline characteristics, by previous 1AF exposure

Table B.1: Baseline balance (primary sample)

	(1)		(2)		
	Control		Treatment		Difference
Variable	N/[Clusters]	Mean/SE	N/[Clusters]	Mean/SE	(1)-(2)
Panel A: Binary variables					
Married (0/1)	396	0.904	361	0.881	0.023
	[58]	(0.016)	[59]	(0.016)	
Household head has secondary school	396	0.346	361	0.377	-0.031
	[58]	(0.030)	[59]	(0.026)	
Household income >50% from farm labor	396	0.788	361	0.753	0.034
	[58]	(0.032)	[59]	(0.027)	
Used improved ag technology in 2016	396	0.619	361	0.651	-0.032
	[58]	(0.037)	[59]	(0.031)	
Reports knowledge of 1AF practices	396	0.061	361	0.141	-0.081***
	[58]	(0.015)	[59]	(0.020)	
Intercropped maize and beans in 2016	396	0.472	361	0.560	-0.087
	[58]	(0.034)	[59]	(0.041)	
Reports having credit access in 2016	396	0.705	361	0.709	-0.005
	[58]	(0.030)	[59]	(0.028)	
Panel B: Continuous variables					
Household size	396	6.616	361	6.662	-0.046
	[58]	(0.154)	[59]	(0.153)	
Acres under maize cultivation in 2016	396	0.994	361	0.936	0.058
	[58]	(0.064)	[59]	(0.050)	
Maize yield (kg/acre) in 2016	396	427.257	361	443.707	-16.450
	[58]	(32.866)	[59]	(35.653)	
F-statistic (test of joint significance)					2.066**
Number of observations					757

Notes: Field office fixed effects are included in all estimation regressions. Standard errors clustered at farmer group cluster level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table B.2: Baseline balance (pre-exposed sample)

Variable	(1) Control N/[Clusters]	(1) Control Mean/SE	(2) Treatment N/[Clusters]	(2) Treatment Mean/SE	Difference (1)-(2)
Panel A: Binary variables					
Married (0/1)	660	0.862	705	0.879	-0.017
	[60]	(0.015)	[59]	(0.013)	
Household head has secondary school	660	0.409	705	0.451	-0.042
	[60]	(0.024)	[59]	(0.030)	
Household income >50% from farm labor	660	0.771	705	0.791	-0.020
	[60]	(0.021)	[59]	(0.020)	
Used improved ag technology in 2016	660	0.883	705	0.879	0.004
	[60]	(0.014)	[59]	(0.014)	
Reports knowledge of 1AF practices	660	0.712	705	0.723	-0.011
	[60]	(0.024)	[59]	(0.023)	
Intercropped maize and beans in 2016	660	0.477	705	0.434	0.043
	[60]	(0.032)	[59]	(0.030)	
Reports having credit access in 2016	660	0.712	705	0.735	-0.023
	[60]	(0.023)	[59]	(0.023)	
Panel B: Continuous variables					
Household size	660	6.665	705	6.901	-0.236
	[60]	(0.122)	[59]	(0.113)	
Acres under maize cultivation in 2016	660	1.010	705	1.068	-0.059
	[60]	(0.040)	[59]	(0.040)	
Maize yield (kg/acre) in 2016	660	599.304	705	649.640	-50.336**
	[60]	(32.313)	[59]	(30.483)	
F-statistic (test of joint significance)					1.300
Number of observations					1365

Notes: Field office fixed effects are included in all estimation regressions. Standard errors clustered at farmer group cluster level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table B.3: Baseline comparison of primary and pre-exposed samples

	(1) Primary Sample		(2) Pre-Exposed Sample		Difference
Variable	N/[Clusters]	Mean/SE	N/[Clusters]	Mean/SE	(1)-(2)
Panel A: Binary variables					
Married (0/1)	757	0.893	1365	0.871	0.022
	[117]	(0.011)	[119]	(0.010)	
Household head has secondary school	757	0.361	1365	0.431	-0.070**
	[117]	(0.020)	[119]	(0.019)	
Household income >50% from farm labor	757	0.771	1365	0.782	-0.010**
	[117]	(0.021)	[119]	(0.015)	
Used improved ag technology in 2016	757	0.634	1365	0.881	-0.247***
	[117]	(0.025)	[119]	(0.010)	
Reports knowledge of 1AF practices	757	0.099	1365	0.718	-0.619***
	[117]	(0.013)	[119]	(0.017)	
Intercropped maize and beans in 2016	757	0.514	1365	0.455	0.059***
	[117]	(0.027)	[119]	(0.022)	
Reports having credit access in 2016	757	0.707	1365	0.724	-0.017***
	[117]	(0.020)	[119]	(0.016)	
Panel B: Continuous variables					
Household size	757	6.638	1365	6.787	-0.149**
	[117]	(0.108)	[119]	(0.083)	
Acres under maize cultivation in 2016	757	0.966	1365	1.040	-0.073***
	[117]	(0.041)	[119]	(0.029)	
Maize yield (kg/acre) in 2016	757	435.102	1365	625.302	-190.200***
	[117]	(24.250)	[119]	(22.143)	
F-statistic (test of joint significance)					86.345***
Number of observations					2122

