Relaxing Credit and Information Constraints: Five-Year Experimental Evidence from Tanzanian Agriculture*

Abdulrazzak Tamim Aurélie P. Harou Marshall Burke
David Lobell Malgosia Madajewicz Christopher Magomba
Hope Michelson Cheryl A. Palm Jiani Xue

October 3, 2022

*Tamim: Department of Economics, University of California, Berkeley, 530 Evans Hall, Berkeley, CA 94720 (e-mail: atamim@berkeley.edu); Harou: Department of Natural Resource Science, McGill University, 21,111 Lakeshore Drive, Macdonald-Stewart 3-035, Sainte-Anne-de-Bellevue, QC H9X 3V9 (e-mail: aurelie.harou@mcgill.ca); Burke: Department of Earth System Science, and Center on Food Security and the Environment, Stanford University, 473 Via Ortega, Stanford, CA 94305 (e-mail: mburke@stanford.edu); Lobell: Department of Earth System Science, and Center on Food Security and the Environment, Stanford University, 473 Via Ortega, Stanford, CA 94305 (e-mail: dlobell@stanford.edu); Madajewicz: Center for Climate Systems Research. Columbia University, 2880 Broadway, Rm 500, New York, NY 10025 (e-mail: mm1174@columbia.edu); Magomba: Department of Food and Resource Economics, Sokoine University of Agriculture, Box 3007, SUA-Morogoro (e-mail cmagomba@sua.ac.tz); Michelson: Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, 314 Mumford Hall 1301 W. Gregory, Urbana 61801 (e-mail: hopecm@illinois.edu); Palm: Food Systems Institute and Agricultural and Biological Engineering, University of Florida, Gainesville, FL 32611-0570 P.O. Box 110570 (e-mail: cpalm@ufl.edu); Xue: Wharton School, University of Pennsylvania, 700 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104 (e-mail: jennyxue@wharton.upenn.edu). This research is funded by the SSHRC grant 430-2018-1121 and the USAID Bureau for Food Security. We thank Matthieu Chemin, Xavier Gine, and Sean Higgins for providing helpful suggestions. This paper also benefited from discussions at the 2021 Pacific Conference for Development Economics, the 2020 Virtual North East Universities Development Consortium, the 2020 Agricultural & Applied Economics Association Virtual Meeting, and the Canadian Agricultural Economics Society Webinar Series. The data used in this paper were collected by Aika Aku, Venacky Daniel, Ambonisye Haule, Litwagile Kajela, Vulfrida Kessy, Sweetness Kileo, Steven Kuhanwa, Habibu Lupatu, Josephat Mbepera, Lilian Mduda, Saidi Mhina, Angel Minja, Paschal Mmelo, Anna Muffui, Burton Mwambokela, Rajabu Mwanyika, and Gloria Peter from Sokoine University of Agriculture under the supervision of Abdulrazzak Tamim, Christopher Magomba, and Revocatus Kayaga Ntengo. The original study was registered in the AEA RCT Registry and the unique identifying number is: "AEARCTR-0005470". This study was approved by the Research Ethics Board at McGill University, file number 16-0619.

Abstract

Low fertilizer application by small farmers continues to inhibit crop yields around the world. The reasons behind low application rates continue to be debated. We study the longer-term outcomes of a field experiment which focused on increasing fertilizer use. The original experiment showed that plot-specific fertilizer recommendations combined with a subsidy increase amounts of applied fertilizer and maize yields relative to either intervention alone. We show that these effects dissipate once the subsidy is discontinued. Our results indicate that ability to pay for fertilizer continues to limit fertilizer use even when farmers have information about appropriate fertilizer types and amounts, and even after farmers have learned that fertilizer use is profitable. (*JEL* O13, Q16, Q18)

I Introduction

Although many regions of the world have achieved dramatic increases in agricultural productivity since the Green Revolution, staple cereal yields remain low throughout sub-Saharan Africa (SSA). One important contributor to this low productivity is limited use of mineral fertilizer (Duflo et al., 2008; Xu et al., 2009; Harou et al., 2017, 2022). Researchers have studied a range of explanations for low investment in mineral fertilizer, including unresponsive soils, high transaction costs, farmers' time preferences, non-adoption among peers, uncertainty, and absence of information about the suitability of fertilizer types and specific quantities for a farmer's soils (Marenya and Barrett, 2009; Duflo et al., 2011; BenYishay and Mobarak, 2018; Harou et al., 2022; Emerick and Dar, 2021).[1]

This study extends the work of Harou et al. (2022), which evaluated effects of two interventions to increase fertilizer use and maize yields in Tanzania: plot-specific fertilizer recommendations and vouchers to purchase agricultural inputs.

[1]For a summary, see Suri and Udry (2022).

Harou et al. (2022) evaluated the impact of information provision and vouchers individually and in combination relative to a control group. Concluded in 2016, the study demonstrated three results. First, only farmers who received vouchers increased amounts of fertilizer they purchased and applied. Second, the increase in fertilizer use among farmers who received both vouchers and plot-specific recommendations was more than twice the increase among farmers who received only vouchers. Third, only the combination of vouchers and plot-specific recommendations increased maize yields, increasing per acre profits by an average amount equivalent to 7-9 days of wage work.

While the majority of studies related to evaluating policies and programs to increase agricultural input use in SSA have focused on time spans of one to three years[2], in this paper, we estimate the effects of the Harou et al. (2022) experiment five years on: three years after the vouchers were discontinued and after plot-level fertilizer recommendations were extended to all farmers – including the group who received only vouchers and the control group. In August 2019, we successfully resurveyed almost 90% of the original 1,050 households from Harou et al. (2022) to determine how fertilizer applications and yields had evolved. We also collected polygons of farmers' plots using GPS devices to compare farmers' self-reported yields and plot areas with satellite-derived yield estimates and GPS-based plot sizes.

We find that the treatment effects from 2016 have dissipated.[3] In 2019, fertilizer applications and maize yields are similar across all treatment groups as well as the control group. This lack of long-term effect on fertilizer use and yields is driven

[2]Duflo et al. (2011) collect data on three seasons in their behavioral bias experiment; Beaman et al. (2021) follow farmers for three years to study adoption of pit planting and crop residue management; Carter et al. (2021) collect data for two years after the implementation of an input subsidy program; and Karlan et al. (2014) follow farmers receiving cash grants or insurance or both for three years.

[3]Our findings are similar to Fishman et al. (2021) who do not find impacts on mineral fertilizer adoption three and six seasons after phaseout of a program that included extension and a subsidy. However, they do find positive impacts on long term adoption of improved seeds.

by the fact that farmers who received vouchers and plot-level recommendations have reverted to baseline levels of fertilizer use. Control farmers did not increase fertilizer use after receiving plot-level recommendations in 2016.

The results suggest that farmers cannot afford fertilizer in the absence of an intervention that supports their ability to pay for fertilizer. The Harou et al. (2022) experiment indicated that farmers need both plot-level fertilizer recommendations and a subsidy to increase application of fertilizer. However, another potential reason for low investment in the recommendation-only and control groups may have been that farmers did not know if increasing fertilizer would be profitable. In the original experiment, farmers in the voucher and recommendation group learned from their experience that the recommended fertilizer is profitable. Therefore, the results in this study strongly suggest that ability to pay at the beginning of the planting season continues to be a constraint even once farmers know how much and which fertilizers they should apply, and they know that fertilizer is profitable.

The question remains how much financial support farmers need to continue investing in fertilizer. Interventions may need to address insufficient liquidity, uninsured risk, and/or preference profiles that prevent farmers from investing early in the season, when they have cash on hand (Duflo et al., 2011). The needed support is likely to depend on realizations of climate conditions. A significant drought affected outcomes during the follow-up season in the Harou et al. (2022) experiment, in 2016. Farmers who received both vouchers and recommendations experienced significantly smaller yield declines relative to baseline than did farmers in the remaining study groups. However, the challenge of coping with the drought may have prevented even the farmers in the voucher and recommendation group from accumulating the cash reserves necessary to purchase fertilizer in subsequent seasons.

We also compare results obtained with self-reported yields and satellite-based measures of maize yields. Self-reported plot area and crop production may be subject to misreporting and non-classical measurement error (Abay et al., 2019), contributing to researchers' interest in GPS-based area measures and satellite-based production assessments (Lobell et al., 2020). We show that the satellite-based yield measures confirm that treatment and control groups experienced similar yields in 2019. However, satellite-based yield measures also suggest that no treatment group experienced higher maize yields in 2016, despite substantially increased use of fertilizer in the voucher plus recommendations group. We hold that these results do not warrant revision of Harou et al. (2022) for at least two reasons.

First, satellite-based yield measures in our data use plot area measured by GPS in 2019. In order to obtain plot areas for 2016, to re-estimate the treatment effects in the original experiment, we asked farmers to recall the boundaries of their main maize plots in 2016. We show that the GPS-based measure matches self-reported plot area well in 2019 but it does not match well plot areas that farmers self-reported in 2016. The reason may be, not surprisingly, that farmers' recall in 2019 of the 2016 plot boundaries is worse than their recall of the 2019 boundaries. The process of obtaining GPS measures for prior years may include measurement error, biasing yields downward.

Second, research has not yet addressed concerns about accuracy of satellite measures over small spatial scales. For example, satellite-based yield measures may be more accurate on pure-stand farm plots than on intercropped plots (Lobell et al., 2020). In our data, pure-stand plots exhibit heterogeneous yield effects based on satellite data that are consistent with the self-reported yields. We find a significant impact of the voucher and recommendation treatment pure-stand plots in 2016 that are above median size with an effect size of 0.25 standard deviation units. Impacts in 2016 are concentrated among and driven by the most productive farmers. Overall, these results are consistent with Suri (2011) who finds important heterogeneity in returns to adoption.

The remainder of this paper is organized as follows: section II details our experimental design and data sources. Section III presents the estimation strategies and section IV reports the main results. We present heterogeneity in section V and discuss the results in section VI. Section VII concludes.

II Research Design

II.1 The 2014 Randomized Experiment

We begin with an overview of the design and results of the original 2014 RCT.[4] Farmers were randomly invited to participate in the original study in two stages. First, out of all maize-growing and accessible villages in Morogoro Rural District, Morogoro Region, Tanzania, 27 villages were assigned randomly as control and 20 as treatment. Second, within the 20 treatment villages, farmers were randomly allocated to one of four groups (recommendations, vouchers, both treatments, and control). Researchers designed the study this way to permit testing of spillovers by comparing control farmers in control villages and control farmers in treatment villages. As in Harou et al. (2022), we focus in this paper on the treatment effects on the main maize plot (MMP), the plot identified by the farmer as most important for household food security and income.

Figure 1 shows the number of participants per treatment (source: Harou et al. (2022)). Ten eligible farmers were randomly selected to participate in the study in each control village. In treatment villages, the farmers were assigned randomly to one of the following arms (10 farmers per arm):

1. Plot-specific recommendations (Group R): Farmers received information about which fertilizer types and quantities they should apply on their 2014 MMP.

[4] Full details for the 2014 study can be found in Harou et al. (2022).

Farmers were provided with recommendations both for one acre of maize culitvation and for one half an acre of maize. Each farmer's recommendations were based on analysis of soil samples collected from the farmer's plot and tested by a team of agronomists and soil scientists from Sokoine University of Agriculture (SUA). Agronomists then met with treated farmers and presented them with a card explaining their soils' deficiencies and what fertilizers were recommended for their MMP.

2. Vouchers (Group V): Farmers in this group were given a voucher valued at 80,000 TZ Shillings (about 40 USD at the time of the study) that they could redeem to purchase any fertilizer they wanted from a specific agro-input dealer. The voucher was sufficient to cover most fertilizer recommendations for a 0.5-acre plot.

3. Plot-specific recommendations and vouchers (Group V+R): This group received both a voucher and recommendations as described above (treatments V and R).

4. Control farmers (Group C): Control farmers received neither recommendations nor vouchers in 2016. This group consisted of control farmers in treatment villages only since we exclude control villages, as detailed in section II.4.

Farmers in the voucher and control groups received plot-specific fertilizer recommendations at the end of the experiment and the collection of the 2016 first endline data. Thus, all farmers had received plot-specific fertilizer recommendations by the time we collected data in 2019.

