A panel analysis of the impact of KickStart irrigation pumps in Kenya

Jeremy Shapiro, Busara Centre for Behavioral Economics

Johannes Haushofer, Busara Centre for Behavioral Economics

Beatrice Sakwa, KickStart International

Julian Dyer, University of Toronto

Grantee Final Report

Accepted by 3ie: June 2017

International Initiative for Impact Evaluation

Note to readers

This impact evaluation has been submitted in partial fulfilment of the requirements of grant OW1.45 issued under Open Window 1. 3ie is making this final report version available to public as it was received from the authors. All content is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or its board of commissioners. Any errors and omissions are the sole responsibility of the authors. All affiliations of the authors listed in the title page are those that were in effect at the time the report was accepted. Any comments or queries should be directed to the corresponding author, Jeremy Shapiro at jeremypshapiro@gmail.com.

Suggested citation: Shapiro, J, Haushofer, J, Sakwa, B and Dyer, J, 2017, *A panel analysis of the impact of KickStart irrigation pumps in Kenya, 3ie Grantee Final Report.* New Delhi: International Initiative for Impact Evaluation (3ie)

Funding for this impact evaluation was provided by 3ie's donors, which include UK aid, the Bill & Melinda Gates Foundation and the Hewlett Foundation. A complete listing of all of 3ie's donors is available on the 3ie website.

A Panel Analysis of the Impact of KickStart Irrigation Pumps in Kenya*

November 21, 2016

Abstract

This report details the findings of an impact evaluation assessing the effect of Kickstart International's irrigation pumps on farmer welfare in Kenya. The evaluation consists of a panel analysis of small scale farming households across 35 districts that bought a pump in 2009, 2011 and 2015. The pumps were mas-produced and marketed by KickStart and sold through local agri-vet shops across Kenya during all the years covered by the study. Impacts are measured as the difference at endline (2015) in outcomes between cohort 3 (purchasing pumps in 2015) and cohorts 1 and 2 (purchasing pumps in 2009 and 2011, respectively); hence we exploit the time variation in having access to the pump. The analysis provides qualitative evidence that the pumps allowed farmers to increase their incomes, however these impacts are not substantiated in quantitative analysis. It is therefore not possible to conclude from this analysis whether the pumps do or do not have positive impacts overall. The panel approach, however, suffers from several limitations, most notably pre-existing differences across cohorts. An upcoming randomized control trail will likely provide conclusive answers regarding the impacts of these irrigation pumps.

Keywords: irrigation pumps, smallholder farmers, impact evaluation.

* **Acknowledgments**: this evaluation is based on data collected by IFPRI and the Busara Center for Behavioral Economics. Financial support for the evaluation was provided by 3ie in a grant to KickStart International.

Contents

1 Executive Summary

KickStart's mission is to lift millions of people out of poverty quickly, cost-effectively and sustainably. In pursuit of this mission, KickStart has developed and distributed several models of human-powered irrigation pumps since 1998. KickStart's flagship products are the MoneyMaker Max and MoneyMaker Hip pumps. This study aims to reliably establish the impact of the pump on the welfare of small-scale farmer households in Kenya. While intuitive direct effects such as changes in land management practices, assets and income are the focus, secondary effects such as food security and consumption, education, health and time allocation are also considered. As a large number of small-scale farmers tend to be women, the study further sought to understand the pump's ability to shift meaningful dimensions of women's empowerment in terms of intra-household decision-making, female psychological well-being and levels of intimate partner violence.

This study consists of a panel analysis of small scale farming households across 35 districts that bought a pump in 2009, 2011 and 2015. Impacts are measured as the difference in outcomes at endline (2015) between cohort 3 (purchasing pumps in 2015) and cohorts 1 and 2 (purchasing pumps in 2009 and 2011, respectively); hence we exploit the time variation in having access to the pump. In addition propensity score matching is used to control for differences across cohorts not attributable to pump purchase.

Overall, it is not possible to come to strong conclusions. Qualitative evidence suggests that the pumps allowed farmers to increase their incomes, however these impacts are not substantiated in quantitative analysis. It is therefore not possible to conclude from this analysis whether the pumps do or do not have positive impacts overall. The panel approach, however, suffers from several limitations - an upcoming randomized control trail will likely provide conclusive answers regarding the impacts of these irrigation pumps.

2 Introduction

An estimated 70 percent of the world's poor are reliant on agriculture (World Bank). Consequently, improving agricultural productivity is a promising poverty reduction strategy. In this context, KickStart's mission is to lift millions of people out of poverty quickly, cost-effectively and sustainably. In pursuit of this mission, KickStart designs, promotes and mass markets simple moneymaking tools. In particular, KickStart has developed and distributed several models of human-powered irrigation pumps since 1998. KickStart's flagship products are the MoneyMaker Max and MoneyMaker Hip pumps, which allow farmers to pull and pressurize water, irrigate up to two acres of land and grow crops year-round.

While KickStart has conducted internal evaluations of the impact of their pumps from the beginning, KickStart sought to complete an external independent evaluation. To that end, IFPRI collected baseline and midline data from small-scale farming households across 35 districts, who had purchased the pump in 2009 and 2011 respectively. Subsequently, the Busara Center for Behavioral Economics was contracted to collect endline data for the prior cohorts (purchasing pumps in 2009 and 2011) and a third cohort that purchased a pump in 2015.

This study aims to reliably establish the impact of the pump on the welfare of small-scale farmer households in Kenya using the panel dataset described above. While intuitive direct effects such as changes in agricultural output, assets and income are the focus, secondary effects such as food security and consumption, education, health and time allocation are also considered. As a large number of small-scale farmers tend to be women, the study further seeks to understand the pump's ability to shift meaningful dimensions of women's empowerment in terms of intra-household decision-making, female psychological wellbeing and levels of intimate partner violence. The long-term purpose of the evaluation is to contribute to knowledge about the role of small-scale irrigation technology in rural development.

The primary questions addressed in this study are: (i) What is the overall effect of owning a pump vs. not owning a pump on various dimensions of household welfare? (ii) What are the long-term effects of owning a pump on household welfare?

To address these questions, this study compares three cohorts of farmers each that have purchased the pump at different time points – in 2009, 2011 and 2015. Allocation to each cohort is not random, therefore this study rests on the assumption that the farmers across all three cohorts are similar to one another. Moreover this study is based on the assumption that farmers who recently purchased the pump (in 2015) have not benefited significantly from the pump due to limited time during which they owned a pump. The data does not robustly support these two assumptions, which contributes to the inability to draw strong conclusions.

If both assumptions hold, farmers who bought the pump in 2015 serve as a quasi-control group in this design. Question (i) can then be answered by comparing cohort 3 to cohorts 1 and 2, and question (ii) is answered by comparing cohort 1 to cohort 2. This study employs direct comparisons, and also uses propensity score matching to control for non-random assignment to cohorts. This analytical approach was described in a pre-registered analysis plan. [1]

The following sections discuss the evaluation design and analytical approach before proceeding to a discussion of the results.

2.1 Literature Review

Sub-Saharan Africa currently lags behinds other regions in irrigation coverage, with only 7.3 million hectares of the 40 million suitable hectares currently irrigated. Smallholder agriculture in Africa is characterized by low productivity, and reliance on rain-fed agriculture, despite the many potential benefits of irrigation. Reliance on rain-fed agriculture limits production to short rainy seasons which causes strong seasonality in prices, leading to price spikes during the dry season, and price slumps during the harvest season as markets are flooded with output. Irrigation weakens this seasonality as farmers are no longer constrained by the rainy seasons. It also allows farmers to grow higher-value crops that require larger amounts of water (Burney, Naylor, and Postel 2013). An analysis of case studies of different irrigation systems in Sub-Saharan Africa found that distributed irrigation systems (comprised of many small-scale irrigation systems) that covered a broad area were more effective than large-scale centralized irrigation (B. Van Koppen, R. Namara, and Safilios-Rothschild 2005). There are a number of other benefits to small-scale irrigation. Implementing small-scale irrigation does not require huge public investment, subject to corruption or bureaucratic inertia (Burney, Naylor, and Postel 2013). In a study focusing specifically on individual small-scale irrigation systems, de Fraiture and Giordano find that privately purchased irrigation systems are highly profitable compared to underperforming public irrigation systems (Fraiture and Giordano 2014). A study in Tigray, a drought-prone region of Northern Ethiopia, employing a propensity matching approach compares households participating in small-scale irrigation schemes with non-participants, and finds significant positive effects on income, overall expenditure, asset accumulation and expenditure on agricultural inputs (Zeweld et al. 2015). However, another study of irrigation in Tigray finds that, while irrigated farms have higher productivity than rain-fed, there is significant scope for improving the productivity of irrigated plots using complementary input technologies (Gebregziabher, R. E. Namara, and Holden 2012).

A recent study developed a model combining GIS data, biophysical and economic

[1] https://www.socialscienceregistry.org/trials/1002

predictions and crop optimization techniques and found that there is significant potential for the expansion of small scale irrigation in Sub-Saharan Africa (Xie et al. 2014). The potential for expansion in smallholder irrigation is further motivation for understanding the effects on farmer livelihoods.

A study of small-scale private irrigation technology in Ghana and Zambia finds a number of interesting gendered dimensions to irrigation use. Female headed households tend to have lower adoption rates than male headed households, and that men generally provide most of the irrigation labour. They do find, however, that female decision making is stronger for irrigated plots than rainfed (Barbara Van Koppen, Hope, and Colenbrander 2012).

A pilot study for Shamba Maisha tested the feasibility of a combined agricultural and microfinance intervention to address the joint challenges of food insecurity and HIV/AIDS. Despite some difficulty in implementing the microfinance arm of the intervention, there was strong willingness to join and qualitative evidence suggests that agricultural and financial training were beneficial and should be included in any combined intervention package. This pilot study will be scaled up in a larger cluster randomized trial to rigorously test the effect of a joint intervention. (Cohen et al. 2015)

3 Intervention, theory of change and research hypotheses

This study evaluates the impact of manually powered irrigation technology on a range of outcomes for small-scale farmers. Importantly, the farmers in this study self-selected into purchasing the pumps. The MoneyMaker pumps were mas-produced and marketed by KickStart and sold through local agri-vet shops across Kenya during all the years covered by the study. The MoneyMaker pumps were locally available for sale to anyone who had (or could otherwise access) the money to buy them. They were locally promoted over the years through live demonstrations by KickStart staff at the retail shops and local farmer field days, and occasionally through local-language radio advertisements.â Therefore the results are interpretable as the impact of pump purchase, which is related to but not the same as the impact of pumps provided through alternative means (e.g., free distribution).

Irrigation enables farmers to grow multiple cycles of crops, produce larger yields and, most importantly, harvest and sell their crops in the off seasons when prices are high, ensuring year-round incomes. Therefore these farmers would increase their food security and incomes, leading to improved education opportunities for children, improved healthcare, investments in new productive assets, investments in new farming businesses, empowerment of women, and overall life improvements.

It is likely that farmers who acquire a pump will change their farming practices in response to the acquisition of new technology. This study investigates impacts on investment in farming inputs and time spent in agriculture.

Any increase in farming productivity or inputs deriving from the pump should be reflected in increased output and changes in the farmers' economic situation. This study investigates changes in revenue, assets, consumption and food security.

Finally, based on other research (Haushofer and Shapiro 2013) it is possible that changes in a farmer's economic situation may lead to improved mental health, female empowerment and a decline in domestic violence. The latter effects are primarily hypothesized to be driven by an improved economic situation as much intrahousheold conflict it driven by economic scarcity.

These outcomes consititute the primary variables considered in this analysis.

4 Context and timeline

As noted above, the panel dataset was conducted across 35 districts in Kenya. Farmers who had purchased a KickStart pump were eligible for inclusion in the study. This sample provides a robustly representative sample of farmers adopting the intervention under study.