Notes: Field office fixed effects are included in all estimation regressions. Standard errors clustered at farmer group cluster level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.
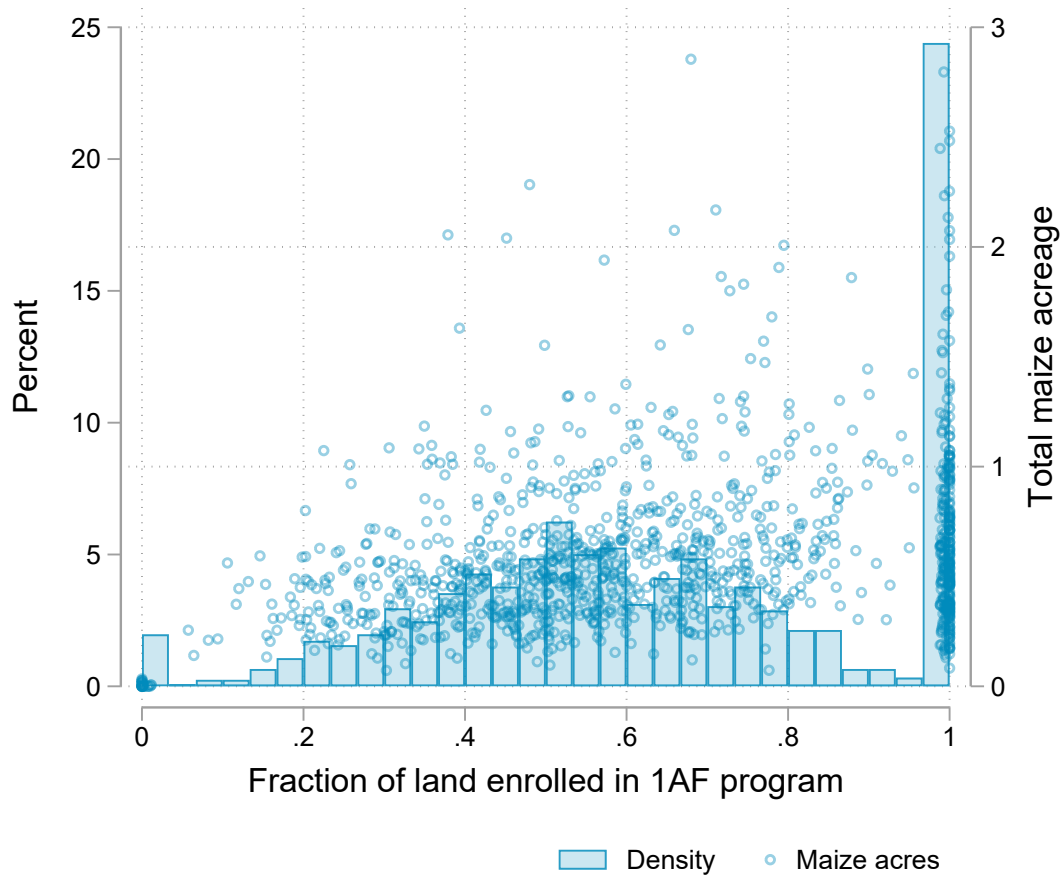
C Additional results