II.2 Data Collection

In 2019, we resurveyed the group of 1,050 households who participated in the original 2014-2016 RCT. The questionnaires included modules on maize yields, mineral fertilizer use, questions about farmers' retention of the fertilizer recommendations, characteristics of each year's MMP, assets, credit, and land tenure. To remind farmers of the 2014 MMP from which soil samples were taken, we showed them a map of their plots and their house location drawn by enumerators in 2014. We also reminded the farmers how they had referred to that plot in 2014. Because the original project was highly respected by extension agents and farmers for its impact and from SUA's reputation, we had good success reaching and interviewing the same farmers from the original study. We were able to visit 920 households out of the 1,050 who participated in the original study by Harou et al. (2022), resulting in an overall high tracking rate of around 88% since baseline in 2014 – or 93.5% compared to the 984 farmers surveyed in 2016 – despite a lack of communication with respondents since the initial study was concluded in 2016.

In 2019, we also collected GPS polygons of farmers' 2014 MMP, 2016 MMP and 2019 MMP. Farmers were asked during the survey in 2019 if they cultivated the same plot in 2014 (2016) as their 2019 MMP. If not, the enumerator asked to visit the 2014 (2016) MMP to collect the new GPS coordinates. In addition to collecting GPS coordinates, the enumerators administered a short questionnaire, including details about plot intercropping, whether the plot included any large objects such as buildings or trees inside its perimeter, and the clarity of the plot's boundaries.

II.3 Outcome Variables

Our main outcomes of interest in this study are fertilizer adoption and maize yields. We measure fertilizer adoption at the extensive margin using a binary variable that takes the value one if the farmer applied any mineral fertilizer and zero otherwise. At the intensive margin, we measure adoption as the quantity of fertilizer applied per acre of maize planted. The acreage we use is either self-reported in the socioeconomic survey or measured using handheld GPS devices, as specified further below.

We measure yields in three different ways. First, we simply use self-reported output in kilograms divided by self-reported planted area in acres. Second, we use analysis of satellite imagery to estimate productivity. Each plot is mapped using a GPS device. Then, for each plot polygon, we extract for each available image during the growing season the average pixel values within the polygon for the visible and near-infrared (NIR) bands of Sentinel-2 satellite measurements, which have 10m spatial resolution. From these we compute for each date the Green Chlorophyll Vegetation Index (GCVI), which is defined as:

$$GCVI = (NIR/Green) - 1 \quad (1)$$

From the time series of GCVI values for each field, we fit a recursive harmonic regression to make the data more robust to missing observations during the peak of the growing season. Then, using the 10th iteration of the harmonic curve we extract the maximum GCVI value over the growing season, which has been shown to correlate well with crop yields in the region (Jin et al., 2019). Our third productivity measure uses self-reported output divided by GPS-measured planted area, which we refer to as GPS-corrected yields.

II.4 Balance

Table 1 displays sample size and retention rates by treatment group and year. To test for baseline differences between the treatment and control groups on outcome variables of interest and on control variables, we regress baseline outcome and control variables on treatment indicators using ordinary least squares (OLS) with village fixed effects (FE) according to the following specification:

$$b_{iv} = \alpha_0 + \sum_{k=1}^{3} \theta_k TREAT_i^k + d_v + \varepsilon_{iv} \tag{2}$$

where b_{iv} is a baseline variable for farmer i in village v, α_0 is a constant, $TREAT_i^k$ is a binary variable that takes the value one for each farmer i assigned to one of the k treatment arms (V, R and V+R) and zero otherwise, d_v is village FE to control for the initial village sampling, and ε_{iv} is the associated idiosyncratic error term. The omitted category is the control group (either pooled from control and treatment villages or from treatment villages only), and standard errors are clustered at the village-level.

Table 2 presents balance tests for the sample excluding control villages, which is our preferred specification, and restricting to the sample for which we have both satellite and self-reported yields (as detailed in section III). Most covariates are fairly balanced except for the livestock ownership indicator, which is imbalanced between all treatments and the control group. Out of all 126 comparisons, 11 are significant at a level less than 10%, representing 8.7% of all comparisons, which is close to what would be expected by chance.

We exclude control villages from all regressions because there are significant differences between means of outcome variables in control villages compared to other groups, see Table A1.This point is discussed in Harou et al. (2022). We conduct robustness checks that include control villages.

II.5 Attrition

From the 1,050 farmers who participated in the study at baseline in 2014, we were able to revisit 920 farmers in 2019, resulting in an attrition rate of 12.4% (or 13% excluding control villages), which is relatively low given that five (three) years had passed since the baseline (endline) data collection in Harou et al. (2022). The main reasons for not being able to locate farmers were migration (38.8%) and death (15.5%) (retention rates by treatment and control groups are displayed in Table 1).

To test whether attrition may have differentially impacted the treatment groups thereby introducing bias, we first regress treatment on whether households attrited, shown in Table A2 in the Appendix. Attrition is unlikely to bias our estimates since the probability of attrition does not differ significantly between all treatments and the control group. Second, we look at retention, defined as having a retained, non-missing observation in the sample, regardless of the reason why the observation is missing. Given that fewer farmers cultivated their 2014 MMP in 2016 and 2019, and because GPS polygons were only collected in 2019, it is possible to have non-classical attrition if polygons were more likely to be collected on cultivated plots. We find that the treatments do not predict retention in the panel when we define it as having a missing value, regardless of the reason, as shown in Table A3. Only one point estimate – for the voucher group in 2016 – is significant but at the 10% level. Third, we follow Ghanem et al. (2021) to assess internal validity following two different assumptions: random assignment conditional on attrition, and independence between unobservables that affect attrition and our outcomes. We still find no significant difference between any treatment and the control group (Table A4). Because attrition is balanced, we do not adjust for it.

III Estimation

In Table 3, we present descriptives regarding which plot farmers cultivated with maize and how cultivation of the soil-tested plot changed over time. The number of farmers who cultivated their 2014 MMP (from which soil samples were taken in 2014) over 2016-2019 decreased from 822 in 2016 to 416 in 2019. Interestingly, more farmers switched to cultivate new main maize plots, with an increase from 81 in 2016 to 369 in 2019 (Table 3). Since these new cultivated plots were not a part of the original experiment, we do not study the 2016 MMP and 2019 MMP due to potential selection bias and instead focus on the tested 2014 MMP.

We are interested in measuring the effects of the voucher, recommendations, and voucher plus recommendations treatments on fertilizer adoption and maize productivity over 2016-2019. We estimate the average treatment effects using analysis of covariance (ANCOVA) with village fixed effects (FE):

$$Y_{iv} = \alpha_1 + \varphi Y_{i2014} + \sum_{k=1}^{3} \gamma_k TREAT_i^k + d_v + e_{iv} \tag{3}$$

where Y_{iv} is either fertilizer use for farmer i in village v measured as a binary variable or in kg/acre, or yields measured in kg/acre or as a z-score, α_1 is a constant, Y_{i2014} is the outcome variable measured at baseline, $TREAT_i^k$ are dummy variables for the three treatment arms, V, R, and V+R, the d_v's are village FE to control for the initial village sampling, and e_{iv} is the associated idiosyncratic error term that varies across individuals and between villages. The main coefficients of interest are γ_k, which are the ANCOVA estimator for treatment k (V, R or V+R). We estimate equation (3) in 2016 only and in 2019 only. We cluster standard errors at the village level to account for potential within-village correlation.

We also estimate the treatment impact using satellite images of farmers' MMPs following the procedure outlined in Lobell et al. (2020). Because reliable satellite images are only available starting in 2016, we are unable to predict yields via satellite at baseline. We therefore have to rely on OLS to estimate the treatment impact on satellite-derived productivity measures instead of using ANCOVA. We estimate the following OLS specification using post-intervention data only:

$$Y_{iv} = \alpha_2 + \sum_{k=1}^{3} \beta_k TREAT_i^k + d_v + u_{iv} \tag{4}$$

where Y_{iv} is yields of farmer i in village v in 2016 or 2019, transformed by subtracting the control group's mean from each observation and then dividing the result by the control group's standard deviation so that it is possible to compare self-reported and satellite yields using z-scores, α_2 is a constant, and u_{iv} is the idiosyncratic error term. The remaining variables are the same as defined in equation (3). We restrict the sample to plots that have both satellite-based and self-reported yields in order to compare results. Furthermore, because satellite-based yield measures in intercropped fields are a combined measure of productivity on all crops, and thus are less directly comparable to self-reported yields of a single crop (maize), we include an indicator variable that equals one if a plot is pure-stand and zero if it is intercropped. We also include an indicator variable for the existence of buildings or large trees or a banana or coffee plantation inside the perimeter of the plot, as these might affect satellite-derived yield estimates. From the 782 original surveys, we are left with 491 and 294 observations for the 2014 MMP in 2016 and 2019, respectively, for which we have both satellite and self-reported yields.

IV Results

IV.1 Technology Adoption

We find that fertilizer application does not differ between any of the groups in 2019, as shown in Table 4. Therefore, the combined information and voucher intervention conducted in 2014 does not confer long-term benefits on farmers in the V+R group relative to farmers in the other groups. Such benefits could occur for example if the one-time voucher and the increased yields allowed farmers to earn extra profits, which could then be invested in subsequent seasons in continuing to increase fertilizer application and thereby productivity and profits. We do confirm that fertilizer application increased in the V+R and V groups in 2016, as shown previously in Table 4.

We also find that the use of fertilizer has declined back to 2014 baseline levels, as shown in Table 4. Thus fertilizer use among R and control groups has not increased in response to plot-specific fertilizer recommendations provided in 2016. Rather the fertilizer use among V+R and V groups has declined to baseline levels. Therefore, the farmers have not sustained the 2016 gains.

IV.2 Agricultural Productivity

IV.2.1 Medium-term effects of information on yields using self-reported data

We find that farmers in 2019 do not maintain the yield gains they experienced from the intervention in 2016, shown in Table 5, column 3. This is not surprising given fertilizer levels in 2019 returned to baseline levels (Table 4). This result holds when controlling for pure-stand plots and plantations (Table 5, column 7), for reasons explained earlier in section III. We corroborate the findings of Harou et al. (2022) that yields in 2016 increased for the V+R treatment (Table 5, columns 1 and 5), even when restricting our sample to plots that have both satellite-based and self-reported yields, as explained in section III.

IV.2.2 Comparing self-reported, GPS-corrected, and satellited-derived yields

On the other hand, satellite-derived yield estimates indicate no significant treatment effect in 2019 or 2016 (Table 5, columns 2 and 4). However, we note that when controls are included (columns 6 and 8), the 2016 treatment impact on self-reported and satellite yields have 95 percent confidence intervals of (0.11, 0.53) and (-0.11, 0.23) standard deviation units, respectively, suggesting that the upper bound of the satellite estimates lies within the 95 percent confidence interval of the self-reported results.

Self-reported yields measured in kg/acre can suffer from two distinct sources of measurement error: misreporting total yields and/or misreporting plot size. Crop-cuts are typically seen as the gold standard of estimating yields, while GPS measures are considered the most objective measure of plot size (Abay et al., 2019; Carletto et al., 2013, 2015) [5] Due to budgetary constraints, we were not able to collect crop cuts in any of the years, but, as previously described, we did collect GPS coordinates of plots in 2019. Hence, we can examine whether misreporting of plot size plays a role in our results. Figure 2 shows the correlation between self-reported and GPS areas by treatment and year. The self-reported and GPS areas are closest to each other in 2019, the year in which the GPS polygons were collected, almost consistently across the different treatment groups, except for the V+R group. The self-reported and GPS measures diverge in earlier years. The team collected GPS coordinates of the main maize plots planted in previous years in 2019. The divergence suggests that the GPS coordinates may be affected by poor recall of what the

[5] While GPS measures are considered more precise and objective measures, they nonetheless contain their own shortcomings (see Cohen (2019)).

boundaries of the planted plot were in previous years.

When comparing self-reported productivity (kg/acre) measures using respondent-reported and GPS plot size (Table 6), we again find no sustained V+R effect in 2019 (column 4)[6]. Furthermore, in 2016, the self-reported yields per GPS plot size are no longer statistically significant (column 2), unlike the significant treatment impact in 2016 of around 100 kg/acre (25% of the control groups' average at baseline) (column 1). The reason for this may be that the GPS-based plot size is inaccurate for 2016 because of poor recall of plot boundaries.

IV.3 Robustness of Results

We test the robustness of our results in two ways. First, since all farmers were given the recommendations after 2016, having received the voucher in 2016 distinguishes the treatment groups in 2019. We therefore test the robustness of our results by pooling the recommendations and control groups together, and the voucher and the voucher plus recommendations groups together. We find that farmers in the pooled V and V+R group apply more fertilizer in 2016 (Table A5 in the Appendix), but no longer in 2019, corroborating our findings above. Likewise, the treatment impact using satellite estimates in both 2016 and 2019 remains insignificant when we pool the voucher and voucher plus recommendations groups (Table A6). The pooled V and V+R groups have a statistically significant effect on self-reported yields in 2019, however, suggesting that vouchers may have long-term, minor impacts on self-reported productivity. This is a surprising result given that the same group of households did not apply more fertilizer. Finally, using GPS-corrected productivity measures, our results stay insignificant in both 2016 and 2019 as in the unpooled analysis presented above (Table A7).