Also as noted above, the panel dataset was collected in three waves. The baseline survey (covering Cohort 1) was conducted by IFPRI in 2009. This sample includes farmers who had recently purchased a pump. The midline survey, also conducted by IFPRI in 2011, included a follow up survey for Cohort 1, as well as Cohort 2, who had recently purchased pumps. The endline data was conducted by Busara in 2015. The endline includes a follow up with a sub-set of Cohorts 1 and 2 as well as a survey of Cohort 3, which had recently purchased a pump in 2015.

5 Evaluation Design

5.1 Sampling and Identification Strategy

This study follows three cohorts of farmers who have purchased Moneymaker pumps. Cohort 1 is made up of farmers who purchased pumps in 2009, Cohort 2 purchased pumps in 2011 and Cohort 3 purchased pumps in 2014/15. KickStart and IFPRI, who initiated the impact evaluation, interviewed a total of 1230 farmers across 35 Kenyan districts: 585 in 2009 (Cohort 1) and an additional 645 in 2011 (Cohort 2). The endline sample includes a subset of the total Cohort 1 and Cohort 2 sample as well as an additional cohort - Cohort 3 households. A random sample out of the existing dataset, stratified by district, was drawn to decide on which Cohort 1 and 2 households to follow up with. Cohort 3 households were randomly selected from incoming sales data from KickStart on a monthly basis. To qualify for Cohort 3 inclusion from the sales list, a household had to have bought a pump no longer than 6 months before the interview. Cohort 3 sample size per district was determined in proportion to the combined Cohort 1 and 2 sample in each location. To control for seasonal effects, the interviews of all three cohorts in each district were completed within two weeks.

The identification strategy is to compare cohort 3 (which has had limited time to benefit from the pump) to cohorts 1 and 2 (which have had the pump for some time). This study rests on the assumption that the farmers across all three cohorts are similar to one another. The data, however, do not support this assumption (see Table 3), which makes it difficult to draw strong conclusions from this analysis. In particular, farmers may self-select into purchasing pumps earlier than other farmers, which would make the different cohorts different in terms of observable and unobservable characteristics. This study attempts to adjust for this with propensity score matching. But this approach is also limited by the limited number of observable characteristics captured in the data and by the fact that there is no distinct baseline and outcome data for cohort 3.

5.2 Data Collection Methods and Instruments

During the 2015 endline, trained interviewers visited each the households. Both the primary male and the primary female of the household were interviewed (separately). Surveys were administered on tablets using the SurveyCTO survey software. To ensure data quality, backchecks, with a focus on non-changing information, were conducted on 10% of all interviews. This procedure was known to field officers ex ante. The instrument, which was based on the baseline and midline instrument and updated to account for additional hypotheses noted above, is included as an appendix.

5.3 Risk and Treatment of Attrition

Attrition was a concern for Cohorts 1 and 2, since some of the tracking information, especially for Cohort 1, was not accurate and did not involve sufficient detail. Three approaches were used to control for attrition. First, the data collection team used all pieces of information available to find respondents for follow-up surveys: phone numbers, GPS co-ordinates and the farmer's name. The first method was to contact farmers a week prior to the field team visiting a region to confirm their location and arrange a convenient time for the survey to be administered. The standard protocol in case of difficulty contacting farmers was to call three times a day for the week prior to the field team visiting. In the event that it was impossible to make contact by phone GPS coordinates were used to locate the respondent's residence. When using both methods, phone or GPS, the field team confirmed this information by asking for the respondent by name in the nearest village. Farmers who were traceable but not available to participate in the survey at the time the field team was visiting their district were revisited at another time. Secondly, survey completion was incentivized through a small appreciation gift (spare pump parts and 2 kg of maize flour).

See Table 0.A in in Appendix A.1 for a detailed breakdown on sample size and attrition level, by cohort. The p-value for a difference attrition for treatment and control groups is 0.158, meaning there is no statistically significant difference in attrition by treatment status. See Table 0.B in the Tables section for the regression table.

6 Evaluation Methods and Implementation

6.1 Basic Specification

For the cohort study, the specifications used to identify differences between the cohort groups are

$$y_{\text{iv}} = \beta_0 + \alpha_d + \beta_1 C12 + \epsilon_{iv} \quad (1)$$

$$y_{\text{iv}} = \beta_0 + \alpha_d + \beta_1 C1 + \epsilon_{iv} \quad (2)$$

where y is the outcome of interest, *C1* is a dummy for whether a household belongs to Cohort 1, and *C12* a pooled dummy for the two older cohorts. α_d represent district fixed effects.

Equation 2 is restricted to Cohorts 1 and 2 test for a difference between long-term and short-term effects, where β_1 is the coefficient of interest. As Cohort 2 is the excluded category in this specification, β_1 tests for differences in effect between those who owned the pumps the longest, Cohort 1, and those who owned the pumps for an intermediate time, Cohort 2.

6.2 Accounting for Outliers in Data

To account for the possibility that a small number of respondents will have extreme values that will be large enough to skew our results, several variables are *top-coded*, meaning that if a respondent's value for a variable is in the top 5%, it is replaced with the value for the 95th percentile. This ensures that the results and cohort means presented in the following analysis are representative of the sample as a whole, and not driven by a small number of observations. Extreme values were noted and top-coded for the following variables before constructing the outcome variables (described in detail in below in Outcome Variables) or the propensity scores.

- Productive Assets, Vehicles, Household Durables, Livestock, Savings and Land & Buildings
- Total physical, emotional and sexual violence as well as domestic violence against children
- Land inputs on fertilizer, seeds & planting materials, irrigation water

- Average revenue from enterprises, monthly agricultural revenue and wage labour revenue

As a robustness check estimates which are not top-coded and estimates where outliers are dropped instead of top-coded are included in the Supplementary Tables Appendix (Tables 7.1 and 7.2).

6.3 Accounting for Multiple Inference

As household water pumps are likely to impact a large number of economic behaviors and dimensions of welfare, and given that the survey instrument often included several questions related to a single behavior or dimension, multiple inference with a large number of outcomes is a concern (see Romano and Wolf 2005). The study accounts for multiple hypotheses by using outcome variable indices and family-wise p-value adjustment.

Primary outcomes of interest, including indices and variable groups, were pre-specified and are discussed in the appendix. This study reports both unadjusted p-values as well as p-values corrected for multiple comparisons using the Family-Wise Error Rate for index variables and reports only unadjusted p-values for individual components of each index.

6.3.1 Construction of Indices

To keep the number of outcome variables low, allowing for greater statistical power even after adjusting p-values to control for multiple inference, indices for several groups of outcome variables were constructed. The procedure outlined in Anderson (2008) was used to construct these indeces. First, for each outcome variable y_{jk}, where j indexes the outcome group and k indexes variables within outcome groups, the variables is recoded such that high values correspond to positive outcomes. The covariance matrix Σ_j for outcomes in outcome group j then consists of elements:

$$\hat{\Sigma}_{jmn} = \sum_{i=1}^{N_{jmn}} \frac{y_{ijm} - \bar{y}_{jm}}{\sigma^y_{jm}} \frac{y_{ijn} - \bar{y}_{jn}}{\sigma^y_{jn}} \tag{3}$$

Here, N_{jmn} is the number of non-missing observations for outcomes m and n in group j, $\bar{y}_{jm}$ and $\bar{y}_{jn}$ are the means for outcomes m and n in outcome group j,and $\sigma^y_j m$ and $\sigma^j_j n$ are the standard deviations in the control group for the same outcomes.

The covariance matrix is then inverted, defining weights w_{jk} for each outcome k in outcome group j by summing the entries in the row of the inverted covariance matrix corresponding to that outcome:

$$\hat{\mathbf{\Sigma}}_j^{-1} = \begin{bmatrix} c_{j11} & c_{j12} & \cdots & c_{j1K} \\ c_{j21} & c_{j22} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots \\ c_{jK1} & \vdots & \ddots & c_{jKK} \end{bmatrix} \tag{4}$$

$$w_{jk} = \sum_{l=1}^{K_j} c_{jkl} \tag{5}$$

Here, K_j is the total number of outcome variables in outcome group j. Finally, each of the outcome variables is transformed by subtracting its mean and dividing by the control group standard deviation, and then weighting it with the weights obtained using the method above. The resulting transformed variable is denoted $\hat{y}_{ij}$, which yields a generalized least squares estimator (Anderson 2008).

$$\hat{y}_{ij} = \left(\sum_{k \in \mathbb{K}_{ij}} \right)^{-1} \sum_{k \in \mathbb{K}_{ij}} w_{jk} \frac{y_{ijk} - \bar{y}_{jk}}{\sigma^y_{jk}} \tag{6}$$

Here, $\mathbb{K}_{ij}$ denotes the set of non-missing outcomes for observation i in outcome group j.

6.3.2 Family-Wise Error Rate

Because combining individual outcome variables in indices as described above still leaves multiple outcome variables (viz. separate index variables for health, education, etc.), p-values for coefficients of interest are adjusted for multiple statistical inference. These coefficients are those on the treatment dummies in the basic specifications, or those on the dummies for individual treatment arms. The procedure for this adjustment, from Anderson 2008, is as follows:

1. Compute naïve p-values for all index variables $\hat{y}_j$ of j main outcome groups, and sort these in order of decreasing significance, i.e in order of increasing p-values such that $p_1 < p_2 < \cdots < p_J$.

2. Follow Anderson's (2008) variant of Efron & Tibshirani's (1993) non-parametric permutation test for each of the indices representing main outcome groups.

This permutation test is used in place of the standard t test, allowing us to calculate p-values that do not rely on assumptions about the distribution of the test statistic. This resampling involves random draws of treatment assignment, in order to sample the data under the assumption of no treatment effect. Then estimate the simulated p-value for difference in means for treatment and control.

3. Impose the original monotonicity (from ordering in step 1) in the resulting vector of p-values $[p_1^*, p_2^*, \cdots, p_J^*]'$ by computing $p_r^{**} = \min\{p_r^*, p_{r+1}^*, \cdots, p_J^*\}$, where r is the position of the outcome in the vector of naïve p-values.

4. Repeat steps 2-4 of the procedure 100,000 times and compute the fraction of iterations where the simulated p-value is lower than the observed p-value and define this as the non-parametric p-value, p_r^{fwer*}.

5. Enforce monotonicity again: $p_r^{fwer} = \min\{p_r^{fwer*}, p_{r+1}^{fwer*}, \cdots, p_J^{fwer*}\}$, to ensure that the largest unadjusted p-values correspond to the largest adjusted p-values.

This yields the final vector of family-wise error-rate corrected p-values. This study reports both these p-values and the naïve p-values. Within outcome groups, this study reports naïve p-values for individual outcome variables other than the indices.

6.4 Propensity Matching

The regression results from the specifications above are supplimented with a propensity matching approach. The intuition behind propensity matching is to match treated individuals (i.e., those in the earlier cohorts) with the most similar non-treated individual and compare their outcomes. Propensity matching relies on the assumption that – conditional on observed characteristics – treatment assignment (being in Cohort 1 or 2 for Equation 1, and being in Cohort 1 for Equation 2) is independent of potential outcomes, and the comparison of treated and untreated groups can be interpreted as the treatment effect with no selection bias. The potential outcomes are also then independent of treatment status conditional on the propensity score $p(X_i)$, where the propensity score is the probability of being assigned to treatment conditional on observables: $p(X_i) = E[T_i|X_i]$.

To implement this, treated individuals are matched with non-treated individuals based on their propensity scores. Propensity scores measure the predicted likelihood of being in the treatment group, given observable baseline characteristics.