Figure C.1: Share of land that treatment farmers enroll in the program

Table C.1: Primary outcomes with controls, primary sample

	(1) Program Maize Yields	(2) Total Maize Output	(3) Profit
1AF participant	318.90***	264.92***	54.80**
	(36.05)	(75.75)	(27.22)
Married (0/1)	112.42*	19.75	3.80
	(59.08)	(147.01)	(53.83)
Household head has	-0.76	259.84***	86.90***
secondary school	(45.96)	(88.10)	(32.70)
Household income	7.86	24.51	16.12
>50% from farm labor	(47.17)	(79.03)	(28.78)
Used improved ag	-94.27**	-142.42*	-61.67**
technology in 2016	(42.62)	(85.34)	(30.87)
Reports knowledge of	-10.18	25.51	-8.90
1AF practices	(69.24)	(145.85)	(53.47)
Intercropped maize	-23.53	-82.25	-24.97
and beans in 2016	(43.68)	(90.91)	(33.36)
Reports having	-90.85**	-45.97	-19.97
credit access in 2016	(35.37)	(78.80)	(28.12)
Household size	8.06	33.46**	12.99**
	(7.44)	(15.28)	(5.34)
FAW Incidence	-5.98*	1.79	1.37
	(3.52)	(7.71)	(2.87)
Acres under maize	69.31**	361.85***	107.06***
cultivation in 2016	(26.98)	(73.64)	(26.21)
Maize yield	0.18***	0.45***	0.14***
(kg/acre) in 2016	(0.05)	(0.10)	(0.04)
Observations	757	691	691
R^2	0.198	0.248	0.207

Results in this table are from linear regressions of the outcome variables on the treatment dummy. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. All regressions include field office (site) fixed effects.

Table C.2: Primary outcomes with controls, full sample

	(1) Program Maize Yields	(2) Total Maize Output	(3) Profit
1AF participant	290.61***	200.22***	28.48*
	(23.35)	(44.76)	(16.74)
Past 1AF participant	-28.73	20.07	7.90
	(27.12)	(66.28)	(23.54)
Married (0/1)	65.57**	69.06	14.47
	(31.89)	(68.69)	(25.54)
Household head has	6.76	181.69***	60.21***
secondary school	(25.00)	(51.81)	(18.85)
Household income	45.89*	132.78**	53.98***
>50% from farm labor	(26.35)	(52.30)	(19.40)
Used improved ag	-29.88	-58.41	-28.68
technology in 2016	(33.70)	(63.26)	(23.16)
Reports knowledge of	21.80	74.35	17.53
1AF practices	(28.99)	(55.58)	(20.85)
Intercropped maize	-13.86	-19.31	0.63
and beans in 2016	(24.37)	(56.28)	(20.73)
Reports having	-23.96	-17.76	-14.26
credit access in 2016	(22.32)	(46.40)	(17.13)
Household size	4.81	42.29***	14.21***
	(4.31)	(9.72)	(3.51)
FAW Incidence	-3.99*	3.84	1.44
	(2.12)	(4.18)	(1.56)
Acres under maize	33.11**	368.87***	114.25***
cultivation in 2016	(15.68)	(54.14)	(19.47)
Maize yield	0.14***	0.36***	0.11***
(kg/acre) in 2016	(0.03)	(0.08)	(0.03)
Observations	2122	1919	1919
R^2	0.164	0.214	0.170

Results in this table are from linear regressions of the outcome variables on the treatment dummy. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. All regressions include field office (site) fixed effects.

D   Variable construction and measurement

Maize yield measurements were taken by enumerators on two 40-square-meter areas selected from farmer plots. Enumerators marked each area before farmers harvested any maize. For treated farmers, two areas were marked each on enrolled and non-enrolled land. Additionally, if farmers planned to harvest green maize from some parts of their plot, enumerators marked areas within these parts for measurement. Yield variables use the average dry weight from the two marked areas. For treated farmers, when we consider farmer-level outcomes, we average yields across enrolled and non-enrolled land, weighted by the proportion of land farmers enrolled.

We calculate projected revenues using average market prices from nearby vendors covering post-harvest months, multiplied by 1.08 to account for typical price increases over the consumption/selling season. We calculate farmer costs using program costs and self-reported costs for treated farmers, and self-reported seed and fertilizer costs for control farmers. Labor costs include land prep, plowing, and planting costs, collected in a survey after planting, as well as post-planting costs collected at harvest time. For paid labor, we use farmer self-reported costs by planting phase. To include the opportunity cost of unpaid labor use, we calculate the mean day wage reported within the sample, devalue this mean wage by 50% (roughly the rural unemployment rate according to DHS data), and multiply this devalued mean by total person-days of unpaid labor for each planting phase. Profit is simply the difference between projected farmer revenues and costs.

Note that in the original PAP, all labor costs were specified to be devalued by 50%. We feel that it is more appropriate to only make this correction for unpaid labor, as this more appropriately reflects the expected wage a household laborer could earn in the market. However, we show in Table D.1 and D.2 that our profit results are robust to valuing all labor at the market rate, although unsurprisingly the treatment effect does decrease slightly in magnitude.