[6] We use the ANCOVA estimator employed in equation (3), but again restrict the sample to plots that have both GPS-based and self-reported areas

As a second robustness check, we include households in control villages and focus on cross-village estimates to examine if baseline imbalances may be driving some of the results. As Tables A8, A9 , and A10 show, our results are robust as all results are qualitatively similar.

V Heterogeneity

As seen in section IV.2, self-reports indicate a significant treatment impact on productivity in 2016 but no longer in 2019, while remote sensing and GPS-corrected measures show no effect in either year. In this section, we examine whether our findings are driven by certain sub-groups.[7]

V.1 Productivity

If modern agricultural technologies are risky and susceptible to price sensitivity, then returns to adoption would be divergent among different types of farmers (Karlan et al., 2014). Thus, it may be rational for some farmers in our study not to invest in this new technology, especially that only one percent of them applied fertilizer at baseline. We examine the treatment effect on yields in 2016 and 2019 by quantile using a recentered influence function following Firpo et al. (2009):

$$RIF(Y; q_\tau, F_Y) = q_\tau + IF(Y; q_\tau) = q_\tau + \frac{\tau - \Gamma\{Y \leq q_\tau\}}{f_Y(q_\tau)} \tag{5}$$

where $IF(Y; q_\tau)$ is the influence function of the quantile specified q_τ, $\Gamma\{Y \leq q_\tau\}$ is an indicator that equals unity if productivity is less than or equal to q_τ and zero otherwise, and $f_Y(q_\tau)$ is the density of the unconditional distribution of productivity

[7]Tables A11 and A12 in the Appendix show treatment impacts on organic inputs and labor. Households in the V+R group were not more likely to use complementary inputs in 2016 (organic fertilizer and manure) in Table A11 and labor in Table A12).

evaluated at q_τ. Then, we regress RIF on the treatments using OLS. The V+R results are displayed in figures 3 and 4 that replicate columns (1) and (2) in Tables 5 and 6, respectively, and display point estimates at various deciles along with their 95% confidence intervals.[8]

Figure 3 shows that for the V+R treatment, self-reported impacts are concentrated among the top 20% of the distribution, and a similar pattern is observed in Figure 4. However, quantile treatment impacts on satellite-based productivity (Figure 3, part b) and GPS-corrected productivity (Figure 4, part b) are insignificant. These findings suggest that the V+R treatment impact on self-reported yields is driven by a relatively small number of highly productive farmers who may be able to bear the risks associated with fertilizer adoption.[9]

V.2 Soil Fertility

Farmers may be aware of their soil deficiencies and decide rationally not to invest in fertilizer if they perceive the returns to be low. Hence, we examine treatment impacts by underlying soil quality by estimating the following equation:[10]

$$Y_{iv} = \beta_0 + \sum_{k=1}^{3} \theta_k TREAT_i^k + \gamma RichSoil_i + \sum_{k=1}^{3} \gamma_k TREAT_i^k \times RichSoil_i + d_v + \epsilon_{iv} \tag{6}$$

where $RichSoil_i$ an indicator variable that equals unity if a household has pre-treatment fertile soils with optimal salinity (slightly or very saline), and zero for

[8]Note that the sample size is the same as the ATE impacts – namely, $N = 491$ in Figure 3 and $N = 431$ in Figure 4.

[9]Using data from the Living Standards Measurement Surveys–Integrated Surveys of Agriculture (LSMS-ISA), Gollin and Udry (2021) find a 2,558 percent difference in maize yields between the 5th and 95th percentile in Tanzania, indicating large dispersion across plots. Our results indicate a similar pattern: the treatment impact is concentrated among the most productive farmers (Figure 3) and those who have good soils (Table 9).

[10]We use ANCOVA for adoption since, unlike for satellite yields, we have baseline data.

soils that have low, medium, or severe salinity, β_0 is a constant term, and ϵ_{iv} is the idiosyncratic error term. All other variables and parameters are the same as those in equation (3).

We find increased adoption among V+R farmers with good soils: farmers in the V+R group who cultivate poor soils applied 17-22 kg/acre more than the control group compared to those who have rich soils who applied 33-35 kg/acre more fertilizer than the control group, both statistically significant at the 1% level (Table 8).

Table 9 shows that higher self-reported yields are observed for households with rich soil quality with an estimate of 0.71 standard deviation units (column 1 in Panel B), which is higher than the 0.44 homogeneous impact in Table 5. Moreover, this impact seems to be persistent in the longer-term in 2019 as seen in column 4, with a point estimate of 0.43. On the other hand, satellite-based and GPS-corrected estimates remain insignificant. Our (self-reported) results suggest there are heterogeneous returns to fertilizer based on soil health, consistent with previous findings in the literature (Marenya and Barrett, 2009).

V.3 Pure-Stand Plots

Third, we assess the sensitivity of satellite yields to the exclusion of intercropped plots (Lobell et al., 2020) and small plots (defined as plots smaller than 1.5-2 acres, see below, using the GPS areas) that may suffer from higher measurement error (Cohen, 2019). Of the 491 households in the 2016 sample, 387 did not intercrop their maize plots. Out of those 387, 208 have GPS areas of more than 1.5 acres, 183 of more than 1.75 acres, and 156 of more than 2 acres.[11] We display these results in Figure 5 which shows 95 percent confidence intervals of treatment impacts on the

[11] Median GPS area among pure-stand plots is around 1.65 acres in 2016.

subsample of pure-stand plots, first for all GPS areas (unconditional), and then by restricting the analysis to larger plots.

When the sample is restricted to pure-stand plots with GPS acreage of more than 1.75 acres, the satellite measures indicate a significant treatment impact of 0.25 standard deviation units. Therefore, satellites may be unable to capture yields on smaller, intercropped plots and/or the treatment impact may be concentrated on larger plots.

Overall, our results suggest that there may indeed be a detectable V+R treatment impact among satellite derived yields, but that it is concentrated among larger pure-stand plots. It is possible that satellite estimates are not able to capture productivity of all plots if this productivity, as proxied by greenness of the leaves, is not high enough for the vegetation index to capture the light reflected.

VI Discussion

There may be two reasons for the lack of sustained fertilizer use over time. First, farmers may not have the liquidity to purchase fertilizers. Second, farmers may not have learned, or updated their beliefs, about fertilizer effectiveness. Our results suggest that the former reason is more likely. We show substantial heterogeneity in the treatment effects. Among farmers in the voucher and recommendation group, those in the top 20% of the yield distribution and those with better quality soils experienced yield increases. These farmers would have learned that fertilizer was profitable. However, farmers across all quintiles of productivity almost uniformly return to the very low baseline levels of fertilizer application by 2019, including the farmers in the top 20% of the distribution. The fact that high productivity farmers discontinued using fertilizer suggests that the ability to finance fertilizer is an obstacle even for farmers who have learned that fertilizer is profitable and even for the most productive farmers.

The difference between results based on self-reported yields and yields derived from satellite data suggest caution in using satellite data for several reasons. First, self-reported data may be needed as an input into satellite-based calculations, for example farmers need to define plot boundaries in order to calculate plot size based on GPS. Mismatch in time periods for which self-reported and satellite-based data are available can introduce errors, for example through difficulties with recall of plot boundaries from past years in the case of our project. Recall issues may be the reason why satellite-based estimates show no average treatment effect in 2016, because GPS estimates of plot size based on recalled boundaries for 2016 are larger on average than are plot sizes self-reported in 2016. This seems to be a recall issue rather than a bias in self-reporting because GPS estimates based on boundaries reported for plots cultivated in 2019 match self-reported plot sizes in 2019 well.

Second, satellite-based yield estimates on small plots should be used with caution pending further research on the validity of such estimates. The discrepancy between the satellite-derived and GPS-corrected yield and self-reported yields for 2016 may be explained by the lack of accuracy over small spatial scales and/or pure-stand versus intercropped plots. Indeed, when we restrict our sample to pure-stand plots that are greater than 1.75 acres, we do also see a significant treatment impact in 2016 for the voucher and recommendation group using satellite-derived yields. Further refining the methodology to collect satellite-derived yields on small plots is an important avenue for future research.

Third, availability of satellite-based data can place limitations on the analysis. We could not compare changes over time, analyze first differences, or control for baseline yields because we lack satellite-based estimates for our baseline year, 2014.

VII Conclusion

We examine the longer-term effects of alleviating credit and information constraints on fertilizer adoption and maize productivity. The results indicate that fertilizer use reverts to baseline as early as one year after the intervention concludes.

The low percentages of farmers who apply fertilizer at baseline and in 2019, three years after the intervention ended, highlight a persistent lack of fertilizer use in rural Tanzania and raise concerns about current programs that find short-term impacts on fertilizer adoption and productivity. Only 1.4% of farmers in our sample applied fertilizer on their 2014 main maize plot in 2019, and 0.8% applied it at baseline in 2014.

Our results suggest that credit constraints, uninsured risk, and/or preferences that result in lack of liquidity at the beginning of the planting season continue to constrain fertilizer use even once information about appropriate fertilizer types and amounts is available. Future research should examine effective ways to relax these constraints over the long run, and how effective approaches depend on changing climate risks.

Finally, future research is needed to better understand the potential for remote sensing to provide more objective measurements of yields and land area. In our study, the discrepancies between farmer reports of yields and satellite based measures, which contradict substantial increases in fertilizer use, confirm questions about reliability of satellite measures over small areas raised in the literature. Future research should examine the reliability of satellite-based yield estimates on individual farm plots.

References

Abay, K. A., G. T. Abate, C. B. Barrett, and T. Bernard (2019). Correlated non-classical measurement errors, 'second best' policy inference, and the inverse size-productivity relationship in agriculture. *Journal of Development Economics 139*, 171–184.

Beaman, L., A. BenYishay, J. Magruder, and A. M. Mobarak (2021). Can network theory-based targeting increase technology adoption? *American Economic Review 111*(6), 1918–43.

BenYishay, A. and A. M. Mobarak (2018). Social learning and incentives for experimentation and communication. *The Review of Economic Studies 86*(3), 976–1009.

Carletto, C., S. Gourlay, and P. Winters (2015). From Guesstimates to GPStimates: Land Area Measurement and Implications for Agricultural Analysis. *Journal of African Economies 24*(5), 593–628.

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. Laajaj, and D. Yang (2021). Subsidies and the African Green Revolution: Direct effects and social network spillovers of randomized input subsidies in Mozambique. *American Economic Journal: Applied Economics 13*(2), 206–29.

Cohen, A. (2019). Estimating farm production parameters with measurement error in land area. *Economic Development and Cultural Change 68*(1), 305–334.

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 98*(2), 482–88.

Duflo, E., M. Kremer, and J. Robinson (2011). Nudging farmers to use fertilizer: Theory and experimental evidence from Kenya. *American Economic Review 101*(6), 2350–90.

Emerick, K. and M. H. Dar (2021). Farmer field days and demonstrator selection for increasing technology adoption. *The Review of Economics and Statistics 103*(4), 680–693.

Firpo, S., N. M. Fortin, and T. Lemieux (2009). Unconditional quantile regressions. *Econometrica 77*(3), 953–973.

Fishman, R., S. C. Smith, V. Bobić, and M. Sulaiman (2021). Can agricultural extension and input support be discontinued? Evidence from a randomized phase-out in Uganda. *The Review of Economics and Statistics*, 1–45.

Ghanem, D., S. Hirshleifer, and K. Ortiz-Becerra (2021). Testing attrition bias in field experiments. *CEGA Working Paper Series No. WPS-113*.

Gollin, D. and C. Udry (2021). Heterogeneity, measurement error, and misallocation: Evidence from African agriculture. *Journal of Political Economy 129*(1), 1–80.

Harou, A. P., M. Madajewicz, H. Michelson, C. A. Palm, N. Amuri, C. Magomba, J. M. Semoka, K. Tschirhart, and R. Weil (2022). The joint effects of information and financing constraints on technology adoption: Evidence from a field experiment in rural Tanzania. *Journal of Development Economics 155*, 102707.

Harou, A. P., T. F. Walker, and C. B. Barrett (2017). Is late really better than never? The farmer welfare effects of pineapple adoption in Ghana. *Agricultural Economics 48*(2), 153–164.