The estimate of interest is the Average Treatment Effect (ATE):

$$\tau = \mathbb{E}\left[Y(1) - Y(0)\right] \tag{7}$$

Under the assumption made above:

$$\tau = \mathbb{E}\left[\bar{\mu}\left(1, p(X)\right) - \bar{\mu}\left(0, p(X)\right)\right] \tag{8}$$

where $\bar{\mu}\left(1, p(X)\right)$ is the conditional mean of Y given treatment status and $p(X) = p$. Therefore, it is possible to estimate the ATE by looking at groups with similar estimated likelihood of treatment $p(X)$ and comparing conditional means of individuals those who were treated and those who were not. See Abadie and Imbens 2015 for details.

The following variables are used to create propensity scores:

- Time invariant covariates from endline data
- Baseline assets (IFPRI)
- Baseline irrigated land area (IFPRI)
- Baseline rainfed land area (IFPRI)

Note that the structure of the data (a single observation for cohort 3 vs. multiple observations for cohort 1 ans 2) makes it impossible to conduct certain checks on the propensity score method. Details of propensity score matching are discussed below.

7 Impact analysis and results of the evaluation questions

7.1 Main Specification Results

Table 1.A presents the results from estimating Equation 1 for the main outcome variables to test for treatment effects. This table shows results after adjusting for outliers by topcoding at the 95th percentile, where any values of non-categorical outcome variables that were above the 95th percentile were replaced with the value at the 95th percentile. The tests of significance that adjust for multiple inference (see the FWER p-values in the second from bottom row) find that none of the

estimates are significantly different from zero once adjustment for multiple inference is made.

Table 1.B shows the results from estimating Equation 2 on the main outcome variables, to test for differences between long-term and short-term treatment effects. After adjusting for multiple testing, for none of the estimates is it possible to reject the hypothesis that the treatment effect is the same in the long- and short-term. This is shown by the FWER p-values reported in the second from bottom row.

The tables discussed above capture impacts of having purchased a pump, regardless of use. This is distinct form the impact of having purchased and used a pump. Both quantities are of interest but are likely distinct as only between 50 and 75 percent of farmers actually report using the pump. An instrumental variables two stage least squares (2SLS) approach is used to test for impacts of actually using the pump while controlling for selection bias. The first stage is to predict whether a farmer uses a pump based on which cohort they belong to (which this analysis assumes is as good as random). Then outcomes are related to actual use of the pump as predicted by cohort status in the second stage. The first stage, shown in Table 1.C, shows that cohort status is a robust predictor of pump usage. As would be expected since less than 100 percent of farmers use the pumps the coefficients in the 2SLS estimates are larger, as seen in Table 1.D. However, the results are not robustly statistically significant and is is not possible to conclusively say if using the pump has positive or null impacts on pump users.

Tables 2.1-9 shows which components of the main outcome variables are driving the overall estimates by estimating Equation 1 on each component individually. These results report p-values that are unadjusted for multiple testing, as this is an exploratory analysis and not a main result. Consistent with the lack of observed impacts on indices, there is no indication of significant differences when assessing components of the indices.

When considering the robustness of the main results to not top-coding or simply dropping outliers, the monetary outcomes shift somewhat in these specification (see Appendix B). For example the assets coefficient becomes highly negative and statistically significant according to unadjusted p-values without top-coding and land inputs becomes positive and statistically significant according to unadjusted p-values when dropping outliers. The other point estimates are not substantially changed and these checks do not suggest markedly different conclusions than those indicated by the main effects table. Moreover, it is not wise to place emphasis on results driven by select observations.

7.2 Propensity Score Results

One limitation of the analysis above (comparing simple endline means across cohorts) is that it does not account for potential differences in the farmers included in each cohort. While selection bias is, to some extent, mitigated by the fact that all farmers in the sample self-selected into pump purchase, there may be differences in the type of farmer who is an early adopter (e.g., purchasing a pump in 2009) vs. a late adopter (purchasing a pump in 2015). This can be seen in Table 3, which shows the mean of various farmer characteristics across cohorts, measured at the respective "baseline" for each cohort. There are, for example, differences in assets across the cohorts. Propensity score matching is used to control for such differences to the extent possible.

Table 4.1 restricts to Cohorts 1 and 2 and presents estimates of a logit regression of treatment status (being in Cohort 1) on various baseline characteristics.[2] The invariant characteristics come from the endline data. Time-varying baseline characteristics come from the baseline survey for Cohort 1 and from the midline survey for Cohort 2. The table shows the Average Marginal Effects of baseline characteristics on the probability of being in Cohort 1. Recall that the Asset variables are in KSH when interpreting the coefficients. The tables shows that productive assets, household durables, vehicles and total rainfed acres are all significant predictors of probability of treatment. These variables, therefore, drive the matching used for propensity score estimation. Figures 1 and 2 show the distribution of propensity scores for the treated and non-treated groups. These histograms show good, if not complete, overlap in propensity scores between treatment and control. This means that there are enough households across groups with similar treatment probability based on baseline and time-invariant characteristics that they can reasonably be compared.

Table 4.2 shows the estimated differences between Cohort 1 and Cohort 2 for primary outcome variables when making the propensity score adjustment. The table does not reveal any significant differences across cohorts when making this adjustment.

Table 5.1 presents estimates of a logit regression of treatment status (being in Cohort 1 or 2) on various baseline characteristics when using data from all three cohorts. In this specification, it is not possible to use assets in the propensity score

[2] A *logit* regression model is used when the dependent variable is categorical: in the case it is treatment status, which can only take the values 1 and 0.

matching. The reason is that it is not valid to include assets both as an outcome as well as a variable for matching.[3] The table shows the Average Marginal Effects of baseline characteristics on the probability of being in the treatment group (being in Cohort 1 or 2). Baseline irrigated acres and total rainfed acres are significant predictors of probability of treatment. These variables, therefore, drive the matching used for propensity score estimation. The sample sizes in this section are somewhat smaller than the earlier tables, as the baseline variables are missing for some observations.

Figures 3 and 4 show the distributions of propensity score for the treated and non-treated cohorts. These plots show that the distribution of propensity scores for non-treated cohorts overlaps with that for the treated cohort to a large extent, meaning that members of the treated cohort can be matched to a member of the non-treated cohort with a similar likelihood of being treated based on baseline characteristics. However, it is impossible to test whether the propensity score method corrects baseline characteristics since there is no baseline data for Cohort 3.

Table 5.2 presents the propensity score matching estimates of the Average Treatment Effect of being in Cohort 1 or 2 compared to Cohort 3. There is a positive and weakly significant (10 percent level) estimate for the KSH value of land inputs used. However, this is to be expected due to chance when assessing impacts across 10 outcomes. When correcting for multiple testing, this results is no longer significant, as shown by the FWER p-value of 0.46. Is is therefore not possible to conclude whether there were or were not significant effects from this analysis.

In Appendix section B.2 we provide a number of robustness checks to our use

[3]To see this, assume that farmers receive random, identically distributed productivity/economic shocks in each period, and that these productivity shocks are absorbed by asset holdings. If we then match based on baseline asset holdings, those who received highly negative shocks in the baseline period will be matched to others with low asset holdings at baseline. This is not a problem when we observe a new round of data at endline for treatment and control, because the expected value of the random shock is the same for all individuals, so we introduce no bias. However, we observe only one round of data for Cohort 3 so the baseline data and endline data are the exact same so, by construction, those who received very negative random shocks at baseline have extremely negative shocks in the endline data. This means that, even with no treatment effect, if we were to match based on baseline values and compare endline outcomes for Cohorts 1 and 2 to Cohort 3, we would expect that those in Cohorts 1 and 2 who received negative shocks at baseline received less extreme shocks at endline which would appear as a positive treatment effect. Another potential issue we face when using baseline data is a potential secular time trend. If asset values increase over time for all farmers in our sample independently of the Kickstart treatment (what we refer to as a "secular" time trend) then if farmer A has the same value of assets today that farmer B had three years ago, then we would expect farmer B to have higher value of assets today, even with no treatment effect. This is a potential issue because our baseline data come from different time periods. If there is a secular time trend in baseline characteristics then individuals with similar baseline values, but measured at different times, may be systematically different.

of propensity score methods to estimate treatment effects. We restrict to the most conservative model for treatment status, using only baseline total acres farmed, and show that the estimated propensity scores show good overlap across treatment status. Using this specification for propensity scores, Table 8.2 shows that we estimate no statistically significant treatment effects.

In Table 8.3 we show that this robust propensity score specification achieves good balance in covariates, with small standardized differences in a number of demographics across treatment and control after matching. Despite this propensity score matching could fail if it does not create balance on time-varying outcome variables - we are unable to test this since we lack a baseline for cohort 3.

In Tables 8.4 to 8.7 we replicate the main propensity score specifications in the earlier main tables section, using a doubly robust treatment effects estimator (the Inverse-Probability- Weighted Regression Adjustment (IPWRA) estimator) and find weak patterns of treatment effects. We find a statistically significant reduction in domestic violence (with no FWER adjustment) when estimating treatment effects across all cohorts when using the full treatment model, but not the robust propensity score specification. When comparing long vs. short run effects (Cohort 1 v. Cohort 2) we find a statistically significant decrease in time spent in agriculture and a statistically significant decrease in psychological wellbeing. When estimating the long v. short run treatment effect using the robust specification we again find the significant, negative estimate for time spent in agriculture, but also find a statistically significant negative effect on the asset index and children's food security. This inconsistent pattern makes it difficult to draw strong conclusions from this analysis, in line with what has been found in the other analysis here conducted.

7.3 Qualitative Analysis

The endline survey also included a number of measures to supplement quantitative results with qualitative insights. These questions were designed as "quasi-open questions", where subjects were asked an open-ended question, but given the option to select multiple responses from a list of many, or describe a response other than those included in the list provided to them.

The three quasi-open questions asked during the endline are the following:

I. How has your main pump changed your investments and lifestyle?

II. Please list the most important problems that you encounter in using the Moneymaker pump.

III. What are your biggest worries about next year?

Question I: *How has your main pump changed your investments and lifestyle?*

Tables 6.1A - C show responses to the question "How has your main pump changed your investments and lifestyle?" for each cohort. Across cohorts, the primary way that owning a pump impacts subjects' investments and lifestyle is by enabling expansion of crop area, followed by an increase in overall financial savings and sending their children to a better school. These three effects of pump ownership were cited at a much higher rate than any of the other options listed in the survey, which suggests that pump use may have enabled farmers to generate excess income. Notable is the lack of differences across cohorts, which undermines the assumption that cohort 3 had not yet benefited from the pump upon which the quantitative analysis rests. This may contribute to the lack of quantitative results.

The analysis shows that the total range of sources for changes in lifestyle and investments experienced by the study participants was captured only partially by the list of responses available in the choices available to them. Many responses included separate causes for investment and lifestyle changes in the "Other" option. The central theme that emerged from these open-ended responses was the food security impacts of using the pump, with a large number of participants stating that acquiring a pump enabled their families to grow excess food for subsistence, increasing food security particularly during dry seasons. As well, many participants elaborated on the pump having greatly eased access to water generally, and particularly through permitting them to easily draw water from deep wells. One respondent mentioned that with the pump she is now able to save time to fetch water, and is able to get quality clean water from a borehole, instead of the river, which is contaminated. Relatedly, many participants explained that the pump enabled them to feed water to their livestock with more ease.

Question II: *Please list the most important problems that you encounter in using the Moneymaker pump.*

Table 6.2 presents the results for the responses given when participants were asked to list the most important problems they have encountered using the pump. When asked to select the most important issues when using the pump, 13.1% of responses cited having never encountered any problems. This is encouraging, signaling that many participants from varying cohorts, having acquired the pumps at different time intervals, enjoy continued satisfaction with their pump. Nevertheless, a majority of participants were able to list at least one major issue they have experienced using the pump. The most-frequently listed reason was that the pump requires two people to operate it, comprising of 13.1% percent of all responses. This was closely followed by 9.9% of responses specifying that the pump was generally "hard to operate". Finally, 9.5% of responses indicated that the rubber caps for the pumps wore out too quickly, and 6.2% that they suffered from a lack of access to spare parts.