Table D.1: Profit with different labor cost definitions, primary sample

	(1) Profit (PAP def)	(2) Profit (Mkt + 50% own)	(3) Profit (Mkt rate labor)
1AF participant	57.435**	54.800**	48.865*
	(27.497)	(27.224)	(27.119)
Observations	691	691	691
R^2	0.215	0.207	0.203
Control Mean Dep. Var	353.189	337.624	323.719

This table presents results from linear regressions of the outcomes in each column on the treatment dummy. Column (1) uses the pre-registered definition of profit which devalued all labor costs from market rate by 50% following local unemployment rate estimates. Column (2) uses our preferred definition, which only devalues unpaid labor at 50% of the market rate. Column (3) values all labor at the market rate. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. All regressions include field office (site) fixed effects and the full set of pre-specified controls.

Table D.2: Profit with different labor cost definitions, full sample

	(1) Profit (PAP def)	(2) Profit (Mkt + 50% own)	(3) Profit (Mkt rate labor)
1AF participant	31.617*	28.476*	21.871
	(16.756)	(16.739)	(16.627)
Observations	1919	1919	1919
R^2	0.177	0.170	0.167
Control Mean Dep. Var	379.453	362.975	349.813

This table presents results from linear regressions of the outcomes in each column on the treatment dummy. Column (1) uses the pre-registered definition of profit which devalued all labor costs from market rate by 50% following local unemployment rate estimates. Column (2) uses our preferred definition, which only devalues unpaid labor at 50% of the market rate. Column (3) values all labor at the market rate. Standard errors (in parentheses) are clustered at the treatment assignment (farmer group cluster) level. All regressions include field office (site) fixed effects and the full set of pre-specified controls.

Note in Panel B of Tables 2 and 3, we consider input use valued in USD. Tables D.3 and D.4 show that we also detect a sizeable increase in fertilizer use when measured in kilograms. We can also break down fertilizer use by phase, and here we see the underlying substitution behind the effect on fertilizer timing noted in Panel A of Table 2. Farmers in the treatment group are not only using more fertilizer at the "correct" time, but also using less fertilizer at incorrect times.

Table D.3: Quantity of fertilizer used (kgs) by planting phase, primary sample

	At Planting		Post Planting	
	(1) DAP	(2) CAN	(3) DAP	(4) CAN
1AF participant	21.653***	-0.110	-6.853***	19.801***
	(1.780)	(0.133)	(1.689)	(1.809)
Observations	757	757	757	757
R^2	0.232	0.011	0.202	0.192
Control Mean Dep. Var	8.938	0.212	10.053	10.782

Results in this table are from linear regressions of the outcomes in each column on the treatment dummy. Standard errors (in parentheses) clustered at the treatment assignment (farmer group cluster) level. All regressions include field office (site) fixed effects.

Table D.4: Quantity of fertilizer used (kgs) by planting phase, full sample

	At Planting		Post Planting	
	(1)	(2)	(3)	(4)
	DAP	CAN	DAP	CAN
1AF participant=1	21.694***	-0.094	-6.943***	20.074***
	(1.855)	(0.165)	(1.752)	(1.871)
Past 1AF participant=1	7.653***	0.174	-2.613*	6.277***
	(1.621)	(0.235)	(1.544)	(1.461)
1AF participant=1 × Past 1AF participant=1	-3.255	0.237	2.794	-0.333
	(2.434)	(0.300)	(1.710)	(2.494)
Observations	2122	2122	2122	2122
R^2	0.171	0.008	0.160	0.169
Control Mean Dep. Var	8.938	0.212	10.053	10.782

Results in this table are from linear regressions of the outcomes in each column on the treatment dummy. Standard errors (in parentheses) clustered at the treatment assignment (farmer group cluster) level. All regressions include field office (site) fixed effects.

E   Sample Attrition

In this section, we aim to provide a careful accounting of sample attrition across surveys and stages of the intervention. The first level of potential attrition was between the baseline survey and pre-enrollment qualification. To qualify for the study, farmers had to pay a 500 KES deposit and form farmer groups. Farmers who did not complete the prepayment or farmers who failed to form a group of at least 3 members were dropped. This meant that of 3137 farmers originally surveyed at baseline, 662 did not end up enrolling in 1AF for the study season. Note that treatment was assigned after farmers were dropped at this stage, so we should not expect any threats to internal validity from post-baseline attrition. However, investigating this should allow us to better understand how the results may generalize outside the scope of this study.

Table E.1 shows how the enrolled and dropped samples differ, and show that almost across the board, the farmers who managed to enroll were significantly different than dropped farmers. Enrolled farmers were more likely to be pre-exposed, had better knowledge of 1AF practices, were more likely to use improved seeds and fertilizer, and were more likely to intercrop. Enrolled farmers also seem more specialized in farming, being more likely to receive more than half their income from farm labor, farming more acres for maize in 2016 and 2015, and harvesting more maize in both previous seasons. Finally, enrolled farmers seem more wealthy, reporting higher acreage for owned land, and a higher assets score.