Jin, Z., G. Azzari, C. You, S. Di Tommaso, S. Aston, M. Burke, and D. B. Lobell (2019). Smallholder maize area and yield mapping at national scales with google earth engine. *Remote Sensing of Environment 228*, 115 – 128.

Karlan, D., R. Osei, I. Osei-Akoto, and C. Udry (2014). Agricultural decisions after relaxing credit and risk constraints. *The Quarterly Journal of Economics 129*(2), 597–652.

Lobell, D. B., G. Azzari, M. Burke, S. Gourlay, Z. Jin, T. Kilic, and S. Murray (2020). Eyes in the sky, boots on the ground: Assessing satellite- and ground-based approaches to crop yield measurement and analysis. *American Journal of Agricultural Economics 102*(1), 202–219.

Marenya, P. P. and C. B. Barrett (2009). State-Conditional fertilizer yield response on Western Kenyan farms. *American Journal of Agricultural Economics 91*(4), 991–1006.

Suri, T. (2011). Selection and comparative advantage in technology adoption. *Econometrica 79*(1), 159–209.

Suri, T. and C. Udry (2022). Agricultural technology in Africa. *Journal of Economic Perspectives 36*(1), 33–56.

Xu, Z., Z. Guan, T. S. Jayne, and R. Black (2009). Factors influencing the profitability of fertilizer use on maize in Zambia. *Agricultural Economics 40*(4), 437–446.

Figures

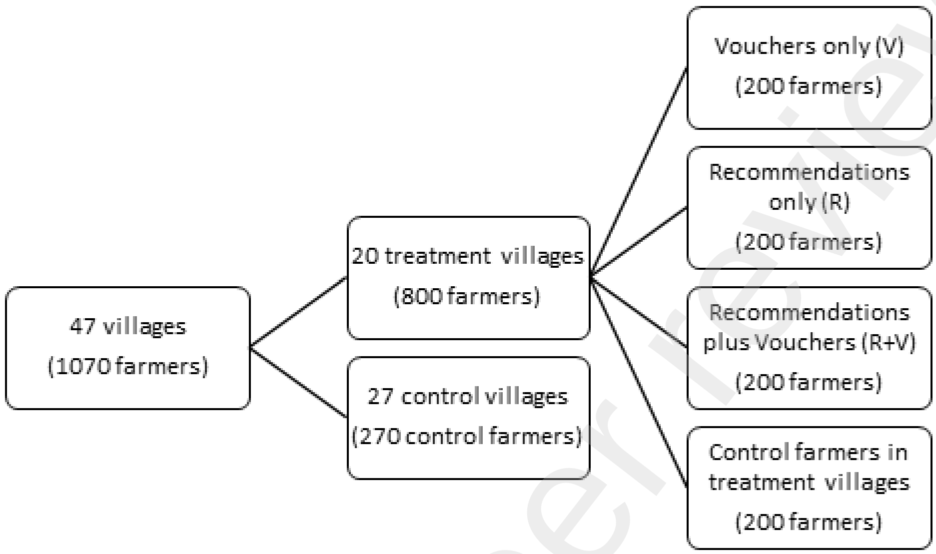


Figure (1) Randomization

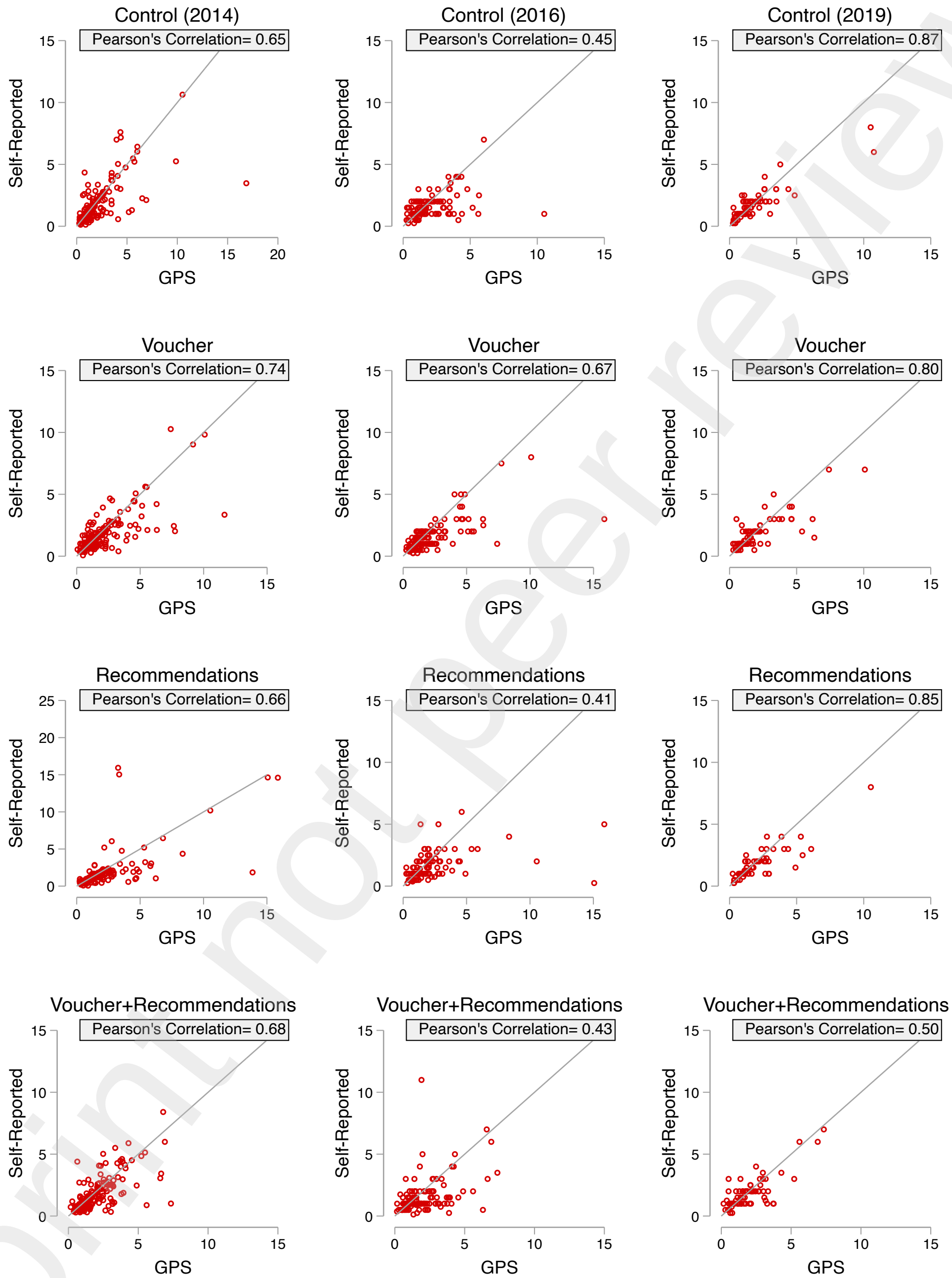


Figure (2) GPS and Self-Reported Areas by Treatment and Year (Acres)

Notes: This figure shows Pearson's Correlation coefficients for each treatment-year combination. These coefficients are displayed at the top of each figure.

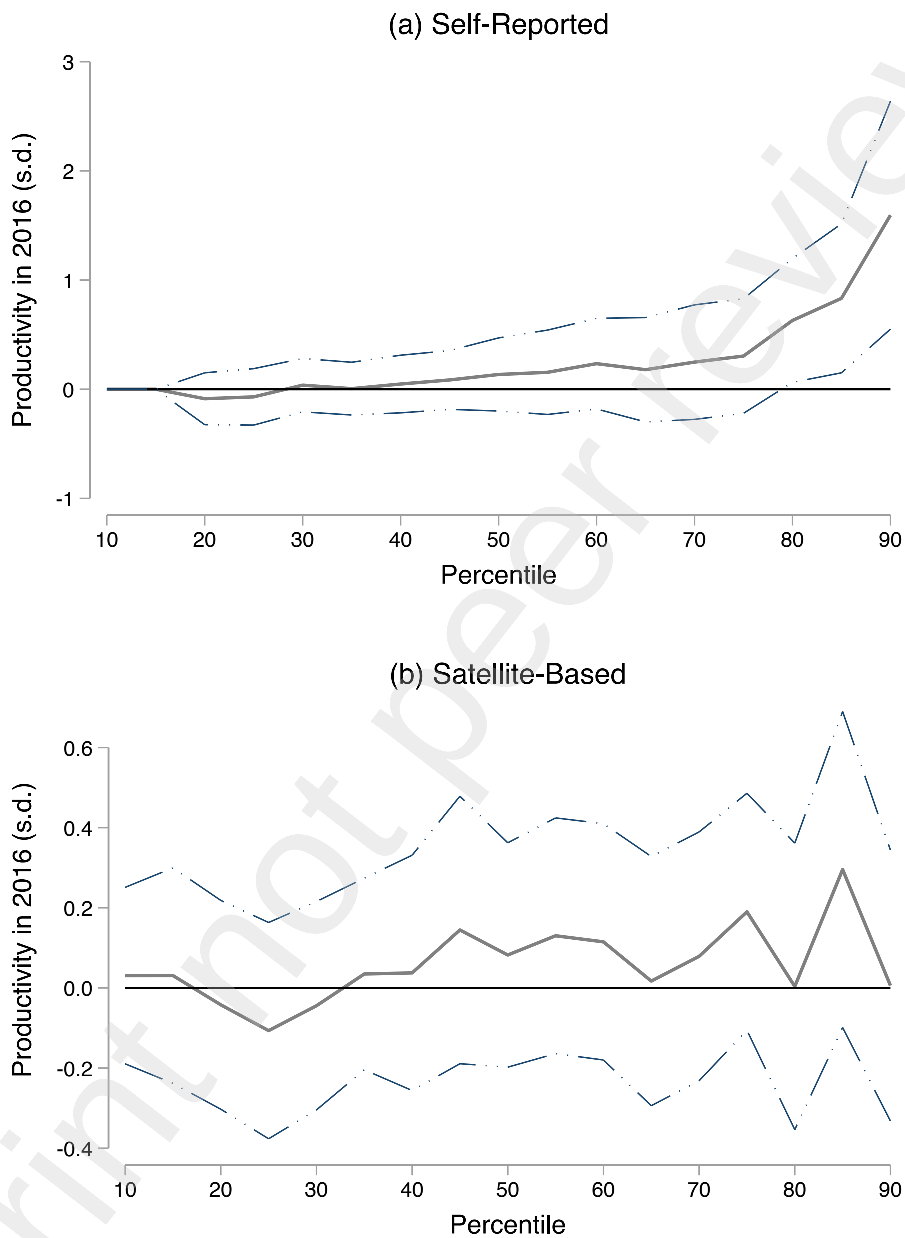
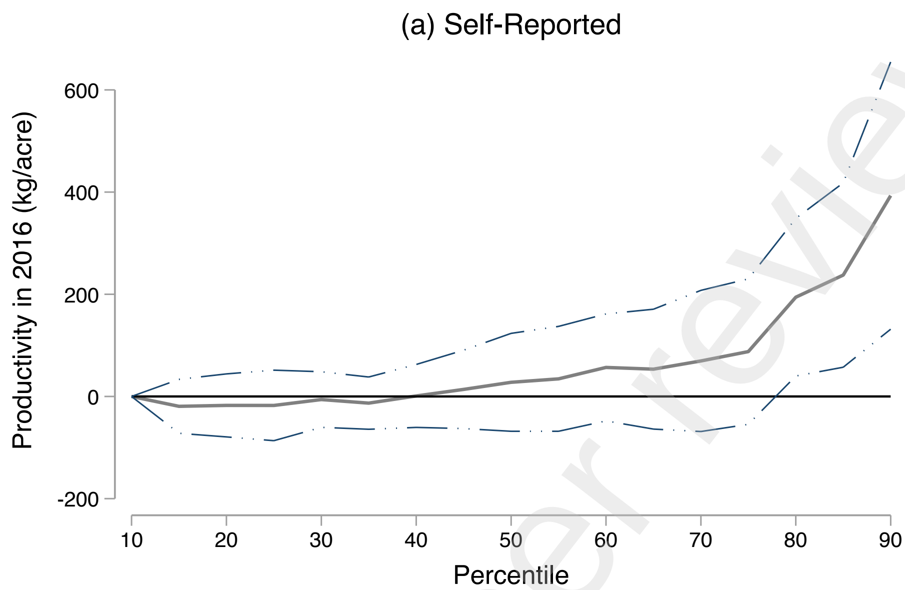


Figure (3) Quantile Treatment Effect on Self-Reported and Satellite-Based Productivity
Notes: Vertical lines are 95% confidence intervals. Each coefficient represents a quantile treatment effect for the 2014 MMP in 2016 in columns (1) and (2) of Table 5, following Firpo et al. (2009). We display the V+R treatment estimates only. The regression employs the same OLS used in the main analysis when comparing self-reported and satellite-derived productivity ($N = 491$). All regressions control for village fixed effects. Standard errors are clustered at the village-level.