Question III: *What are your biggest worries about next year?*

Finally, respondents were asked to select what their most significant worries were for the coming year. This question was asked as a proxy for study participants' psychological health. Table 6.3 presents a breakdown of the responses. With 24.7% of overall responses, school fees/tuition emerged as the most frequently cited concern across study participants. This was followed by season-related productivity concerns, with 19.5% of responses indicating worries around rainfall yield, and 11.1% of responses citing fear of drought as a major concern for the upcoming year.

8 Discussion and implications for policy and practice

Overall, it is not possible to reach conclusions based on the analysis in this study. Neither the basic treatment effect specifications or propensity score matching results indicate impacts after taking into account multiple testing adjustments. The assumptions required for this analysis to produce valid estimates of impact, however, were not met. For one thing, there were clear differences across the 3 cohorts, suggesting they are not directly comparable. Further, qualitative analysis suggests that all three cohorts in the study benefited from pump ownership. This poses a challenge for the design of the quantitative analysis, which relies on Cohort 3 being a valid control group who have not yet experienced the benefit of the pumps. If, as the qualitative evidence suggests, Cohort 3 farmers benefited significantly from their Kickstart pump, it is not possible to consier them as a control group. Therefore these results do not conclusively indicate either that the irrigation pumps do or do not have positive impacts overall.

There is a qualitative indication that the pumps improve the economic situation of those that purchase and use them. It will be necessary to reconcile the positive qualitative effects with the lack of observed quantitative effects by testing the impact of KickStart's pumps with a randomized control trial KickStart will launch in the future. The RCT, which does not suffer from the same limitations as panel analysis, will provide more robust quantitative evidence on the impact of irrigation technology for small-scale farmers.

References

Abadie, Alberto and Guido W. Imbens (2015). "Matching on the Estimated Propensity Score". In: *NBER Working Paper*.

Anderson, Michael L. (2008). "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects". In: *Journal of the American Statistical Association* 103.484, pp. 1481–1495.

Burney, Jennifer A., Rosamond L. Naylor, and Sandra L. Postel (2013). "The case for distributed irrigation as a development priority in sub-Saharan Africa". In: *Proceedings of the National Academy of Sciences* 110.31, pp. 12513–12517.

Cohen, Craig R. et al. (2015). "Shamba Maisha: Pilot agricultural intervention for food security and HIV health outcomes in Kenya: design, methods, baseline results and process evaluation of a cluster-randomized controlled trial". In: *SpringerPlus* 4.1, pp. 1–17. ISSN: 2193-1801. DOI: 10.1186/s40064-015-0886-x. URL: http://dx.doi.org/10.1186/s40064-015-0886-x.

Fraiture, Charlotte de and Meredith Giordano (2014). "Small private irrigation: A thriving but overlooked sector". In: *Agricultural Water Management* 131, pp. 167–174.

Gebregziabher, Gebrehaweria, Regassa E. Namara, and Stein Holden (2012). "Technical Efficiency of Irrigated and Rain-Fed Smallholder Agriculture in Tigray, Ethiopia: A Comparative Stochastic Frontier Production Function Analysis". In: *Quarterly Journal of International Agriculture* 51.3, pp. 203–226.

Haushofer, Johannes and Jeremy Shapiro (2013). "Household Response to Income Changes: Evidence from an Unconditional Cash Transfer Program in Kenya". In: *Working Paper*.

Romano, Joseph P. and Michael Wolf (2005). "Stepwise Multiple Testing as Formalized Data Snooping". In: *Econometrica* 73.4, pp. 1237–1282. ISSN: 0012-9682.

Van Koppen, B., R. Namara, and C. Safilios-Rothschild (2005). *Reducing Poverty through Investments in Agricultural Water Management: Poverty and Gender Issues and Synthesis of Sub-Saharan Africa Case Study Reports*. International Water Management Institute.

Van Koppen, Barbara, Lesley Hope, and Willem Colenbrander (2012). "Gender Aspects of Small-scale Private Irrigation in Africa". In: *International Water Management Institute Working Paper*.

Xie, Hua et al. (2014). "Estimating the potential for expanding smallholder irrigation in Sub-Saharan Africa". In: *Agricultural Water Management* 131, pp. 183–193.

Zeweld, Woldegebrial et al. (2015). "Adoption of Small-Scale Irrigation and Its Livelihood Impacts in Northern Ethiopia". In: *Irrigation and Drainage* 64, pp. 655–668.

A Main Tables

A.1 Attrition

Table 0.A. Sample and Attrition by Cohort.

	Reschedule	Refused	Untraceable	Mteja	Relocated	Other	**Total**	$\frac{\text{Total}}{\text{Cohort Target Sample}}$
Cohort 1:	58	33	80	15	22	23	231	0.53
Cohort 2:	48	32	98	15	22	16	215	0.51
Cohort 3:	76	41	46	52	23	42	238	0.54
All Cohorts:	182	106	224	82	67	81	661	0.51
	Target	Completed	Success Rate					
Cohort 1	432	306	0.71					
Cohort 2	426	352	0.83					
Cohort 3	439	351	0.80					
Total:	1297	1009	0.78					

The upper panel of this table records the number of unsuccesful attempts to reach respondents in each cohort, including the total number of unsuccessful attempts and this total as a fraction of the total target sample by cohort. The bottom panel compares the target sample by cohort to the actual number of surveys completed.

Table 0.B. Attrition Test

	Attrition
Treatment Cohorts	0.033
	(0.024)
Constant	0.200
	(0.020)***
R^2	0.00
N	1,297

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

This table tests for attrition rates generated by assigning the outcome $attrition_i = 1$ if the respondent was in the target group of respondents but they were not successfully surveyed and $attrition_i = 0$ if they were in the target sample and were surveyed. This regression specification therefore corresponds to the bottom panel of Table 0.A, and is a test of our overall success in reaching the target number of respondents in each cohort. The discrepancy with the figures reported in the upper panel of Table 0.A is due to the fact that the figures in the upper panel reflect unsuccessful attempts to reach respondents, and that the number of attempts made was adjusted to try and reach the cohort targets.

A.2 Main Specification – Indices

Table 1.A. Equation 1

Outcome Var	(1) Asset Index	(2) Tot. Ag Time	(3) Land Inputs	(4) Average Monthly Revenue	(5) Weighted Consumption Index	(6) FS: Household	(7) FS: Children	(8) Psych Wellbeing	(9) Female Empowerment	(10) Domestic Violence
Treatment Cohorts	-27,504	-6.32	946	-4,756	-0.18	0.01	0.01	0.03	0.12	-0.35
	(18,709)	(9.70)	(1,895)	(17,481)	(6.68)	(0.06)	(0.06)	(0.07)	(0.07)*	(0.25)
Naive P-Value	[0.14]	[0.51]	[0.62]	[0.79]	[0.98]	[0.83]	[0.91]	[0.71]	[0.07]	[0.17]
R^2	0.05	0.07	0.13	0.02	0.11	0.09	0.04	0.04	0.03	0.05
N	995	995	995	995	995	993	993	995	995	995
Control Cohort Mean	224,408	36	24,336	88,933	19.16	0.00	-0.00	0.00	0.00	1.59
FWER p-value	0.73	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.51	0.76
FDR p-value	0.73	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.51	0.76

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Sample data top-coded at 95th percentile to account for outliers. Standard Error in parentheses below estimate. Controls for all specifications: District Fixed Effects. Units for Columns 1,3 and 4 in KSH. Column 2 in units of hrs/day during farming seasons (long rains, short rains and irrigated seasons). Column 5 is an index of consumption (in kg) by category, weighted by that category's share of consumption by expenditure. Columns 6-9 are weighted standardized averages of component variables. Column 10 is the total number of instances in the past 6 months.

Estimates treatment effect on the top-coded sample, using Equation 1 to identify overall treatment effect by comparing Cohorts 1 and 2 to Cohort 3.

Table 1.B. Equation 2

Outcome Var	(1) Asset Index	(2) Tot. Ag Time	(3) Land Inputs	(4) Average Monthly Revenue	(5) Weighted Consumption Index	(6) FS: Household	(7) FS: Children	(8) Psych Wellbeing	(9) Female Empowerment	(10) Domestic Violence
Cohort 1	-23,026	-21.31	375	-7,050	-5.94	-0.03	-0.08	-0.17	0.02	-0.30
	(22,655)	(11.36)*	(2,303)	(20,790)	(8.50)	(0.07)	(0.08)	(0.09)*	(0.09)	(0.30)
Naive P-Value	[0.31]	[0.06]	[0.87]	[0.73]	[0.48]	[0.70]	[0.29]	[0.06]	[0.79]	[0.33]
R^2	0.04	0.09	0.12	0.05	0.22	0.11	0.06	0.07	0.06	0.04
N	656	656	656	656	656	655	655	656	656	656
Control Cohort Mean	216,356	38.26	27,460	91,335	19.12	0.14	0.08	0.12	0.08	1.33
FWER p-value	0.92	0.44	0.99	0.99	0.97	0.99	0.92	0.44	0.99	0.92
FDR p-value	0.92	0.45	0.99	0.99	0.96	0.99	0.92	0.45	0.99	0.92

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Sample data top-coded at 95th percentile to account for outliers. Standard Error in parentheses below estimate. Controls for all specifications: District Fixed Effects. Units for Columns 1,3 and 4 in KSH. Column 2 in units of hrs/day during farming seasons (long rains, short rains and irrigated seasons). Column 5 is an index of consumption (in kg) by category, weighted by that category's share of consumption by expenditure. Columns 6-9 are weighted standardized averages of component variables. These weighted standardized averages are normalized to be zero for Cohort 3, but in this table Cohort 2 is the control group. This explains why the Control Cohort Means are not zero for Columns 6-9. Column 10 is the total number of instances in the past 6 months.

Estimates difference in long-term and short-term treatment effect on the top-coded sample, using Equation 2 to identify differential treatment effect by comparing Cohorts 1 to Cohort 2.

Table 1.C: 2SLS – First Stage

Outcome Var	(1) Used Pump (1/0)
Treatment Cohorts	0.25
	(0.03)***
Naive P-Value	[0.00]
R^2	0.11
N	958
Control Cohort Mean	0.47

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Standard Error in parentheses below estimate. Controls for all specifications: District Fixed Effects.

This is the first stage regression in the Two-Stage Least Squares regression where Cohort status is used as an instrument for using the pump. There is a positive and significant relationship between being in the treatment cohorts and using the pump.

Table 1.D: 2SLS – Second Stage

Outcome Var	(1) Asset Index	(2) Tot. Ag Time	(3) Land Inputs	(4) Avg Monthly Revenue	(5) Weighted Cons Index	(6) FS: Household	(7) FS: Children	(8) Psych Wellbeing	(9) Female Empowerment	(10) Domestic Violence
Pump Use	-105,394	-29.13	2,401	-10,839	-4.61	0.00	0.02	0.10	0.48	-1.36
	(75,860)	(39.17)	(7,510)	(71,079)	(25.78)	(0.24)	(0.26)	(0.28)	(0.28)*	(1.03)
Naive P-Value	[0.16]	[0.46]	[0.75]	[0.88]	[0.86]	[1.00]	[0.95]	[0.71]	[0.08]	[0.19]
R^2	0.03	0.05	0.13	0.02	0.18	0.08	0.04	0.03	0.02	0.01
N	958	958	958	958	958	958	958	958	958	958
Control Cohort Mean	224,408	36.41	24,336	88,933	19.16	0.00	0.00	0.00	0.00	1.59

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Standard Error in parentheses below estimate. Controls for all specifications: District Fixed Effects.

This is the second stage regression in the Two-Stage Least Squares regression where Cohort status is used as an instrument for using the pump.