For the most part, these variables simply suggest that 1AF may struggle to reach the poorest farmers, a challenge that is not unusual for entrepreneurially-focused agriculture programs (Carter et al., 2019). The enrolled farmers cultivate fewer acres than the average farmer in more representative samples like that from Tegemeo Institute, suggesting that although 1AF reaches many small farmers effectively, there may be a role for other approaches to help the very poorest. However, keeping in mind the difference in pre-exposure, it could also be that since past participation may have itself changed many of these variables, it could also simply be that this is driving most of the differences here.

To check this, Table E.2 repeats the balance test for the primary sample. We see that enrolled farmers are still significantly different on many variables, but mostly related to wealth and agricultural specialization. In terms of knowledge of 1AF practices, use of improved seed and fertilizer, credit access, and intercropping, the two groups are statistically indistinguishable. We see also that many differences, while significant, are much smaller in magnitude.

E.1   Harvest Survey Attrition

Once we restrict to the sample assigned to treatment, our primary driver of attrition from the final analysis dataset is missing harvest data. This missingness takes two forms. The first is total attrition: farmers for whom we have no dry weight survey data. This applies

Table E.1: Baseline balance across enrolled and dropped groups

		(1) Enrolled		(2) Dropped	Difference
Variable	N	Mean/SE	N	Mean/SE	(1)-(2)
Pre-exposed	2475	0.627 (0.010)	662	0.521 (0.019)	0.106***
Married	2427	0.876 (0.007)	662	0.881 (0.013)	-0.005
Father has Secondary Education	2429	0.411 (0.010)	634	0.339 (0.019)	0.072***
Receives more than half of income from farm	2427	0.765 (0.009)	662	0.719 (0.017)	0.046**
Used CAN or DAP and improved seeds on maize or beans	2456	0.770 (0.008)	659	0.677 (0.018)	0.094***
Reports Knowledge of OAF practices	2427	0.487 (0.010)	662	0.335 (0.018)	0.152***
Intercrops Maize and Beans	2427	0.473 (0.010)	662	0.432 (0.019)	0.041*
Farmer has access to credit	2427	0.719 (0.009)	662	0.702 (0.018)	0.017
Number of Household Members	2426	6.715 (0.052)	662	6.088 (0.097)	0.628***
Maize acres 2016	2427	1.006 (0.016)	662	0.839 (0.030)	0.167***
Maize acres 2015	2211	1.050 (0.059)	545	0.887 (0.028)	0.163
Acres owned for LR 2017	2387	2.698 (0.058)	644	2.149 (0.100)	0.549***
Acres owned and planted in LR 2016	2420	1.164 (0.022)	659	0.905 (0.032)	0.260***
Maize harvest 2016 (kgs)	2365	551.701 (12.160)	605	430.522 (20.624)	121.179***
Maize harvest 2015 (kgs)	2406	519.241 (12.225)	649	388.405 (23.114)	130.835***
Assets score	2475	18.749 (0.154)	662	16.184 (0.262)	2.565***
F-statistic (test of joint significance)					6.545***
Number of observations					2572

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

Table E.2: Baseline balance across enrolled and dropped groups among primary sample

		(1) Enrolled		(2) Dropped	Difference
Variable	N	Mean/SE	N	Mean/SE	(1)-(2)
Married	874	0.886 (0.011)	317	0.874 (0.019)	0.012
Father has Secondary Education	896	0.363 (0.016)	303	0.284 (0.026)	0.079**
Receives more than half of income from farm	874	0.757 (0.015)	317	0.675 (0.026)	0.082***
Used CAN or DAP and improved seeds on maize or beans	917	0.593 (0.016)	315	0.575 (0.028)	0.019
Reports Knowledge of OAF practices	874	0.097 (0.010)	317	0.104 (0.017)	-0.007
Intercrops Maize and Beans	874	0.509 (0.017)	317	0.467 (0.028)	0.042
Farmer has access to credit	874	0.706 (0.015)	317	0.681 (0.026)	0.025
Number of Household Members	874	6.597 (0.087)	317	5.839 (0.130)	0.758***
Maize acres 2016	874	0.955 (0.027)	317	0.759 (0.036)	0.195***
Maize acres 2015	763	1.143 (0.165)	255	0.834 (0.039)	0.309
Acres owned for LR 2017	859	2.614 (0.093)	306	2.060 (0.178)	0.554***
Acres owned and planted in LR 2016	872	1.128 (0.037)	314	0.842 (0.044)	0.286***
Maize harvest 2016 (kgs)	857	433.838 (19.396)	286	349.549 (23.385)	84.289**
Maize harvest 2015 (kgs)	864	377.908 (17.452)	310	295.306 (22.909)	82.602**
Assets score	922	16.319 (0.264)	317	14.795 (0.351)	1.524***
F-statistic (test of joint significance)					3.088***
Number of observations					940

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

to 102 (137) farmers in the primary (pre-exposed) sample, for a total attrition rate of 11% (9%). Total attrition is uncorrelated with treatment status, but is negatively correlated with pre-exposure.