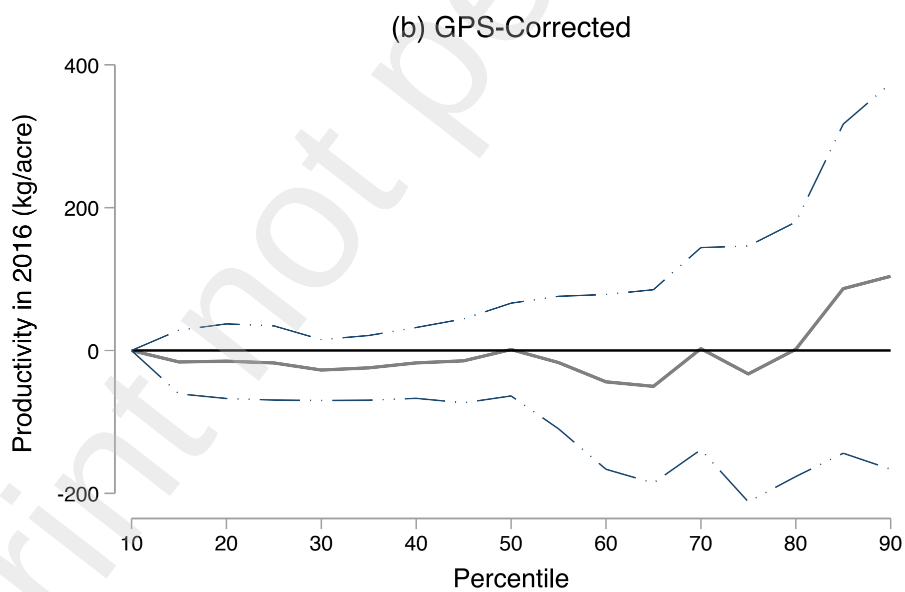


Figure (4) Quantile Treatment Effect on Self-Reported and GPS-Corrected Productivity

Notes: Vertical lines are 95% confidence intervals. Each coefficient represents a quantile treatment effect for the 2014 MMP in 2016 in columns (1) and (2) of Table 6, following Firpo et al. (2009). We display the V+R treatment estimates only. The regression employs the same ANCOVA used in the main analysis when comparing self-reported and GPS-corrected productivity ($N = 431$). All regressions control for village fixed effects. Standard errors are clustered at the village-level.

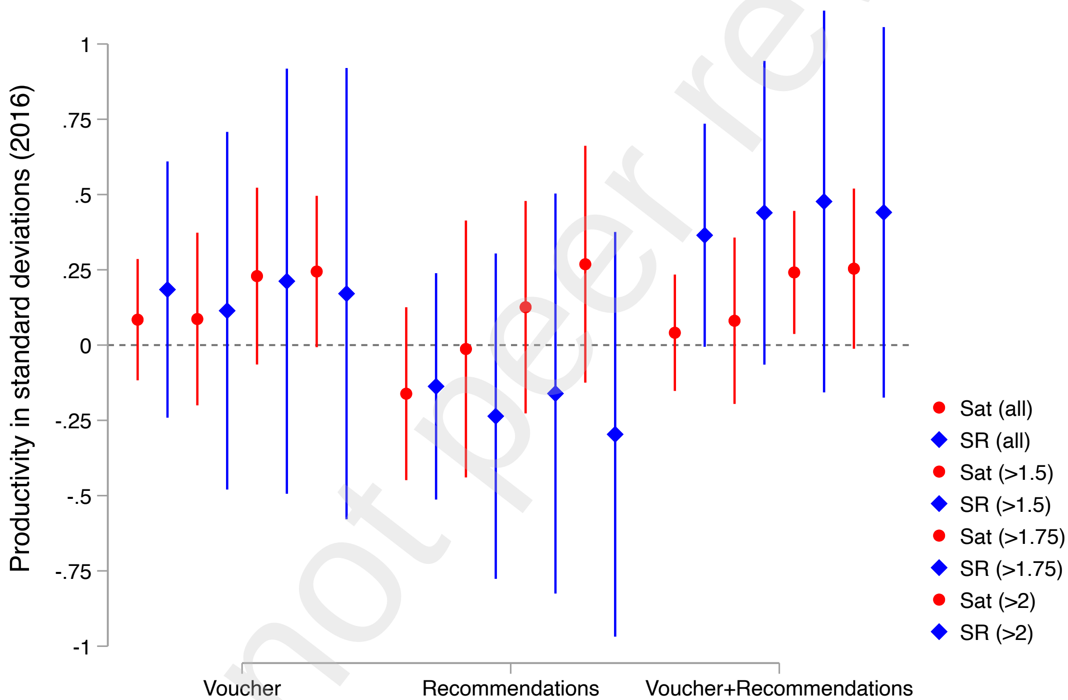


Figure (5) Heterogeneous Treatment Effect on Pure-Stand Plots by GPS Areas

Notes: Vertical lines are 95% confidence intervals. Each coefficient represents an OLS regression of average treatment effect on pure-stand plots only in 2016. The last three regressions restrict the sample to plots of GPS areas larger than 1.5, 1.75, and 2 acres, respectively. All regressions control for village fixed effects. Standard errors are clustered at the village-level.

Tables

Table (1) Panel Retention

	Baseline 2014	During intervention 2015		Immediate Post 2016		Longer-Term Post 2019	
	N	N	Retention	N	Retention	N	Retention
Control	190	147	77.4%	179	94.2%	165	86.8%
Voucher	198	155	78.3%	187	94.5%	178	89.9%
Recommendations	191	138	72.3%	177	92.7%	162	84.8%
Voucher+Recommendations	203	157	77.3%	190	93.6%	175	86.2%
Control Villages	268	244	91.0%	251	93.7%	240	89.6%
Total	1,050	841	80.1%	984	93.7%	920	87.0%

Notes: This table presents numbers of households that were surveyed in the four survey rounds.

Table (2) Within-Village Balance Tests

		Control Mean	Treatment coefficient			F-Test (p-value)			N
			V	R	V+R	V = R	V = V+R	R = V+R	
Panel A: Outcomes									
(1)	Fertilizer (kg/ SR acre)	0.02	0.54	0.30	-0.07	0.7	0.17	0.38	491
			(0.41)	(0.37)	(0.06)				
(2)	Fertilizer (kg/ GPS acre)	0.01	0.21	0.47	-0.07	0.68	0.27	0.35	486
			(0.24)	(0.52)	(0.07)				
(3)	Fertilizer (=1)	0.01	0.00	0.01	-0.00	0.83	0.66	0.51	491
			(0.01)	(0.01)	(0.01)				
(4)	Yields (kg/ SR acre)	403.2	-34.53	50.77	-37.19	0.09	0.91	0.05	435
			(31.37)	(59.71)	(37.35)				
(5)	Yields (kg/ GPS acre)	410.02	-58.12	-36.78	-36.84	0.53	0.54	1	433
			(58.55)	(64.69)	(78.18)				
Panel B: Covariates									
(6)	Male-Head (=1)	0.87	-0.04	-0.01	-0.00	0.46	0.31	0.77	491
			(0.04)	(0.05)	(0.05)				
(7)	Head Age (Years)	44.29	-1.45	1.36	1.61	0.09	0.03	0.88	491
			(1.74)	(1.37)	(1.35)				
(8)	Head Education (=1 if some education)	0.9	0.02	0.00	-0.03	0.57	0.26	0.53	491
			(0.03)	(0.04)	(0.03)				
(9)	Head Education (=1 if beyond primary)	0.05	-0.00	-0.04*	0.00	0.09	0.94	0.03	491
			(0.04)	(0.02)	(0.03)				
(10)	Distance to plot in minutes	34.02	-0.82	-6.13	2.39	0.18	0.35	0.1	458
			(2.36)	(3.82)	(4.78)				
(11)	Credit Access (=1)	0.12	-0.03	-0.05	-0.07*	0.59	0.33	0.58	491
			(0.03)	(0.03)	(0.04)				
(12)	Remittances (=1)	0.14	0.01	-0.03	0.04	0.49	0.55	0.07	491
			(0.05)	(0.06)	(0.06)				
(13)	Asset Index	-.17	0.07	0.06	-0.00	0.96	0.6	0.8	491
			(0.21)	(0.28)	(0.18)				
(14)	Livestock Ownership (=1)	0.82	-0.13**	-0.10**	-0.11*	0.52	0.73	0.86	491
			(0.06)	(0.04)	(0.06)				
(15)	Household Size	5.1	-0.21	0.09	0.10	0.38	0.38	0.97	491
			(0.34)	(0.31)	(0.36)				
(16)	Area Owned (SR acres)	5.59	-0.17	0.13	-0.47	0.54	0.56	0.21	491
			(0.64)	(0.76)	(0.77)				
(17)	Close to Chairman (=1)	0.33	0.00	-0.01	-0.03	0.82	0.49	0.71	491
			(0.05)	(0.05)	(0.06)				
(18)	Received Training (=1)	0.07	0.01	-0.01	0.04	0.32	0.42	0.11	491
			(0.02)	(0.02)	(0.03)				
(19)	Visited by Extension (=1)	0.17	0.02	-0.03	0.01	0.4	0.91	0.52	491
			(0.05)	(0.06)	(0.07)				
(20)	Maize Area (SR acres)	2.11	0.07	-0.13	-0.33	0.34	0.02	0.18	487
			(0.22)	(0.15)	(0.21)				
(21)	Improved Seeds (=1)	0.15	0.01	-0.01	0.02	0.57	0.83	0.46	491
			(0.05)	(0.04)	(0.05)				

Notes: V stands for the Voucher treatment; R for the Recommendations treatment; V+R for the Voucher plus Recommendations treatment. The first three columns report results of baseline balance tests in which we estimate $b_{iv} = \alpha_0 + \sum_{k=1}^{3} \theta_k TREAT_i^k + d_v + \varepsilon_{iv}$. The next three columns test the equality of coefficients between the three treatments. Robust standard errors in parentheses. SR stands for self-reported, as answered by respondents in the socioeconomic survey. Standard errors are clustered at the village-level. *** p<0.01, ** p<0.05, * p<0.1. All regressions control for village fixed effects.

Table (3) Households' Cultivation Behavior

	2014	2016	2019	Total
Cultivated the 2014 MMP	915	822	416	3,273
Regarded the same 2014 MMP as their MMP in 2016 and 2019	915	814	338	
Cultivated a new MMP in 2016 and 2019	0	74	369	
Total	915	888	707	4,002

Notes: MMP stands for the main maize plot from which soil samples were taken in 2014. It is defined as the lot that is most important for the household i terms of food security and income generation. The table shows the cultivation behavior of households in 2014, 2016, and 2019 including the new plots that they switch to cultivate over time.

Table (4) Average Treatment Effect on Fertilizer Adoption

	(1) Fertilizer (kg/SR acre)	(2) Fertilizer (kg/GPS acre)	(3) Fertilizer (=1)	(4) Fertilizer (kg/SR acre)	(5) Fertilizer (kg/GPS acre)	(6) Fertilizer (=1)
	2016	2016	2016	2019	2019	2019
Voucher	8.65***	7.64**	0.30***	0.38	-0.55	-0.00
	(2.77)	(2.79)	(0.07)	(1.45)	(1.42)	(0.02)
Recommendations	-0.15	-0.26	0.04	-1.29	-1.79	-0.01
	(1.19)	(0.86)	(0.02)	(1.02)	(1.44)	(0.03)
Voucher + Recommendations	27.15***	25.07***	0.71***	-1.48	-1.69	-0.02
	(3.24)	(5.75)	(0.03)	(0.94)	(1.15)	(0.02)
Baseline value	0.71**	0.92**	0.19	-0.08***	-0.05***	-0.04
	(0.31)	(0.39)	(0.23)	(0.01)	(0.01)	(0.03)
Control mean	0.02	0.01	0.04	0.00	0.00	0.03
(Std. dev.)	(0.24)	(0.13)	(0.18)	(0)	(0)	(0.16)
V=R (p-value)	0.01	0.01	0	0.15	0.13	0.69
V=V+R (p-value)	0	0.01	0	0.18	0.18	0.23
R=V+R (p-value)	0	0	0	0.63	0.81	0.4
N	472	472	589	292	292	301
Village FE	✓	✓	✓	✓	✓	✓

Notes: This table presents ANCOVA results from estimating $Y_{iv} = \alpha_0 + \varphi Y_{i0} + \sum_{k=1}^{3} \gamma_k TREAT_i^k + d_v + \varepsilon_{iv}$ in which we exclude the pure control villages. Robust standard errors in parentheses and are clustered at the village- level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. SR stands for self-reported, as answered by respondents in the socioeconomic survey. The control mean refers to the average of the control group at baseline. V stands for the Voucher treatment; R for the Recommendations treatment; V+R for the Voucher plus Recommendations treatment. All regressions control for village fixed effects.