A.3 Main Specification – Components

Table 2.1. Components of Assets

Outcome Var	(1) Productive Assets	(2) Vehicles	(3) Durable Assets	(4) Livestock	(5) Savings	(6) Land and Buildings
Treatment Cohorts	-1,302	-21,170	-2,103	1,790	-4,150	87,599
	(1,533)	(15,929)	(1,706)	(1,534)	(3,624)	(72,263)
Naive P-Value	[0.40]	[0.18]	[0.22]	[0.24]	[0.25]	[0.23]
R^2	0.00	0.00	0.00	0.00	0.00	0.00
N	995	995	995	995	995	995
Control Cohort Mean	22,802	130,653	25,625	12,844	31,822	632,313

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Sample data top-coded to account for outliers. Standard Error in parentheses below estimate.

Controls for all specifications: District Fixed Effects.

Estimates treatment effect separately on different asset categories. Land and Buildings are not included in the Total Value of Assets Index. Units are in KSH.

Table 2.2. Components of Agricultural Time Use

Outcome Var	(1) Planting	(2) Weeding	(3) Plant Protection	(4) Fertilizer Application	(5) Irrigation	(6) Harvesting	(7) Other Farm Activities
Treatment Cohorts	-1.13	-2.07	0.43	-0.33	0.46	1.16	-0.17
	(1.72)	(3.60)	(1.01)	(0.70)	(2.86)	(2.04)	(0.38)
Naive P-Value	[0.51]	[0.57]	[0.67]	[0.64]	[0.87]	[0.57]	[0.66]
R^2	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N	995	995	995	995	995	995	995
Control Cohort Mean	6.37	12	2.09	2.42	6.94	5.58	.91

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Sample data top-coded to account for outliers. Standard Error in parentheses below estimate.

Controls for all specifications: District Fixed Effects.

Estimates treatment effect separately on the components that make up the Total Agricultural Time Use Index. Units are hrs/day during the agricultural seasons (long rains, short rains, and irrigated seasons).

Table 2.3. Components of Land Inputs

Outcome Var	(1) Fertilizer	(2) Seeds & Planting Materials	(3) Irrigation Water	(4) Pesticide & Agrochemicals	(5) Other Inputs
Treatment Cohorts	1,102	331	795	288	-124
	(593)*	(656)	(1,014)	(284)	(267)
Naive P-Value	[0.06]	[0.61]	[0.43]	[0.31]	[0.64]
R^2	0.00	0.00	0.00	0.00	0.00
N	995	995	995	995	995
Control Cohort Mean	6,616	7,597	5,041	2,842	2,168

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Sample data top-coded to account for outliers. Standard Error in parentheses below estimate.

Controls for all specifications: District Fixed Effects.

Estimates treatment effect separately on the components that make up the Total Value of Land Inputs Index. Units are KSH.

Table 2.4. Components of Revenue

Outcome Var	(1) Harvest	(2) Livestock	(3) Wage Labour	(4) Forestry	(5) Fish	(6) Enterprises	(7) Other Revenue Source
Treatment Cohorts	2,288	1,630	-4,497	-198	-0	-1,463	469
	(5,018)	(2,552)	(2,893)	(387)	(0)	(13,215)	(5,538)
Naive P-Value	[0.65]	[0.52]	[0.12]	[0.61]	[0.57]	[0.91]	[0.93]
R^2	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N	995	995	995	995	995	995	995
Control Cohort Mean	8,766	16,383	9,773	926	0	24,268	7,578

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Sample data top-coded to account for outliers. Standard Error in parentheses below estimate.

Controls for all specifications: District Fixed Effects.

Estimates treatment effect separately on the components that make up the Average Monthly Revenue. Units are KSH.

Table 2.5. Components of Weighted Consumption

Outcome Var	(1) Cereals	(2) Bread & Pasta	(3) Roots & Tubers	(4) Vegetables	(5) Meat	(6) Fish	(7) Dairy	(8) Oils & Fats	(9) Fruits	(10) Sweets	(11) Beverages	(12) Alcohol
Treatment Cohorts	-0.21	-0.03	-0.06	-0.19	0.93	-0.20	0.21	-0.11	-0.02	-0.02	0.22	0.03
	(4.76)	(0.09)	(0.05)	(0.56)	(0.85)	(0.34)	(0.35)	(0.13)	(0.09)	(0.02)	(0.58)	(0.13)
Naive P-Value	[0.97]	[0.70]	[0.27]	[0.73]	[0.28]	[0.57]	[0.55]	[0.40]	[0.83]	[0.21]	[0.70]	[0.81]
R^2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N	995	995	995	995	995	995	995	995	995	995	995	995
Control Cohort Mean	12.36	.2	.23	2.52	.88	.54	1.03	.45	.23	.04	.54	.09

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Sample data top-coded to account for outliers. Standard Error in parentheses below estimate.

Controls for all specifications: District Fixed Effects.

Estimates treatment effect separately on the components that make up the Total Consumption Index. Units are kg weighted by expenditure share of total consumption.

Table 2.6. Components of Food Security

Outcome Var	(1) HH: Meals Skipped	(2) HH: Days w/o Food	(3) Children: Meals Skipped	(4) Children: Days w/o Food	(5) Less Preferred Food	(6) Buy Food on Credit	(7) Borrow to buy Food	(8) Eat Two Meals/Day	(9) Has Food for Tomorrow
Treatment Cohorts	-0.02	-0.01	-0.01	-0.01	-0.01	0.02	-0.00	0.03	0.03
	(0.02)	(0.02)	(0.02)	(0.01)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Naive P-Value	[0.41]	[0.44]	[0.59]	[0.53]	[0.68]	[0.65]	[0.87]	[0.12]	[0.15]
R^2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
N	995	995	995	995	995	995	995	995	995
Control Cohort Mean	.15	.07	.08	.03	.49	.34	.12	.91	.87

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Sample data top-coded to account for outliers. Standard Error in parentheses below estimate.
Controls for all specifications: District Fixed Effects.

Estimates treatment effect separately on the components that make up the Food Security Index. Each component is a binary (1/0) answer to the question. Questions where a "Yes" answer indicated worse food security were flipped to create the index.

Table 2.7. Components of Psychological Wellbeing

Outcome Var	(1) Perceived Stress Scale	(2) Optimism Scale	(3) Self-Esteem Scale	(4) Depression Scale
Treatment Cohorts	0.22	0.28	0.94	1.25
	(0.25)	(0.35)	(0.64)	(1.07)
Naive P-Value	[0.38]	[0.42]	[0.14]	[0.25]
R^2	0.00	0.00	0.00	0.00
N	995	995	995	995
Control Cohort Mean	3.66	5.39	10.12	16.78

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Sample data top-coded to account for outliers. Standard Error in parentheses below estimate.
Controls for all specifications: District Fixed Effects.

Estimates treatment effect separately on the components that make up the Psych Wellbeing Index. Each scale is the sum of questions on a 4- or 5- point scale.

Table 2.8. Components of Empowerment

Outcome Var	(1) Actual Decisionmaking	(2) Hypothetical Decisionmaking
Treatment Cohorts	0.24	0.38
	(0.228)	(0.362)
Naive P-Value	[0.29]	[0.30]
R^2	0.00	0.00
N	995	995
Control Cohort Mean	3.46	2.88

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Sample data top-coded to account for outliers. Standard Error in parentheses below estimate.

Controls for all specifications: District Fixed Effects.

Estimates treatment effect separately on the components that make up the Empowerment Index. Actual decisionmaking is the sum of eight binary (1/0) variables indicating female participation (on their own or jointly with their spouse) in various houeshold decisions. Hypothetical decisionmaking is the sum of eight variables recording (on a scale from 1-4) to what extent a woman can make her own personal choices on the same eight household decisions, where a higher number indicates greater influence.

Table 2.9. Components of Domestic Violence

Outcome Var	(1) Emotional	(2) Physical	(3) Sexual	(4) Children
Treatment Cohorts	-0.24	-0.30	0.09	0.07
	(1.32)	(0.37)	(0.11)	(0.23)
Naive P-Value	[0.86]	[0.42]	[0.41]	[0.76]
R^2	0.00	0.00	0.00	0.00
N	995	995	995	995
Control Cohort Mean	2.78	.81	.12	.33

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Sample data top-coded to account for outliers. Standard Error in parentheses below estimate.

Controls for all specifications: District Fixed Effects.

Estimates treatment effect separately on the components that make up the Domestic Violence Index. Units are total incidents in each category for the last 6 months.

A.4 Propensity Score Analysis

Table 3 Baseline Balance Check

	(1) Cohort 1 Mean	(2) Cohort 2 Mean	(3) Cohort 3 Mean	(4) Pvalue: $C3 = C1/C2$	(5) Pvalue: $C1 = C2$
Baseline: Productive Assets	18,375	81,366	22,802	0.00	0.00
Baseline: Vehicles	48,710	8,926	130,653	0.00	0.01
Baseline: HHDurables	13,720	9,691	25,625	0.00	0.01
Baseline: Buildings	9,273	166,082	632,313	0.00	0.00
Baseline: TotalAssets	90,078	266,064	811,393	0.00	0.00
Baseline: IrrigatedAcres	0.15	0.13	1.16	0.00	0.93
Baseline: RainfedAcres	2.10	1.68	2.03	0.23	0.16
Baseline: TotalAcres	2.25	1.81	3.19	0.01	0.38
HH Size	4.84	4.94	4.98	0.84	0.61

Displays the means for various baseline characteristics for the three cohorts in the sample, topcoded at th 95th percentile. The two columns on the right display the pvalues for a test of whether the baseline characteristics are equal for the different cohorts. The fourth column presents the test for Cohort 3 v. Cohorts 1 and 2, and the fifth column presents the test for Cohort 1 v. Cohort 2. All comparisons where the pvalue is below 0.1 are statistically significant. Cohort 3 has statistically significant differences in all baseline characteristics other than Total Rainfed Acres and Household Size. Cohorts 1 and 2 are statistically different from one another in all categories of baseline assets.

Table 4.1 - Propensity Score Logit Regression, Cohorts 1 and 2

	Cohort 1
Baseline Productive Assets	-0.00
	(0.00)***
Baseline Vehicles	0.00
	(0.00)***
Baseline HH Durables	0.00
	(0.00)**
Baseline Total Irr Acres	0.077
	(0.06)
Baseline Total Rainfed Acres	0.03
	(0.01)***
HH Size	0.00
	(0.01)
Pseudo R2	0.14
N	578

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Propensity score logit regression, average marginal effects. These coefficients represent the average marginal effect (across respondents) of different baseline characteristics on likelihood of being in the treatment group.

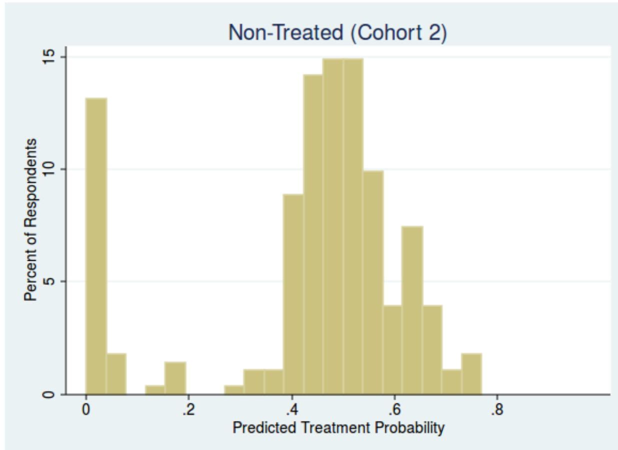

Figure 1: Propensity Score Distribution – Nontreated Cohorts

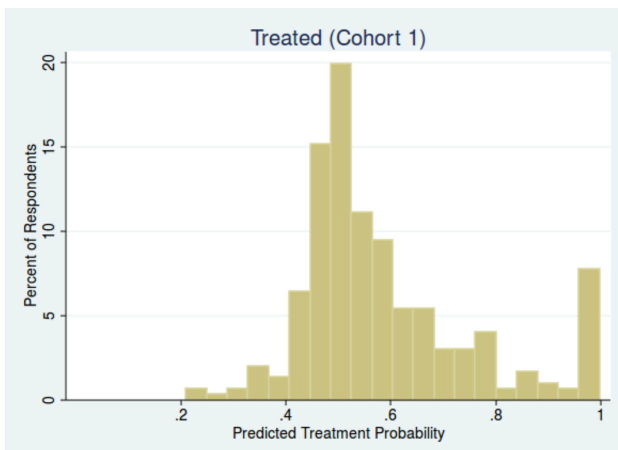

Figure 2: Propensity Score Distribution – Treated Cohorts

Table 4.2 - Treatment Effects: Propensity Score Matching, Cohorts 1 and 2.