The second form of missingness is partial attrition, which by definition is restricted to treated farmers: for some farmers, despite knowing that they cultivated more than zero maize non-enrolled acres, we are missing dry weight survey data (and, sometimes, input survey data). The primary driver of this missingness is due to farmer survey responses: some farmers initially told enumerators they had no non-enrolled maize acres, but for a later (post-harvest) GPS land measurement survey did indeed have non-enrolled acres. After excluding total attrition farmers, partial attrition affects 58 (95) farmers in the primary (pre-exposed) groups, for a partial attrition rate of 7% (7%).

We consider a number of strategies for checking the robustness of our results to partial attrition. The results of these strategies are presented in Figure E.1 and Figure E.2. Our first strategy is to make the simple but extreme assumption that missing yield data actually represents extremely poor yields or an entirely failed harvest and the respondent or the enumerator preferred not to report them. For this exercise, we set the missing yields to zero. We show separately the effects of making this assumption only for full attrition, and then additionally making this assumption for partial attrition. This estimate sets a rather extreme lower bound for the treatment effect given the attrition. For full attrition, the treatment effects we estimate are robust to even this extreme assumption. Given the greater numbers involved, this strategy substantially lowers the magnitude of our estimated treatment effects for partial attrition, especially for profit. This suggests that in this worst-case scenario, the average treatment effect of the program on profit may not be distinguishable from zero.