Table (5) Average Treatment Effect on Self-Reported and Satellited-Based Productivity

	(1) Self-Reported (z-score)	(2) Satellite-Based (z-score)	(3) Self-Reported (z-score)	(4) Satellite-Based (z-score)	(5) Self-Reported (z-score)	(6) Satellite-Based (z-score)	(7) Self-Reported (z-score)	(8) Satellite-Based (z-score)
	2016	2016	2019	2019	2016	2016	2019	2019
Voucher	0.176	0.045	0.167	-0.097	0.183	0.043	0.164	-0.103
	(0.181)	(0.076)	(0.128)	(0.131)	(0.181)	(0.076)	(0.127)	(0.127)
Recommendations	-0.022	-0.057	0.079	-0.022	-0.030	-0.054	0.097	-0.037
	(0.115)	(0.121)	(0.193)	(0.186)	(0.116)	(0.120)	(0.191)	(0.187)
Voucher+Recommendations	0.440**	0.051	0.194	-0.064	0.444**	0.053	0.205	-0.086
	(0.168)	(0.088)	(0.145)	(0.119)	(0.167)	(0.086)	(0.139)	(0.116)
Pure stand (=1)					-0.150	-0.056	0.159	0.028
					(0.175)	(0.073)	(0.161)	(0.108)
Plantations (=1)					-0.110	0.145	-0.086	0.209
					(0.109)	(0.084)	(0.121)	(0.204)
Control mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(Std. dev.)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
V=R (p-value)	0.17	0.38	0.56	0.69	0.14	0.4	0.65	0.73
V=V+R (p-value)	0.09	0.96	0.79	0.84	0.09	0.93	0.7	0.91
R=V+R (p-value)	0	0.39	0.32	0.8	0	0.39	0.35	0.77
N	491	491	294	294	491	491	294	294
Village FE	✓	✓	✓	✓	✓	✓	✓	✓

Notes: This table presents OLS results from estimating $Y_{iv} = \alpha_1 + \sum_{k=1}^{3} \beta_k TREAT_i^k + d_v + \varepsilon_{iv}$ in which we exclude the pure control villages. All yields are measured in standard deviation units to be able to compare self-reported and satellite-based measures that use different units. Robust standard errors in parentheses and are clustered at the village-level. *** p<0.01, ** p<0.05, * p<0.1. The control mean refers to the average of the control group at endline because satellites are unable to predict yields at baseline. V stands for the Voucher treatment; R for the Recommendations treatment; V+R for the Voucher plus Recommendations treatment. All regressions control for village fixed effects.

Table (6) Average Treatment Effect on Self-Reported and GPS-Corrected Productivity

	(1) Self-Reported (kg/acre)	(2) GPS-Corrected (kg/acre)	(3) Self-Reported (kg/acre)	(4) GPS-Corrected (kg/acre)	(5) Self-Reported (kg/acre)	(6) GPS-Corrected (kg/acre)	(7) Self-Reported (kg/acre)	(8) GPS-Corrected (kg/acre)
					>0.5	>0.5	>0.5	>0.5
	2016	2016	2019	2019	2016	2016	2019	2019
Voucher	50.40	28.53	39.44	42.54	45.16	32.74	6.31	12.10
	(46.78)	(45.03)	(37.93)	(45.29)	(49.78)	(47.40)	(34.19)	(40.05)
Recommendations	-22.77	-47.14*	7.80	-42.18	-23.94	-40.30	-37.44	-71.81
	(26.75)	(23.36)	(52.43)	(48.86)	(27.60)	(26.22)	(53.60)	(53.14)
Voucher+Recommendations	100.97**	41.33	24.46	42.82	102.62**	43.18	-7.07	-8.98
	(41.94)	(54.50)	(43.60)	(49.29)	(43.83)	(59.28)	(40.83)	(45.87)
Baseline value	0.15*	0.21**	0.15*	0.17***	0.15	0.23**	0.15	0.17**
	(0.09)	(0.08)	(0.08)	(0.06)	(0.09)	(0.09)	(0.09)	(0.07)
Control mean	403.13	410.02	378.05	361.85	397.68	406.53	373.75	359.06
(Std. dev.)	(311.34)	(466.19)	(328.87)	(331.08)	(314.4)	(474.48)	(334.17)	(336.19)
V=R (p-value)	0.06	0.07	0.44	0.15	0.09	0.09	0.32	0.19
V=V+R (p-value)	0.19	0.75	0.57	0.99	0.21	0.8	0.64	0.45
R=V+R (p-value)	0	0.09	0.63	0.12	0.01	0.15	0.44	0.28
N	431	431	271	271	409	409	250	250
Village FE	✓	✓	✓	✓	✓	✓	✓	✓

Notes: This table presents ANCOVA results from estimating $Y_{iv} = \alpha_0 + \varphi Y_{i0} + \sum_{k=1}^{3} \gamma_k TREAT_i^k + d_v + \varepsilon_{iv}$ in which we exclude the pure control villages. Robust standard errors in parentheses and are clustered at the village-level. *** p<0.01, ** p<0.05, * p<0.1. The control mean refers to the average of the control group at baseline. V stands for the Voucher treatment; R for the Recommendations treatment; V+R for the Voucher plus Recommendations treatment. All regressions control for village fixed effects.

Table (7) Fertilizer Adoption Rates by Productivity Quintiles

	2014			2016			2019		
Quintile	Mean	SD	N	Mean	SD	N	Mean	SD	N
First	0	0	113	28.57	45.36	126	0	0	51
Second	.84	9.17	119	27.34	44.75	128	1.82	13.48	55
Third	.91	9.53	110	28.1	45.14	121	1.85	13.61	54
Fourth	0	0	111	32.52	47.04	123	3.64	18.89	55
Fifth	3.54	18.56	113	49.18	50.2	122	1.75	13.25	57
All	1.06	10.25	566	33.06	47.08	620	1.84	13.46	272

Notes: This tables shows percentages of farmers who applied any mineral fertilizer in all three rounds by quintiles of the 2016 self-reported productivity. SD stands for the standard deviation.

Table (8) Heterogeneous Treatment Effect on Adoption by Soil Quality

	(1) Fertilizer (kg/SR acre)	(2) Fertilizer (kg/GPS acre)	(3) Fertilizer (=1)
	2016	2016	2016
Panel A: Poor soil quality (θ_k)			
Voucher	10.544*	9.399*	0.280***
	(5.767)	(4.718)	(0.094)
Recommendations	-1.257	0.034	0.023
	(1.787)	(1.816)	(0.028)
Voucher + Recommendations	22.357***	16.594***	0.644***
	(3.509)	(3.281)	(0.058)
V=R (p-value)	0.07	0.08	0.02
V=V+R (p-value)	0.12	0.19	0
R=V+R (p-value)	0	0	0
Panel B: Rich soil quality ($\theta_k + \gamma_k$)			
Voucher	8.128***	6.009*	0.314***
	(2.188)	(3.315)	(0.064)
Recommendations	1.053	-1.076	0.054
	(1.154)	(1.581)	(0.040)
Voucher + Recommendations	32.982***	35.239***	0.786***
	(5.784)	(10.774)	(0.049)
V=R (p-value)	0.02	0.03	0
V=V+R (p-value)	0	0.02	0
R=V+R (p-value)	0	0.01	0
N	459	459	573
Control mean	0.02	0.01	0.01
(Std. dev.)	(0.24)	(0.13)	(0.09)
Village FE	✓	✓	✓

Notes: This table presents results of $Y_{iv} = \alpha_0 + \sum_{k=1}^{3} \theta_k TREAT_i^k + \gamma RichSoil_i + \sum_{k=1}^{3} \gamma_k TREAT_i^k \times RichSoil_i + d_v + \varepsilon_{iv}$ estimated using ANCOVA in which we exclude the pure control villages. Robust standard errors in parentheses are clustered at the village-level. *** p<0.01, ** p<0.05, * p<0.1. Soils are classified into fertile or poor based on their electrical conductivity. Fertile soils are the ones that have optimal salinity (slightly or very saline), whereas poor soils have low, medium, or severe salinity. V stands for the Voucher treatment; R for the Recommendations treatment; V+R for the Voucher plus Recommendations treatment.

Table (9) Heterogeneous Treatment Effect on Productivity by Soil Quality

	(1) Self-Reported (z-score)	(2) Satellite-Based (z-score)	(3) GPS-Corrected (z-score)	(4) Self-Reported (z-score)	(5) Satellite-Based (z-score)	(6) GPS-Corrected (z-score)
	2016	2016	2016	2019	2019	2019
Panel A: Poor soil quality (θ_k)						
Voucher	0.243	-0.063	0.103	0.122	-0.307	0.113
	(0.363)	(0.186)	(0.281)	(0.179)	(0.288)	(0.137)
Recommendations	-0.115	-0.226	-0.144	-0.100	-0.477	-0.131
	(0.206)	(0.242)	(0.188)	(0.207)	(0.380)	(0.182)
Voucher+Recommendations	0.196	-0.061	0.004	0.090	-0.117	0.153
	(0.225)	(0.189)	(0.275)	(0.210)	(0.224)	(0.129)
V=R (p-value)	0.3	0.48	0.43	0.24	0.68	0.26
V=V+R (p-value)	0.87	0.99	0.72	0.86	0.47	0.81
R=V+R (p-value)	0.19	0.49	0.58	0.36	0.19	0.21
Panel B: Rich soil quality ($\theta_k + \gamma_k$)						
Voucher	0.189	0.115	0.115	0.243	0.065	0.288
	(0.144)	(0.165)	(0.167)	(0.214)	(0.201)	(0.269)
Recommendations	0.055	0.103	-0.058	0.310	0.354**	0.066
	(0.179)	(0.116)	(0.209)	(0.275)	(0.173)	(0.259)
Voucher+Recommendations	0.709***	0.217	0.344	0.426*	-0.042	0.414
	(0.246)	(0.162)	(0.256)	(0.246)	(0.212)	(0.290)
V=R (p-value)	0.42	0.93	0.35	0.76	0.04	0.3
V=V+R (p-value)	0.02	0.59	0.3	0.39	0.62	0.53
R=V+R (p-value)	0.02	0.44	0.18	0.48	0.07	0.08
N	478	478	478	284	284	284
Control mean	0.00	-0.01	-0.01	-0.05	0.00	-0.05
(std. dev.)	(1.01)	(1.00)	(0.98)	(0.91)	(1.02)	(0.90)
Village FE	✓	✓	✓	✓	✓	✓

Notes: This table presents results of $Y_{iv} = \alpha_0 + \sum_{k=1}^{3} \theta_k TREAT_i^k + \gamma RichSoil_i + \sum_{k=1}^{3} \gamma_k TREAT_i^k \times RichSoil_i + d_v + \varepsilon_{iv}$ estimated using OLS in which we exclude the pure control villages. All yields are measured in standard deviation units to be able to compare self-reported and satellite-based measures that use different units. Robust standard errors in parentheses are clustered at the village-level. *** p<0.01, ** p<0.05, * p<0.1. Soils are classified into fertile or poor based on their electrical conductivity. Fertile soils are the ones that have optimal salinity (slightly or very saline), whereas poor soils have low, medium, or severe salinity. V stands for the Voucher treatment; R for the Recommendations treatment; V+R for the Voucher plus Recommendations treatment.