Outcome Var	(1) Asset Index	(2) Tot. Ag Time	(3) Land Inputs	(4) Average Monthly Revenue	(5) Weighted Consumption Index	(6) FS: Household	(7) FS: Children	(8) Psych Wellbeing	(9) Female Empowerment	(10) Domestic Violence
Average Treatment Effect	-35,258	-8.76	110	-5,224	-2.62	0.04	-0.03	-0.01	0.01	0.56
	(33,809)	(12.88)	(2,460)	(12,577)	(5.89)	(0.10)	(0.10)	(0.23)	(0.10)	(0.47)
Naive p-value	0.30	0.50	0.96	0.68	0.66	0.69	0.73	0.96	0.91	0.24
N	578	578	578	578	578	578	578	578	578	578
Control Cohort Mean	216,356	38.26	27,460	91,335	19.12	0.14	0.08	0.12	0.08	1.33
FWER p-value	0.965	0.999	1.000	1.000	0.999	1.000	1.000	1.000	1.000	0.947
FDR p-value	0.962	0.996	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.919

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Sample restricted to exclude individuals whose propensity scores were outside the best-practice bandwidth from any other observations, meaning they could not be adequately matched for comparison. Standard Error in parentheses below estimate. Controls for all specifications: District Fixed Effects. Units for Columns 1,3 and 4 in KSH. Column 2 in units of hrs/day during farming seasons (long rains, short rains and irrigated seasons). Column 5 is an index of consumption (in kg) by category, weighted by that category's share of consumption by expenditure. Columns 6-9 are weighted standardized averages of component variables. Column 10 is the total number of instances in the past 6 months.

Table 5.1 - Propensity Score Logit Regression, All Cohorts

	Cohort 1 or 2
Baseline Total Irr Acres	-0.25
	(0.02)***
Baseline Total Rainfed Acres	0.02
	(0.01)***
HH Size	0.00
	(0.01)
Pseudo R2	0.1
N	915

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Propensity score logit regression, average marginal effects. These coefficients represent the average marginal effect (across respondents) of different baseline characteristics on likelihood of being in the treatment group.

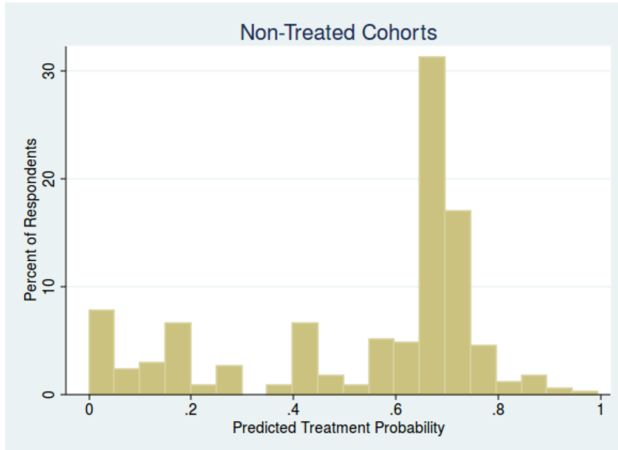

Figure 3: Propensity Score Distribution – Nontreated Cohorts

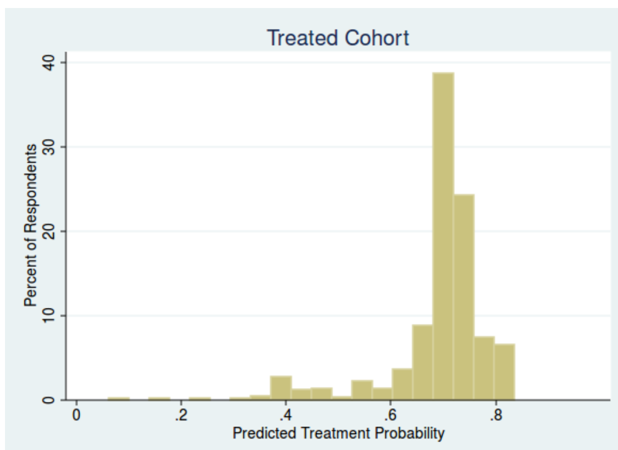

Figure 4: Propensity Score Distribution – Treated Cohorts

Table 5.2 - Treatment Effects: Propensity Score Matching, All Cohorts.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Outcome Var	Asset Index	Tot. Ag Time	Land Inputs	Average Monthly Revenue	Weighted Consumption Index	FS: Household	FS: Children	Psych Wellbeing	Female Empowerment	Domestic Violence
Average Treatment Effect	-10,894	16.64	6,591	-5,306	11.15	-0.02	0.02	-0.04	0.07	-0.42
	(30,126)	(12.87)	(3,494)*	(14,777)	(12.35)	(0.16)	(0.08)	(0.10)	(0.08)	(0.28)
Naive p-value	0.72	0.20	0.06	0.72	0.37	0.88	0.74	0.70	0.40	0.13
N	913	913	913	913	913	913	913	913	913	913
Control Cohort Mean	224,408	36.41	24,336	88,933	19.16	0.00	-0.00	0.00	0.00	1.59
FWER p-value	1.00	0.83	0.49	1.00	0.96	1.00	1.00	1.00	0.96	0.73
FDR p-value	1.00	0.83	0.47	1.00	0.96	1.00	1.00	1.00	0.96	0.73

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Sample restricted to exclude individuals whose propensity scores were outside the best-practice bandwidth from any other observations, meaning they could not be adequately matched for comparison. Standard Error in parentheses below estimate. Controls for all specifications: District Fixed Effects. Units for Columns 1,3 and 4 in KSH. Column 2 in units of hrs/day during farming seasons (long rains, short rains and irrigated seasons). Column 5 is an index of consumption (in kg) by category, weighted by that category's share of consumption by expenditure. Columns 6-9 are weighted standardized averages of component variables. Column 10 is the total number of instances in the past 6 months.

A.5 Qualitative Analysis

Table 6.1.A: Pump Effects – Cohort 1

How has your main pump changed your investment & lifestyle?	Number	Percent
Have expanded my crop area	187	21
Have sent my children to a better school	131	15
Have saved more money	107	12
Have bought more livestock	59	7
Have bought building material	47	5
Have bought better clothes for myself or family	41	5
Have bought a new farm equipment	40	5
Have bought other home appliances	31	4
Have bought new kitchenware	24	3
Have started a new business	22	2
Have invested more in an existing business	21	2
Have bought improved livestock breeds	13	1
Have bought new furniture	13	1
Have bought new phone	12	1
Other	117	13
Don't know	3	0
Not applicable	14	2
Total	882	100

Note: This tabulates **responses** by respondents, where respondents were allowed to select several options.

Table 6.1.B: Pump Effects – Cohort 2

How has your main pump changed your investment & lifestyle?	Number	Percent
Have expanded my crop area	136	21
Have saved more money	96	15
Have sent my children to a better school	84	13
Have bought better clothes for myself or family	33	5
Have bought more livestock	29	4
Have bought building material	28	4
Have bought a new farm equipment	22	3
Have bought other home appliances	21	3
Have bought new kitchenware	18	3
Have started a new business	16	2
Have bought new furniture	14	2
Have bought improved livestock breeds	13	2
Have invested more in an existing business	10	2
Have bought new phone	8	1
Other	98	15
Don't know	1	0
Not applicable	22	3
Total	649	100

Note: This tabulates **responses** by respondents, where respondents were allowed to select several options.

Table 6.1.C: Pump Effects – Cohort 3

How has your main pump changed your investment & lifestyle?	Number	Percent
Have expanded my crop area	157	23
Have saved more money	80	12
Have sent my children to a better school	67	10
Have bought building material	31	5
Have bought better clothes for myself or family	30	4
Have bought more livestock	26	4
Have started a new business	22	3
Have bought improved livestock breeds	21	3
Have invested more in an existing business	18	3
Have bought a new farm equipment	13	2
Have bought other home appliances	13	2
Have bought new furniture	8	1
Have bought new kitchenware	7	1
Have bought new phone	5	1
Other	131	19
Don't know	5	1
Refuse to answer	1	0
Not applicable	37	6
Total	672	100

Note: This tabulates **responses** by respondents, where respondents were allowed to select several options.

Table 6.2: Most Important Problems

Question II	
Most important problems in using the MoneyMaker	Percent
Requires two people to operate	13.11
No problems at all	13.11
Pump hard to operate	9.93
Rubber caps wears out quickly	9.49
Water too deep for pump to reach	7.02
Lack of spare parts	6.24
General muscle pain/cramps/blister	5.27
Cannot irrigate a large area	5.21
Backache resulting from operating pumps	3.78
Lack of hosepipe	3.4
Lack of technical support on use of equipment	3.24
Hip pain resulting from operating pumps	3.18
Loses a lot of water	2.63
lack of local technical skills for service and maintenance	2.52
Respiratory problems resulting from operating pump	1.26
Rusting	0.99
lack of time to operate equipment efficiently	0.55
Other health problems resulting from operating pumps	0.44
Lack of customers to hire/use equipment	0.16
Don't know	0.11
Not applicable	1.32
Other	6.86

Table 6.3: Biggest Concerns

Question III	
What are your biggest worries about next year?	Percent
School fees/Tuition	24.72
Production related (rainfall yield)	19.49
Drought	11.12
Health concerns	7.64
Feeding the family	4.35
Flooding	3.75
Pests	2.88
Political or social unrest	1.34
Marital Strife	0.54
Don't know	3.88
Refuse to answer	0.4
Not applicable	3.75

B Supplementary Tables

B.1 Top-coding Robustness

Table 7.1. Equation 1, No Topcoding.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Outcome Var	Asset Index	Tot. Ag Time	Land Inputs	Average Monthly Revenue	Weighted Consumption Index	FS: Household	FS: Children	Psych Wellbeing	Female Empowerment	Domestic Violence
Treatment Cohorts	-2,015,755	-6.32	3,321	-4,754	-0.18	0.01	0.01	0.03	0.12	-0.01
	(1,104,236)*	(9.70)	(5,492.73)	(16,713)	(6.69)	(0.06)	(0.06)	(0.07)	(0.07)*	(1.62)
Naive P-Value	[0.07]	[0.51]	[0.55]	[0.78]	[0.98]	[0.83]	[0.91]	[0.71]	[0.07]	[0.99]
R^2	0.02	0.07	0.07	0.03	0.11	0.09	0.04	0.04	0.03	0.06
N	995	995	995	995	995	993	993	995	995	995
Control Cohort Mean	106,153	32.47	10,025	25,965	15.25	0.00	0.00	0.00	0.01	0.53
FWER p-value	0.46	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.49	1.00
FDR p-value	0.42	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.45	1.00

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Standard Error in parentheses below estimate. Controls for all specifications: District Fixed Effects. Units for Columns 1,3 and 4 in KSH. Column 2 in units of hrs/day during farming seasons (long rains, short rains and irrigated seasons). Column 5 is an index of consumption (in kg) by category, weighted by that category's share of consumption by expenditure. Columns 6-9 are weighted standardized averages of component variables. Column 10 is the total number of instances in the past 6 months.