In addition, we consider two strategies we consider reasonable if we instead assume partial attrition is primarily or entirely (conditionally) random, rather than driven by negative results. In the first, we use a simple imputation strategy and fill missing non-enrolled yields with the field-office mean yields among control farmers. In the second, we use multiple imputation methods in Stata to impute the missing non-enrolled yields, using the multivariate normal method. Imputation is done exclusively using baseline farming practices and yields, although this result is also robust to using study-season covariates.
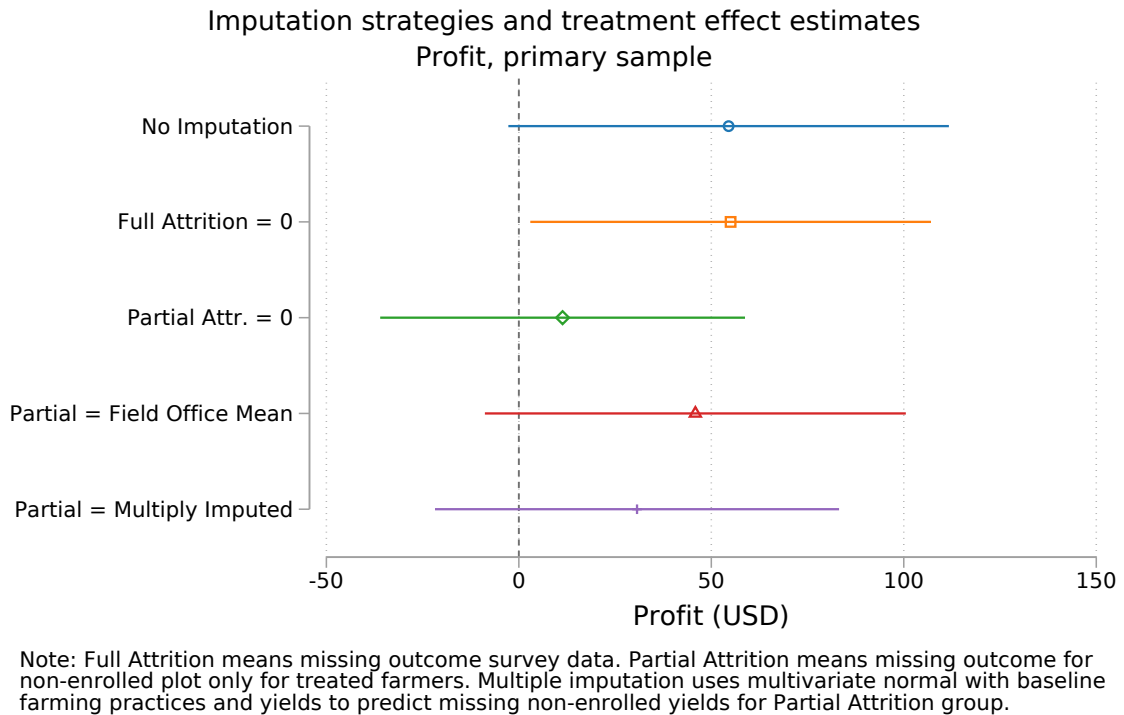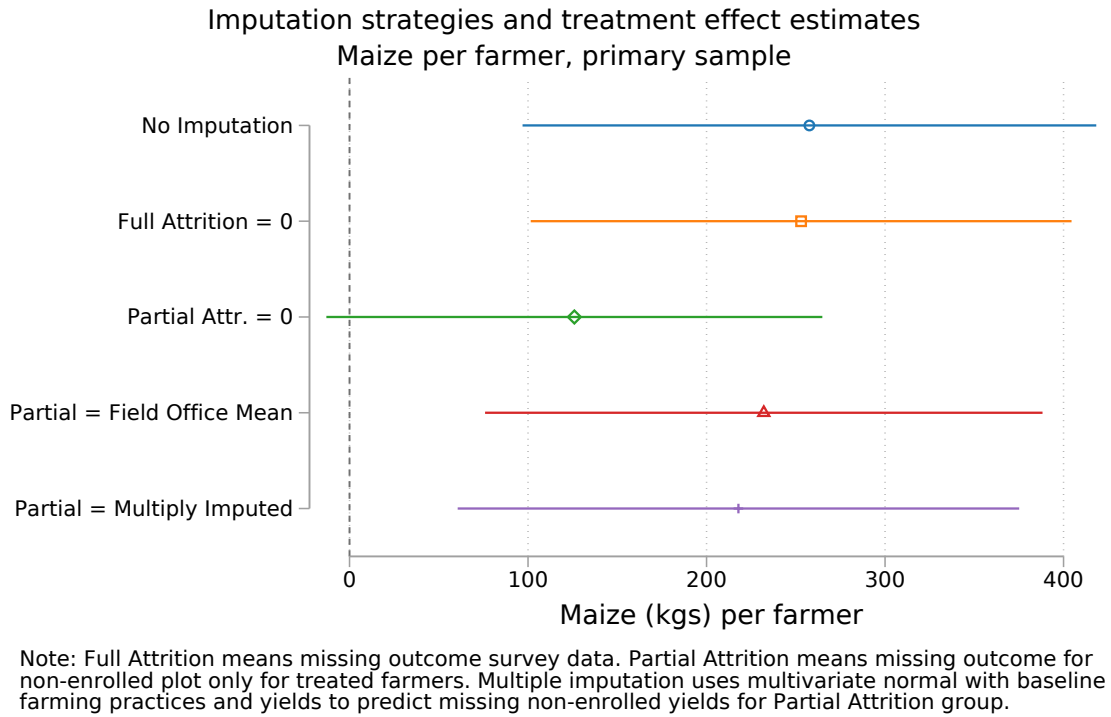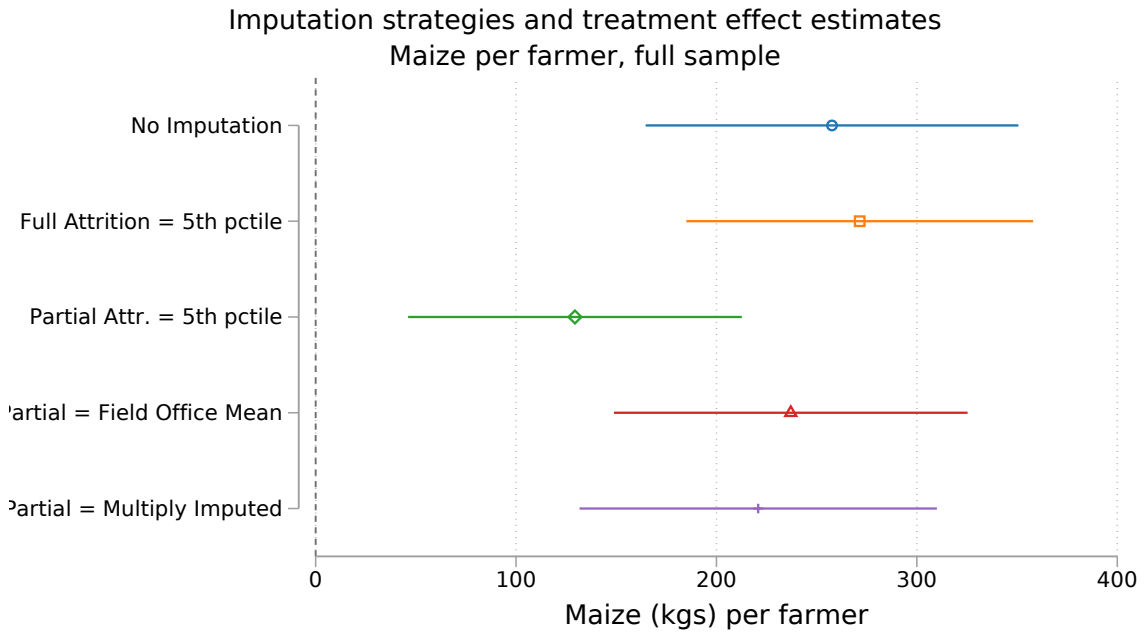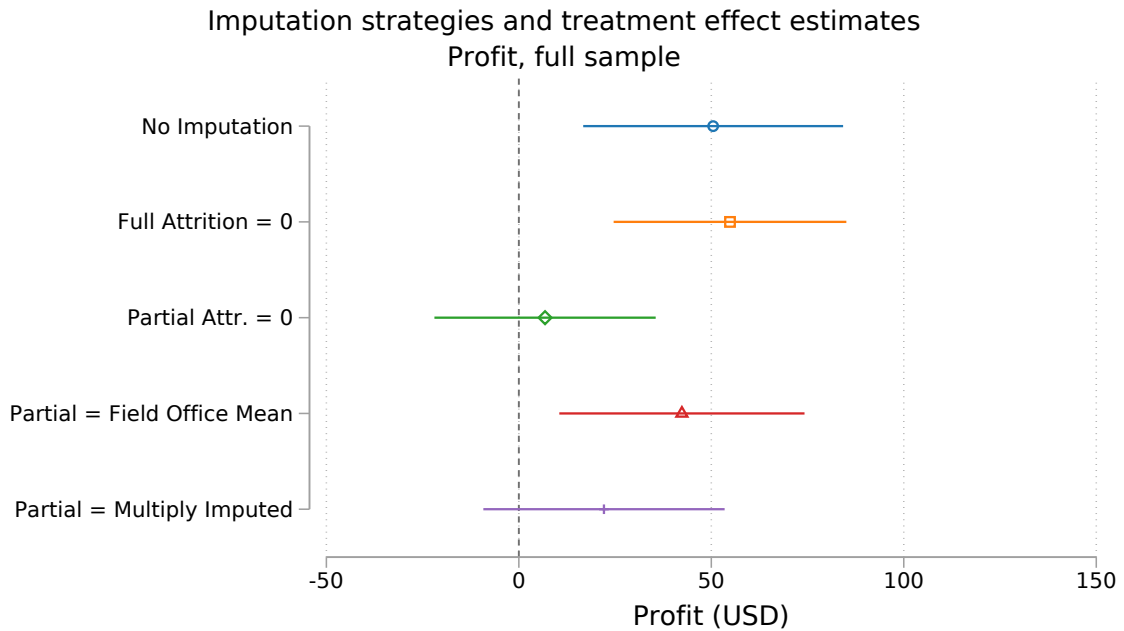
Imputation strategies and treatment effect estimates
Maize per farmer, primary sample