Appendix (For Online Publication)

Balance

Table (A1) Cross-Village Balance Tests

		Control	Treatment coefficient			F-Test (p-value)			N
		Mean	V	R	V+R	V = R	V = V+R	R = V+R	
Panel A: Outcomes									
(1)	Fertilizer (kg/ SR acre)	0.04	0.63	0.41	-0.01	0.73	0.17	0.34	659
			(0.46)	(0.43)	(0.04)				
(2)	Fertilizer (kg/ GPS acre)	0.02	0.31	0.58	-0.00	0.69	0.29	0.33	653
			(0.30)	(0.59)	(0.02)				
(3)	Fertilizer (=1)	0.01	0.00	0.01	-0.00	0.86	0.59	0.49	659
			(0.01)	(0.01)	(0.01)				
(4)	Yields (kg/ SR acre)	481.02	-117.97***	-26.76	-107.82***	0.07	0.68	0.05	578
			(41.55)	(54.81)	(39.23)				
(5)	Yields (kg/ GPS acre)	435.74	-85.30**	-55.85	-48.17	0.41	0.32	0.82	575
			(36.16)	(47.48)	(53.39)				
Panel B: Covariates									
(6)	Male-Head (=1)	0.84	-0.01	0.02	0.03	0.41	0.27	0.77	659
			(0.04)	(0.04)	(0.04)				
(7)	Head Age (Years)	45.92	-2.94*	-0.23	0.15	0.05	0.02	0.81	659
			(1.56)	(1.37)	(1.46)				
(8)	Head Education (=1 if some education)	0.92	0.00	-0.02	-0.05	0.49	0.24	0.56	659
			(0.03)	(0.03)	(0.03)				
(9)	Head Education (=1 if beyond primary)	0.07	-0.01	-0.06**	-0.01	0.08	0.95	0.03	659
			(0.03)	(0.02)	(0.02)				
(10)	Distance to plot in minutes	33.97	-0.77	-6.04*	3.29	0.23	0.24	0.1	587
			(3.24)	(3.36)	(4.73)				
(11)	Credit Access (=1)	0.11	-0.01	-0.03	-0.05*	0.41	0.21	0.56	659
			(0.03)	(0.03)	(0.03)				
(12)	Remittances (=1)	0.18	-0.03	-0.06	-0.00	0.46	0.57	0.08	659
			(0.04)	(0.05)	(0.05)				
(13)	Asset Index	0.19	-0.29	-0.33	-0.36	0.85	0.56	0.89	659
			(0.24)	(0.29)	(0.24)				
(14)	Livestock Ownership (=1)	0.79	-0.09*	-0.07	-0.08	0.67	0.89	0.81	659
			(0.05)	(0.04)	(0.05)				
(15)	Household Size	5.42	-0.62*	-0.30	-0.20	0.36	0.21	0.62	659
			(0.31)	(0.26)	(0.30)				
(16)	Area Owned (SR acres)	5.21	-0.07	0.38	-0.03	0.37	0.94	0.38	659
			(0.44)	(0.65)	(0.62)				
(17)	Close to Chairman (=1)	0.32	0.03	0.02	-0.01	0.9	0.45	0.61	659
			(0.05)	(0.05)	(0.05)				
(18)	Received Training (=1)	0.07	0.01	-0.01	0.04	0.45	0.33	0.11	659
			(0.02)	(0.02)	(0.04)				
(19)	Visited by Extension (=1)	0.21	-0.03	-0.06	-0.02	0.49	0.92	0.48	659
			(0.05)	(0.05)	(0.05)				
(20)	Maize Area (SR acres)	2.07	0.03	-0.11	-0.27*	0.47	0.05	0.28	655
			(0.19)	(0.15)	(0.14)				
(21)	Improved Seeds (=1)	0.21	-0.04	-0.05	-0.03	0.89	0.8	0.67	659
			(0.05)	(0.04)	(0.06)				

Notes: V stands for the Voucher treatment; R for the Recommendations treatment; V+R for the Voucher plus Recommendations treatment. The first three columns report results of baseline balance tests in which we estimate $b_i = \alpha_0 + \sum_{k=1}^{3} \theta_k TREAT_i^k + \varepsilon_i$ including control villages. The next three columns test the equality of coefficients between the three treatments. Robust standard errors in parentheses. SR stands for self-reported, as answered by respondents in the socioeconomic survey. Standard errors are clustered at the village-level.

Attrition Analysis

Table (A2) Probability of Attrition by Treatment

	Attrition (=1 if attrited)			
	2016	2016	2019	2019
Voucher	-0.00234	-0.00326	-0.0306	-0.0314
	(0.0219)	(0.0217)	(0.0294)	(0.0300)
Recommendations	0.0154	0.0157	0.0203	0.0201
	(0.0220)	(0.0220)	(0.0333)	(0.0332)
Voucher+Recommendations	0.00614	0.00539	0.00635	0.00423
	(0.0179)	(0.0180)	(0.0262)	(0.0264)
Constant	0.0579***		0.132***	
	(0.0142)		(0.0231)	
N	782	782	782	782
Village FE		✓		✓

Notes: Robust standard errors in parentheses. Standard errors are clustered at the village-level. *** p<0.01, ** p<0.05, * p<0.1. Control villages are excluded.

Table (A3) Non-Missing Values by Treatment

	(1) Self-Reported (=1)	(2) Satellite-Based (=1)	(3) GPS-Corrected (=1)	(4) Self-Reported (=1)	(5) Satellite-Based (=1)	(6) GPS-Corrected (=1)
	2016	2016	2016	2019	2019	2019
Voucher	0.076*	0.091	0.077	0.016	0.006	0.006
	(0.040)	(0.053)	(0.048)	(0.048)	(0.044)	(0.044)
Recommendations	0.011	0.002	-0.009	-0.050	-0.050	-0.050
	(0.036)	(0.050)	(0.047)	(0.042)	(0.043)	(0.043)
Voucher+Recommendations	0.061	0.055	0.041	-0.043	-0.048	-0.048
	(0.037)	(0.035)	(0.038)	(0.047)	(0.048)	(0.048)
Control mean	0.76	0.62	0.60	0.41	0.40	0.40
N	782	782	782	782	782	782
Village FE	✓	✓	✓	✓	✓	✓

Notes: Notes: Robust standard errors in parentheses. Standard errors are clustered at the village-level. *** p<0.01, ** p<0.05, * p<0.1. All regressions control for village fixed effects. Control villages are excluded.

Table (A4) Internal Validity in the Presence of Non-Classical Attrition

	(1) Self-Reported (kg/acre)	(2) GPS-Corrected (kg/acre)	(3) Self-Reported (kg/acre)	(4) GPS-Corrected (kg/acre)
	2016	2016	2019	2019
Response (= 1)	-26.38	56.55	3.12	-24.24
	(75.79)	(97.48)	(48.42)	(66.34)
Voucher	-4.66	26.81	14.47	-46.70
	(118.56)	(71.77)	(47.54)	(78.26)
Recommendations	-40.20	17.70	23.25	-48.47
	(72.08)	(88.29)	(39.60)	(84.62)
Voucher+Recommendations	-67.91	-26.80	-17.04	-23.79
	(50.37)	(54.67)	(40.98)	(104.25)
Response×Voucher	-1.11	-81.89	-49.78	23.52
	(134.88)	(106.80)	(53.11)	(84.13)
Response×Recommendations	94.33	-53.58	19.91	50.41
	(100.74)	(105.74)	(72.64)	(86.48)
Response×Voucher+Recommendations	57.32	-9.87	-14.35	-18.22
	(70.67)	(96.83)	(59.16)	(111.69)
Mean baseline control attritors	429.07	368.73	404.82	361.38
IV-R Test: V=Response×V=0 (p-value)	0.99	0.64	0.57	0.52
IV-R Test: R=Response×R=0 (p-value)	0.51	0.84	0.62	0.83
IV-R Test: R+V=Response×R+V=0 (p-value)	0.33	0.78	0.64	0.68
IV-P Test: Response=V=Response×V =0 (p-value)	0.97	0.81	0.52	0.69
IV-P Test: Response=R=Response×R =0 (p-value)	0.66	0.94	0.8	0.94
IV-P Test: Response=R+V=Response×R+V =0 (p-value)	0.41	0.67	0.72	0.69
N	688	555	688	555
Village FE	✓	✓	✓	✓

Notes: This table shows tests for attrition bias using a more flexible form detailed in Ghanem et al. (2021) and includes all missing values for not only the classical attritors who were not interviewed, but also non-classical cases such as those who stopped cultivating maize (see Table 3) and those who were unreachable to collect GPS polygons. The estimation is done using OLS following the regression $Y_{iv} = \alpha_0 + \sum_{k=1}^{3} \theta_k TREAT_i^k + Response_i + \sum_{k=1}^{3} \gamma_k TREAT_i^k \times Response_i + d_v + \varepsilon_{iv}$ and excludes the control villages. IV-R refers to internal validity for the respondent subpopulation and IV-P refers to internal validity for the study population as defined in Ghanem et al. (2021). Robust standard errors in parentheses. Standard errors are clustered at the village-level. All regressions control for village fixed effects.

Robustness

Table (A5) Pooled Treatment Effect on Fertilizer Adoption

	(1) Fertilizer (kg/SR acre)	(2) Fertilizer (kg/GPS acre)	(3) Fertilizer (=1)	(4) Fertilizer (kg/SR acre)	(5) Fertilizer (kg/GPS acre)	(6) Fertilizer (=1)
	2016	2016	2016	2019	2019	2019
Treatment	17.78***	16.28***	0.49***	0.11	-0.25	-0.01
	(1.82)	(2.99)	(0.05)	(0.69)	(0.64)	(0.01)
Baseline value	0.59	0.85*	0.16	-0.10***	-0.06***	-0.05
	(0.37)	(0.45)	(0.19)	(0.01)	(0.00)	(0.03)
Control mean	0.16	0.18	0.01	0.23	0.27	0.02
(Std. dev.)	(2.57)	(3.31)	(0.1)	(3.14)	(4.05)	(0.12)
N	472	472	589	292	292	301
Village FE	✓	✓	✓	✓	✓	✓

Notes: This table presents ANCOVA results from estimating $Y_{iv} = \alpha_0 + \varphi Y_{i0} + \gamma TREAT_i + d_v + \varepsilon_{iv}$ in which we pool the voucher only group and the voucher plus recommendations group together and exclude the control villages. The reference category is the pooled control group and recommendations only group. Robust standard errors in parentheses and are clustered at the village-level. *** p<0.01, ** p<0.05, * p<0.1. SR stands for self-reported, as answered by respondents in the socioeconomic survey. The control mean refers to the average of the control group at baseline. All regressions control for village fixed effects.

Table (A6) Pooled Treatment Effect on Self-Reported and Satellited-Based Productivity

	(1) Self-Reported (z-score)	(2) Satellite-Based (z-score)	(3) Self-Reported (z-score)	(4) Satellite-Based (z-score)	(5) Self-Reported (z-score)	(6) Satellite-Based (z-score)	(7) Self-Reported (z-score)	(8) Satellite-Based (z-score)
	2016	2016	2019	2019	2016	2016	2019	2019
Treatment (= 1)	0.318**	0.076	0.143*	-0.072	0.327**	0.075	0.138*	-0.078
	(0.128)	(0.060)	(0.079)	(0.090)	(0.127)	(0.058)	(0.074)	(0.085)
Pure stand (=1)					-0.148	-0.055	0.154	0.029
					(0.178)	(0.074)	(0.157)	(0.107)
Plantations (=1)					-0.113	0.146*	-0.078	0.208
					(0.107)	(0.084)	(0.125)	(0.203)
Control mean	0.03	0.00	0.09	-0.02	0.03	0.00	0.09	-0.02
(std. dev.)	(1.07)	(0.99)	1.05)	(1.00)	(1.07)	(0.99)	1.05)	(1.00)
N	491	491	294	294	491	491	294	294
Village FE	✓	✓	✓	✓	✓	✓	✓	✓

Notes: This table presents OLS results from estimating $Y_{iv} = \alpha_1 + \gamma TREAT_i + d_v + \varepsilon_{iv}$ in which we pool the voucher only group and the voucher plus recommendations group together and exclude the control villages. The reference category is the pooled control group and recommendations only group. All yields are measured in standard deviation units to be able to compare self-reported and satellite-based measures that use different units. Robust standard errors in parentheses and are clustered at the village-level. *** p<0.01, ** p<0.05, * p<0.1. The control mean refers to the average of the control group at endline because satellites are unable to predict yields at baseline. V stands for the Voucher treatment; R for the Recommendations treatment; V+R for the Voucher plus Recommendations treatment. All regressions control for village fixed effects.