Estimates treatment effect on the non top-coded sample, using Equation 1 to identify overall treatment effect by comparing Cohorts 1 and 2 to Cohort 3.

Table 7.2: Equation 1, Outliers Dropped

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Outcome Var	Asset Index	Tot. Ag Time	Land Inputs	Average Monthly Revenue	Weighted Consumption Index	FS: Household	FS: Children	Psych Wellbeing	Female Empowerment	Domestic Violence
Treatment Cohorts	-9,631	-2.29	4,465	-1,091	0.28	0.01	0.01	0.00	0.11	-0.15
	(13,942)	(1.96)	(1,450)***	(4,121)	(0.23)	(0.06)	(0.06)	(0.06)	(0.06)*	(0.16)
Naive P-Value	[0.49]	[0.24]	[0.00]	[0.79]	[0.23]	[0.83]	[0.91]	[0.99]	[0.08]	[0.33]
R^2	0.06	0.07	0.14	0.05	0.11	0.09	0.04	0.04	0.04	0.04
N	946	946	946	946	946	993	993	946	950	958
Control Cohort Mean	73,890	4.83	9,284	23,597	4.68	-0.00	-0.00	-0.14	-0.09	0.19
FWER p-value	0.99	0.85	0.02	1.00	0.82	1.00	1.00	1.00	0.54	0.96
FDR p-value	0.99	0.85	0.02	1.00	0.84	1.00	1.00	1.00	0.57	0.95

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Standard Error in parentheses below estimate. Controls for all specifications: District Fixed Effects.

Estimates treatment effect on the top-coded sample, but where outliers have been removed entirely, instead of having variables censored, using Equation 1 to identify overall treatment effect by comparing Cohorts 1 and 2 to Cohort 3.

B.2 Propensity Score Robustness

Table 8.1: Propensity Score Logit – Total Land

	Cohort 1 or 2
Baseline Total Farm Acres	-0.02
	(0.01)***
Household Size	0.00
	(0.01)
R2_P	0.01
N	915

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Standard Error in parentheses below estimate. Controls for all specifications: District Fixed Effects.

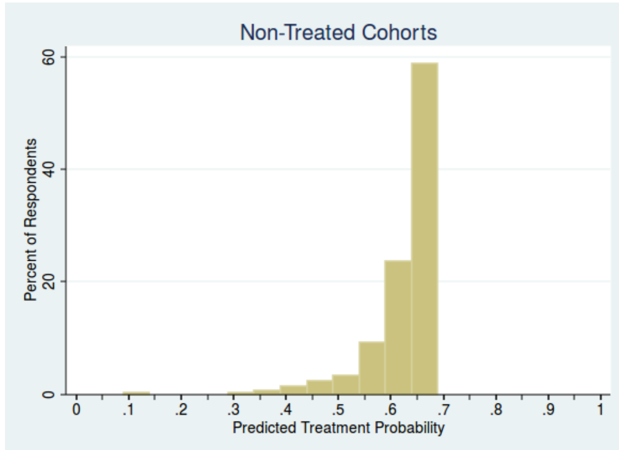

Figure 5: Robust Propensity Score Distribution – Nontreated Cohorts

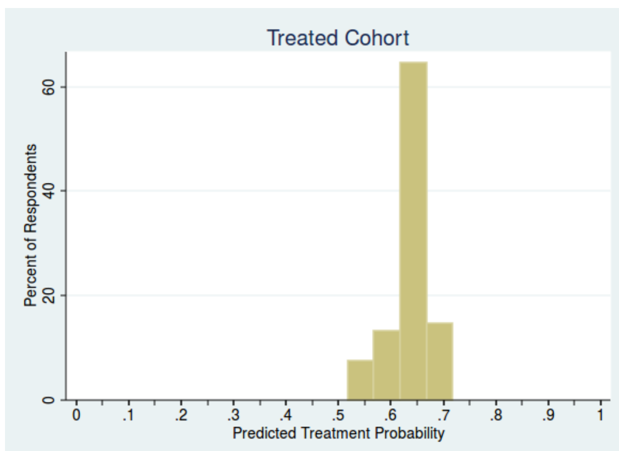

Figure 6: Robust Propensity Score Distribution – Treated Cohorts

Table 8.2: Propensity Score Treatment Effects – Total Land

Outcome Var	(1) Asset Index	(2) Tot. Ag Time	(3) Land Inputs	(4) Average Monthly Revenue	(5) Weighted Consumption Index	(6) FS: Household	(7) FS: Children	(8) Psych Wellbeing	(9) Female Empowerment	(10) Domestic Violence
Average Treatment Effect	-3,537	6.18	2,981	-4,618	4.55	0.07	0.04	0.13	0.05	-0.21
	(26,990)	(10.25)	(2,846)	(17,193)	(5.67)	(0.08)	(0.08)	(0.08)	(0.08)	(0.31)
Naive p-value	0.90	0.55	0.30	0.79	0.42	0.35	0.65	0.12	0.56	0.49
N	913	913	913	913	913	913	913	913	913	913
Control Cohort Mean	223,746	36.30	24,264	88,670	19.10	0.00	0.00	0.00	0.00	1.59
FWER p-value	1.00	0.84	0.52	1.00	0.98	1.00	1.00	1.00	0.98	0.74
FDR p-value	1.00	0.84	0.46	1.00	0.97	1.00	1.00	1.00	0.97	0.72

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Standard Error in parentheses below estimate. Controls for all specifications: District Fixed Effects.

Estimates Propensity Score Treatment Effects using a different specification including only total land used.

Table 8.3: Post Matching Covariate Balance Check

	Standardized Differences		Variance Ratio	
	Raw	Matched	Raw	Matched
Baseline Total Acres	-.194521	-.052459	.2341404	.6085805
HH Head Education	-.1343392	-.0540919	1.04189	1.003715
HH Head Gender	.0107401	-.0058296	1.029208	.9843916
HH Size	-.0028163	-.0452887	.8691142	.7142926

Covariate balance check after matching on baseline total acres farmed.

Table 8.4. Treatment Effects: Doubly-Robust. All Cohorts. Full Treatment Model.

Outcome Var	(1) Asset Index	(2) Tot. Ag Time	(3) Land Inputs	(4) Average Monthly Revenue	(5) Weighted Consumption Index	(6) FS: Household	(7) FS: Children	(8) Psych Wellbeing	(9) Female Empowerment	(10) Domestic Violence
ATE	-33,595.328	-12.026	4,387.664	2,677.374	12.689	0.022	0.024	-0.020	0.040	-0.566
	(23,952.205)	(13.255)	(3,218.613)	(20,445.955)	(6.473)**	(0.092)	(0.083)	(0.079)	(0.073)	(0.270)**
Potential No Treatment Mean	226,564.691	40.898	23,828.619	81,925.599	14.410	0.027	0.035	0.093	0.017	1.681
	(20,928.723)***	(11.734)***	(1,970.454)***	(9,724.735)***	(3.215)***	(0.069)	(0.075)	(0.058)	(0.057)	(0.236)***
N	913	913	913	913	913	913	913	913	913	913
Naive p-value	0.16	0.36	0.17	0.90	0.05	0.81	0.77	0.80	0.59	0.04
Control Cohort Mean	223745.65	36.30	24,263.93	88,670.44	19.10	0.00	-0.00	0.00	-0.00	1.59
Control Std. Dev.	295517.94	159.71	31,003.56	273849.13	84.25	1.00	1.00	1.00	1.00	3.97

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Doubly-robust treatment effects (IPWRA) for all cohorts using the full propensity score treatment model.

B.3 Single Cohort Comparisons

Table 8.5. Treatment Effects: Doubly-Robust. Cohorts 1 and 2. Full Treatment Model.

Outcome Var	(1) Asset Index	(2) Tot. Ag Time	(3) Land Inputs	(4) Average Monthly Revenue	(5) Weighted Consumption Index	(6) FS: Household	(7) FS: Children	(8) Psych Wellbeing	(9) Female Empowerment	(10) Domestic Violence
Treatment Effect	6,778.784	-43.571	4,652.783	-836.759	-4.464	-0.008	0.007	-0.380	-0.215	-0.007
	(46,684.626)	(18.145)**	(5,734.056)	(28,358.776)	(5.422)	(0.113)	(0.108)	(0.158)**	(0.130)*	(0.639)
Potential No Treatment Mean	200,686.481	71.485	24,433.568	82,031.927	13.388	0.135	0.117	0.186	0.330	1.976
	(22,985.028)***	(14.097)***	(3,509.928)***	(23,803.067)***	(4.205)***	(0.057)**	(0.054)**	(0.088)**	(0.085)***	(0.436)***
N	578	578	578	578	578	578	578	578	578	578
Naive p-value	0.88	0.02	0.42	0.98	0.41	0.95	0.95	0.02	0.10	0.99
Control Cohort Mean	223745.65	36.30	24,263.93	88,670.44	19.10	0.00	-0.00	0.00	-0.00	1.59
Control Std. Dev.	295517.94	159.71	31,003.56	273849.13	84.25	1.00	1.00	1.00	1.00	3.97

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Doubly-robust treatment effects (IPWRA) for cohorts 1 and 2 using the full propensity score treatment model.

Table 8.6. Treatment Effects: Doubly-Robust. All Cohorts. Treatment Model: only Baseline Total Acres.

Outcome Var	(1) Asset Index	(2) Tot. Ag Time	(3) Land Inputs	(4) Average Monthly Revenue	(5) Weighted Consumption Index	(6) FS: Household	(7) FS: Children	(8) Psych Wellbeing	(9) Female Empowerment	(10) Domestic Violence
Treatment Effect	-19,206.379	-0.500	2,279.660	6,181.179	-1.549	0.041	0.010	0.024	0.095	-0.341
	(20,760.711)	(8.881)	(1,877.653)	(17,906.200)	(4.553)	(0.063)	(0.066)	(0.069)	(0.069)	(0.254)
Potential No Treatment Mean	216,471.849	32.495	23,834.252	84,860.669	16.858	0.032	0.032	0.030	0.007	1.553
	(17,503.263)***	(7.262)***	(1,600.545)***	(13,975.803)***	(3.857)***	(0.050)	(0.051)	(0.052)	(0.056)	(0.216)***
N	913	913	913	913	913	913	913	913	913	913
Naive p-value	0.36	0.96	0.23	0.73	0.73	0.51	0.87	0.73	0.17	0.18
Control Cohort Mean	223745.65	36.30	24,263.93	88,670.44	19.10	0.00	-0.00	0.00	-0.00	1.59
Control Std. Dev.	295517.94	159.71	31,003.56	273849.13	84.25	1.00	1.00	1.00	1.00	3.97

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Doubly-robust treatment effects (IPWRA) for all cohorts using the restrictive propensity score treatment model with propensity calculated using only baseline total acres farmed.

Table 8.7. Treatment Effects: Doubly-Robust. Cohorts 1 and 2. Treatment model: Baseline Total Acres Only

Outcome Var	(1)	(2) Asset Index	(3) Tot. Ag Time	(4) Land Inputs	(5) Average Monthly Revenue	(6) Weighted Consumption Index	(7) FS: Household	(8) FS: Children	(9) Psych Wellbeing	(10) Female Empowerment	Domestic Violence
Treatment Effect	-49,248.246	-41.546	945.125	-4,627.255	-3.700	-0.050	-0.082	-0.239	-0.054	-0.223	
	(19,857.978)**	(13.783)***	(2,178.836)	(17,276.249)	(5.995)	(0.088)	(0.092)	(0.101)**	(0.089)	(0.309)	
Potential No Treatment Mean	206,581.197	67.408	25,227.443	84,248.350	18.853	0.054	0.088	0.155	0.135	1.391	
	(16,256.401)***	(12.876)***	(1,477.613)***	(15,875.065)***	(4.432)***	(0.054)	(0.052)*	(0.072)**	(0.062)**	(0.251)***	
N	578	578	578	578	578	578	578	578	578	578	
Naive p-value	0.01	0.00	0.66	0.79	0.54	0.57	0.38	0.02	0.55	0.47	
Control Cohort Mean	223745.65	36.30	24,263.93	88,670.44	19.10	0.00	-0.00	0.00	-0.00	1.59	
Control Std. Dev.	295517.94	159.71	31,003.56	273849.13	84.25	1.00	1.00	1.00	1.00	3.97	

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Doubly-robust treatment effects (IPWRA) for cohorts 1 and 2 using the restrictive propensity score treatment model with propensity calculated using only baseline total acres farmed.