Note: Full Attrition means missing outcome survey data. Partial Attrition means missing outcome for non-enrolled plot only for treated farmers. Multiple imputation uses multivariate normal with baseline farming practices and yields to predict missing non-enrolled yields for Partial Attrition group.

Imputation strategies and treatment effect estimates
Profit, primary sample

Note: Full Attrition means missing outcome survey data. Partial Attrition means missing outcome for non-enrolled plot only for treated farmers. Multiple imputation uses multivariate normal with baseline farming practices and yields to predict missing non-enrolled yields for Partial Attrition group.

Figure E.1: Imputation strategies for attrition, primary sample

Imputation strategies and treatment effect estimates
Maize per farmer, full sample

Note: Full Attrition means missing outcome survey data. Partial Attrition means missing outcome for non-enrolled plot only for treated farmers. Multiple imputation uses multivariate normal with baseline farming practices and yields to predict missing non-enrolled yields for Partial Attrition group.

Imputation strategies and treatment effect estimates
Profit, full sample

Note: Full Attrition means missing outcome survey data. Partial Attrition means missing outcome for non-enrolled plot only for treated farmers. Multiple imputation uses multivariate normal with baseline farming practices and yields to predict missing non-enrolled yields for Partial Attrition group.

Figure E.2: Imputation strategies for attrition, full sample