Table (A7) Pooled Treatment Effect on Self-Reported and GPS-Corrected Productivity

	(1) Self-Reported (kg/acre)	(2) GPS-Corrected (kg/acre)	(3) Self-Reported (kg/acre)	(4) GPS-Corrected (kg/acre)	(5) Self-Reported (kg/acre)	(6) GPS-Corrected (kg/acre)	(7) Self-Reported (kg/acre)	(8) GPS-Corrected (kg/acre)
					>0.5	>0.5	>0.5	>0.5
	2016	2016	2019	2019	2016	2016	2019	2019
Treatment (=1)	86.19**	57.55	85.01**	57.41	28.85	63.43	18.93	38.67
	(34.60)	(41.86)	(35.71)	(45.29)	(25.68)	(43.71)	(24.93)	(42.43)
Baseline Value	0.15*	0.21**	0.15	0.24**	0.15*	0.18***	0.15	0.18**
	(0.09)	(0.08)	(0.09)	(0.09)	(0.08)	(0.06)	(0.09)	(0.07)
Control mean	426.51	395.38	400.26	368.8	427.62	395.73	402.15	362.72
(std. dev.)	(318.6)	(404.16)	(288.56)	(318.06)	(322.98)	(410.88)	(293.54)	(318.48)
N	431	431	409	409	271	271	250	250
Village FE	✓	✓	✓	✓	✓	✓	✓	✓

Notes: This table presents ANCOVA results from estimating $Y_{iv} = \alpha_0 + \varphi Y_{i0} + \sum_{k=1}^{3} \gamma_k TREAT_i^k + d_v + \varepsilon_{iv}$ in which we pool the voucher only group and the voucher plus recommendations group together and exclude the control villages. The reference category is the pooled control group and recommendations only group. Robust standard errors in parentheses and are clustered at the village-level. *** p<0.01, ** p<0.05, * p<0.1. The control mean refers to the average of the control group at baseline. V stands for the Voucher treatment; R for the Recommendations treatment; V+R for the Voucher plus Recommendations treatment. All regressions control for village fixed effects.

Table (A8) Cross-Village Treatment Effect on Fertilizer Adoption

	(1) Fertilizer (kg/SR acre)	(2) Fertilizer (kg/GPS acre)	(3) Fertilizer (=1)	(4) Fertilizer (kg/SR acre)	(5) Fertilizer (kg/GPS acre)	(6) Fertilizer (=1)
	2016	2016	2016	2019	2019	2019
Voucher	8.66***	7.64***	0.30***	0.38	-0.54	-0.00
	(2.73)	(2.75)	(0.07)	(1.43)	(1.40)	(0.02)
Recommendations	-0.15	-0.26	0.04*	-1.28	-1.79	-0.01
	(1.17)	(0.84)	(0.02)	(1.00)	(1.42)	(0.03)
Voucher + Recommendations	27.15***	25.07***	0.71***	-1.48	-1.69	-0.02
	(3.18)	(5.66)	(0.03)	(0.93)	(1.13)	(0.02)
Baseline value	0.70**	0.91**	0.15	-0.09***	-0.05***	-0.06*
	(0.30)	(0.38)	(0.17)	(0.02)	(0.01)	(0.03)
Control mean	0.04	0.02	0.03	0.04	0.02	0.02
(Std. dev.)	(0.38)	(0.2)	(0.18)	(0.42)	(0.23)	(0.12)
V=R (p-value)	0.01	0.01	0	0.13	0.11	0.7
V=R+V (p-value)	0	0.01	0	0.16	0.16	0.21
R=R+V (p-value)	0	0	0	0.61	0.81	0.35
N	638	638	786	406	406	416
Village FE	✓	✓	✓	✓	✓	✓

Notes: This table presents ANCOVA results from estimating $Y_{iv} = \alpha_0 + \varphi Y_{i0} + \sum_{k=1}^{3} \gamma_k TREAT_i^k + d_v + \varepsilon_{iv}$. Robust standard errors in parentheses are clustered at the village-level. *** p<0.01, ** p<0.05, * p<0.1. SR stands for self-reported, as answered by respondents in the socioeconomic survey. The control mean refers to the average of the control group at baseline. V stands for the Voucher treatment; R for the Recommendations treatment; V+R for the Voucher plus Recommendations treatment. All regressions control for village fixed effects.

Table (A9) Cross-Village Treatment Effect on Self-Reported and Satellite-Derived Productivity

	(1) Self-Reported (z-score)	(2) Satellite-Based (z-score)	(3) Self-Reported (z-score)	(4) Satellite-Based (z-score)	(5) Self-Reported (z-score)	(6) Satellite-Based (z-score)	(7) Self-Reported (z-score)	(8) Satellite-Based (z-score)
	2016	2016	2019	2019	2016	2016	2019	2019
Voucher	0.115	0.045	0.111	-0.096	0.115	0.045	0.108	-0.097
	(0.116)	(0.075)	(0.084)	(0.128)	(0.117)	(0.075)	(0.084)	(0.122)
Recommendations	-0.014	-0.057	0.053	-0.022	-0.015	-0.057	0.061	-0.048
	(0.074)	(0.119)	(0.126)	(0.182)	(0.075)	(0.116)	(0.126)	(0.175)
Voucher+Recommendations	0.286**	0.051	0.129	-0.064	0.287**	0.055	0.132	-0.086
	(0.108)	(0.087)	(0.095)	(0.117)	(0.107)	(0.084)	(0.093)	(0.112)
Pure stand(=1)					-0.033	-0.109	0.099	-0.153
					(0.096)	(0.080)	(0.107)	(0.109)
Plantations(=1)					0.012	0.120	-0.022	0.196
					(0.090)	(0.076)	(0.098)	(0.157)
Control mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
(Std. dev.)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
V=R (p-value)	0.15	0.37	0.55	0.68	0.15	0.36	0.63	0.79
V=V+R (p-value)	0.08	0.96	0.79	0.83	0.08	0.93	0.72	0.94
R=V+R (p-value)	0	0.38	0.3	0.8	0	0.35	0.34	0.81
N	659	659	409	409	659	659	409	409
Village FE	✓	✓	✓	✓	✓	✓	✓	✓

Notes: This table presents OLS results from estimating $Y_{iv} = \alpha_1 + \sum_{k=1}^{3} \beta_k TREAT_i^k + d_v + \varepsilon_{iv}$. All yields are measured in standard deviation units using z-scores to be able to compare self-reported and satellite-based measures that use different units. Robust standard errors in parentheses are clustered at the village-level. *** p<0.01, ** p<0.05, * p<0.1. The control mean refers to the average of the control group at baseline. V stands for the Voucher treatment; R for the Recommendations treatment; V+R for the Voucher plus Recommendations treatment. All regressions control for village fixed effects.

Table (A10) Cross-Village Treatment Effect on Self-Reported and GPS-Corrected Productivity

	(1) Self-Reported (kg/acre)	(2) GPS-Corrected (kg/acre)	(3) Self-Reported (kg/acre)	(4) GPS-Corrected (kg/acre)	(5) Self-Reported (kg/acre)	(6) GPS-Corrected (kg/acre)	(7) Self-Reported (kg/acre)	(8) GPS-Corrected (kg/acre)
					>0.5	>0.5	>0.5	>0.5
	2016	2016	2019	2019	2016	2016	2019	2019
Voucher	49.83	29.78	40.30	42.76	44.81	32.56	5.36	11.98
	(45.77)	(44.28)	(37.62)	(44.47)	(48.92)	(46.39)	(34.52)	(39.56)
Recommendations	-22.01	-46.36*	7.20	-42.07	-22.73	-40.38	-35.61	-71.89
	(26.22)	(23.46)	(51.33)	(47.84)	(27.01)	(25.70)	(53.86)	(52.46)
Voucher + Recommendations	100.36**	42.12	25.39	43.59	102.15**	43.04	-8.29	-9.26
	(40.22)	(53.40)	(42.98)	(48.54)	(42.38)	(57.80)	(40.74)	(45.49)
Baseline value	0.14**	0.23***	0.17*	0.19***	0.13**	0.23***	0.11	0.16***
	(0.06)	(0.07)	(0.09)	(0.07)	(0.06)	(0.07)	(0.07)	(0.05)
Control mean	481.13	435.74	456.39	431.86	485.28	427.91	454.14	408.48
(Std. dev.)	(399.58)	(411.28)	(347.3)	(392.38)	(405.52)	(409.08)	(353.55)	(345.4)
V=R (p-value)	0.05	0.05	0.42	0.14	0.08	0.08	0.36	0.18
V=R+V (p-value)	0.18	0.75	0.56	0.97	0.2	0.8	0.63	0.44
R=R+V (p-value)	0	0.08	0.59	0.1	0	0.13	0.49	0.26
N	573	573	370	370	545	545	340	340
Village FE	✓	✓	✓	✓	✓	✓	✓	✓

Notes: This table presents ANCOVA results from estimating $Y_{iv} = \alpha_0 + \varphi Y_{i0} + \sum_{k=1}^{3} \gamma_k TREAT_i^k + d_v + \varepsilon_{iv}$. Robust standard errors in parentheses and are clustered at the village-level. *** p<0.01, ** p<0.05, * p<0.1. The control mean refers to the average of the control group at baseline. V stands for the Voucher treatment; R for the Recommendations treatment; V+R for the Voucher plus Recommendations treatment. All regressions control for village fixed effects.

Other Inputs

Table (A11) Average Treatment Effect on Short-Term Organic Input Use

	(1) Organic fertilizer (=1)	(2) Manure area (acres)	(3) Crop residue area (acres)	(4) Total residue area (acres)
Voucher	0.14***	0.02	0.30***	0.32***
	(0.04)	(0.01)	(0.08)	(0.08)
Recommendations	0.05	-0.01	0.18**	0.17**
	(0.04)	(0.01)	(0.08)	(0.08)
Voucher+Recommendations	0.04	-0.01	0.10	0.09
	(0.04)	(0.01)	(0.06)	(0.07)
Baseline Value	0.02	0.00	0.15	0.14
	(0.06)	(0.00)	(0.09)	(0.09)
Control mean	0.11	0.00	0.21	0.21
(std. dev.)	(0.31)	(0.00)	(0.67)	(0.67)
V=R (p-value)	0.09	0.12	0.22	0.12
V=V+R (p-value)	0.1	0.12	0.08	0.04
R=V+R (p-value)	0.89	0.2	0.41	0.42
N	490	489	489	489
Village FE	✓	✓	✓	✓

Notes: This table shows the treatment impact on organic input application for the 2014 main maize plot in 2016 using ANCOVA by estimating $Y_{iv} = \alpha_0 + \varphi Y_{i0} + \sum_{k=1}^{3} \gamma_k TREAT_i^k + d_v + \varepsilon_{iv}$ and excluding the control villages. Robust standard errors in parentheses and are clustered at the village-level. *** p<0.01, ** p<0.05, * p<0.1. We do not present longer-term results because we did not collect detailed input use data in 2019. The control mean refers to the average of the control group at baseline. V stands for the Voucher treatment; R for the Recommendations treatment; V+R for the Voucher plus Recommendations treatment. All regressions control for village fixed effects.

Table (A12) Average Treatment Effect on Short-Term Agricultural Labor

	(1) Respondent labor/acre	(2) Household labor/acre	(3) Hired labor/acre	(4) Total labor/acre
Voucher	-2.70	-0.03	0.73	-1.87
	(3.10)	(3.20)	(0.55)	(5.57)
Recommendations	2.70	4.10	0.45	4.43
	(4.67)	(4.48)	(0.66)	(8.10)
Voucher+Recommendations	4.62	5.74	0.51	11.27
	(4.67)	(3.68)	(0.46)	(8.25)
Baseline Value	0.11	0.01	0.01	0.05
	(0.07)	(0.05)	(0.02)	(0.05)
Control mean	34.47	19.4	5.73	59.29
(std. dev.)	(33.38)	(23.55)	(26.6)	(59.0)
V=R (p-value)	0.18	0.36	0.61	0.4
V=V+R (p-value)	0.22	0.15	0.5	0.18
R=V+R (p-value)	0.73	0.77	0.91	0.53
N	467	467	478	479
Village FE	✓	✓	✓	✓

Notes: This table shows the treatment impact on labor for the 2014 main maize plot in 2016 using ANCOVA by estimating $Y_{ivt} = \alpha_0 + \varphi Y_{i0} + \sum_{k=1}^{3} \gamma_k TREAT_i^k + d_v + \varepsilon_{ivt}$ and excluding the control villages. Robust standard errors in parentheses and are clustered at the village-level. Self-reported acreage is used in this table. We do not present longer-term results because we did not collect detailed input use data in 2019. The control mean refers to the average of the control group at baseline. V stands for the Voucher treatment; R for the Recommendations treatment; V+R for the Voucher plus Recommendations treatment. All measures are winsorized at the top 1% of values and all regressions control for village fixed effects.