Table 9.1: C1 compared to C3

Outcome Var	(1) Asset Index	(2) Tot. Ag Time	(3) Land Inputs	(4) Average Monthly Revenue	(5) Weighted Consumption Index	(6) FS: Household	(7) FS: Children	(8) Psych Wellbeing	(9) Female Empowerment	(10) Domestic Violence
Cohort 1	-36,458	-11.79	715.07	-6,085	1.82	-0.01	-0.02	-0.04	0.14	-0.42
	(21,055)*	(10.38)	(2,144)	(19,258)	(8.04)	(0.07)	(0.08)	(0.08)	(0.08)*	(0.28)
Naive P-Value	[0.08]	[0.26]	[0.74]	[0.75]	[0.82]	[0.90]	[0.74]	[0.65]	[0.07]	[0.13]
R^2	0.07	0.07	0.15	0.03	0.13	0.09	0.04	0.07	0.03	0.07
N	711	711	711	711	711	709	709	711	711	711
Control Cohort Mean	224,408	36.41	24,336	88,933	19.16	0.00	0.00	0.00	0.00	1.59
FWER p-value	0.45	0.88	1.00	1.00	1.00	1.00	1.00	1.00	0.45	0.61
FDR p-value	0.45	0.85	1.00	1.00	1.00	1.00	1.00	1.00	0.45	0.59

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Standard Error in parentheses below estimate. Controls for all specifications: District Fixed Effects.

This analysis excludes Cohort 2 and directly compares Cohort 1 to Cohort 3.

Table 9.2: C3 compared to C2

Outcome Var	(1) Asset Index	(2) Tot. Ag Time	(3) Land Inputs	(4) Average Monthly Revenue	(5) Weighted Consumption Index	(6) FS: Household	(7) FS: Children	(8) Psych Wellbeing	(9) Female Empowerment	(10) Domestic Violence
Treatment Cohorts	-12,002	5.39	598	-8,563	-5.12	0.01	0.04	0.10	0.07	-0.18
	(24,390)	(13.65)	(2,467)	(23,928)	(7.19)	(0.08)	(0.08)	(0.09)	(0.09)	(0.33)
Naive P-Value	[0.62]	[0.69]	[0.81]	[0.72]	[0.48]	[0.95]	[0.66]	[0.25]	[0.40]	[0.58]
R^2	0.09	0.10	0.14	0.04	0.06	0.12	0.05	0.05	0.06	0.06
N	623	623	623	623	623	622	622	623	623	623
Control Cohort Mean	224,408	36.41	24,336	88,933	19.16	0.00	-0.00	0.00	0.00	1.59
FWER p-value	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.92	0.98	1.00
FDR p-value	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.93	1.00	1.00

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Standard Error in parentheses below estimate. Controls for all specifications: District Fixed Effects.

This analysis excludes Cohort 1 and directly compares Cohort 1 to Cohort 2.

C Outcome Variables

Primary outcome variables will be the following. Construction of outcome variables in more detail below.

1. Total value of assets owned (ksh)
2. Total time spent in agriculture
3. Total investment in agricultural inputs (ksh)
4. Average Monthly Revenue
5. Food Consumption Index (kg)
6. Psychological health Index
7. Female Decisionmaking Index
8. Domestic Violence in past 6 months

C.1 Index Outcome Variables

* Indicates an index variable representing a main hypothesis to be tested, and will be adjusted for multiple inference.
† indicates a secondary outcome or a component variable of an index. Results will be reported but will not be adjusted for multiple inference, as they are not main hypotheses.

C.1.1 Outcomes

1. Total value of Assets Owned (KSH)
   (a) Productive Assets (KSH)†
      i. Hip Pump
      ii. Moneymaker Hip Pump (MMP)
      iii. Super MMP
      iv. Motorized Pump
      v. Hose pipe
      vi. Ox-Ploughs
      vii. Oxen/work bulls
      viii. Knapsack sprayers
      ix. Wheelbarrows
      x. Ox-carts/donkey carts
      xi. Hand carts
      xii. Fishing equipment (boats, canoes, etc)

xiii. Sewing machine

(b) Vehicles (KSH)†

i. Motor vehicle/pickup truck

ii. Bicycle

iii. Motor cycle

(c) Household Durables (KSH)†

i. Refrigerator/cooler

ii. Cellular phone

iii. TV

iv. Radio

v. Other electronic equipment

vi. Satellite dishes

vii. Electric generator

viii. Solar panel

ix. Car battery

(d) Livestock †

i. Cattle

ii. Small Livestock

iii. Birds

(e) Savings†

i. Savings with institution (KSH)

ii. Savings with family/friends (KSH)

(f) Land and Buildings (KSH)†

i. House (main building only)

ii. Zero grazing unit

iii. Boreholes/wells

iv. Fish pond

Index:*: Sum of (a)-(e). (KSH)
Notes: Assets uses the household's reported current value. When this was not available, the household was assigned the median value for that asset category. Value of livestock owned is calculated using the average sale price in the sample for each of the three types of livestock.

2. Time use (HRS)

(a) Time allocation of household members (males, females, externally hired labour) for rainfed plots (during rainy and dry season) and for irrigated crops to:

        i. Planting (hours)†
        ii. Weeding (hours)†
        iii. Plant Protection (hours)†
        iv. Fertilizer Application (hours)†
        v. Irrigation (hours)†
        vi. Harvesting (hours)†
        vii. Other farm operations (hours)†
    (b) Number of people hired for irrigation activities.†
    (c) How the pump has changed the time use of household members, separately by men, women, boys and girls, on the following activities:
        i. Leisure time†
        ii. Non-leisure social activities (e.g. faith-based activities)†
        iii. Rainfed crops†
        iv. Irrigated crops†
        v. Livestock production†
        vi. Non-farm economic activities†
        vii. School activities†
        viii. Household chores†

    Index*: Sum of (a).i-vii: Time working in agriculture (males, females and externally hired labour) (HRS)

3. Total investment in agricultural inputs (KSH)
   For each planting season for rainfed crops (during short and long rains) and for irrigated crops:

    (a) Fertilizer spending (KSH)†
    (b) Seeds & other planting materials (KSH)†
    (c) Water for irrigation (KSH)†
    (d) Pesticides & agrochemicals (KSH)†
    (e) Other inputs (e.g rented equipment) (KSH)†

   Index*: Sum of (a)-(e) Total amount spent on agricultural inputs on all plots in all seasons. (KSH)

4. Average Monthly Revenue (KSH)

    (a) Revenue from livestock (KSH)†
    (b) Revenue from harvest (KSH)†
    (c) Casual labor/ wage jobs (KSH)†
    (d) Revenue from forest and agro-forestry products (KSH)†

(e) Fisheries revenue (KSH)†

(f) Revenue from entrepreneurial activities (KSH)†

(g) Revenue from other sources (remittances, rental income, pension, charitable organizations) (KSH)†

Index*: Sum of (a)-(g), adjusted to monthly revenue: Average Monthly Revenue (KSH)

5. Consumption (KG)

   (a) Cereals (KG or relevant unit)†

   (b) Bread and pasta (KG or relevant unit)†

   (c) Roots and tubers (KG or relevant unit)†

   (d) Vegetables (KG or relevant unit)†

   (e) Meat (KG or relevant unit)†

   (f) Fish (KG or relevant unit)†

   (g) Dairy products (KG or relevant unit)†

   (h) Oils and fats (KG or relevant unit)†

   (i) Fruits (KG or relevant unit)†

   (j) Sweets (KG or relevant unit)†

   (k) Beverages (KG or relevant unit)†

   (l) Alcoholic beverages (KG or relevant unit)†

   Index*: Weighted sum of consumption (KG) in each category, weighted by the category's share of total consumption by expenditure. Expenditure shares are taken from a similar Kenyan dataset.

6. Food security

   (a) Meals skipped last month (adults)†

   (b) Whole days without food last month (adults)†

   (c) Meals skipped last month (children)†

   (d) Whole days without food last month (children)†

   (e) Less preferred foods last month†

   (f) Purchase of food on credit or borrowing last month†

   (g) HH members regularly eat 2 meals a day†

   (h) HH has enough food for tomorrow at home†

   Index*: Children - Weighted standardized average of variables (c), (d)
   Index*: Household – Weighted standardized average of variables (a)-(h)

7. Psychological health (only asked to females)
   (a) Perceived Stress Scale (PSS)†
   (b) Optimism Score (Scheier-Rosenberg)†
   (c) Self-esteem Score (Scheier-Rosenberg)†
   (d) Depression Score (CESD)†

   Index*: Weighted standardized average of measures (a)-(d)
   Notes: Each of these scales was created by summing up the score on each question (on either a four or five point scale) and reversing the scores for negatively coded questions. A higher score on the Optimism and Self-Esteem Scores is an indicator of good psychological health, but a higher score on the Perceived Stress Scale or the Depression Scale indicated poor psychological health. The final index is a weighted average of the four scales, where the scales were flipped such that a higher number always represents better psychological health.

8. Female decision-making. Includes actual ("do you contribute to decisions on ... ?") and potential ("Could you make your own independent choice on ... if you wanted to?") decision-making power of females over:
   (a) Minor household expenditure
   (b) How to respond to serious health problems
   (c) How to protect female household members from violence
   (d) Whether and how to express religious faith
   (e) Fertility decisions
   (f) Household tasks to do
   (g) Saving decisions
   (h) Planting decisions and plot management

   Index*: Empowerment: weighted standardized average of dummies for actual decisionmaking, and answers on a scale from 1-4 denoting hypothetical female decisionmaking power for the categories above.

9. Violence. Count of instances of the following forms of emotional and physical violence against female household members over the past 6 months.
   (a) Emotional Violence†
      i. Jealous/angry for speaking to other men (instances in last six months)
      ii. Accusations of unfaithfulness (instances in last six months)
      iii. Forbidden socialising with friends (instances in last six months)
      iv. Limited contact with family (instances in last six months)
      v. Did not trust her with money (instances in last six months)

vi. Humiliated her in front of friends (instances in last six months)
vii. Threatened her or someone close to her (instances in last six months)
viii. Insults (instances in last six months)

(b) Physical Violence†

i. Push, shake, throw something at her (instances in last six months)
ii. Slapped (instances in last six months)
iii. Twist arm / pull hair (instances in last six months)
iv. Punch with fist or something causing harm (instances in last six months)
v. Kicked, dragged, beaten up (instances in last six months)
vi. Choked or burned on purpose (instances in last six months)
vii. Threaten with a weapon (instances in last six months)

(c) Sexual Violence†

i. Physically forced sexual intercourse against her will (instances in last six months)
ii. Force her to perform sexual acts she did not want (instances in last six months)

(d) Violence Against Children†

i. Beat any of the children under age 12 (instances in last six months)

Index*: Sum of (a)-(d): all acts of domestic violence during the past six months.

10. Financial Variables:

(a) Total amount borrowed†
(b) Proportion of assets owned that were bought with borrowed money†
(c) Amount saved for purchasing assets†
(d) Unable to make repayment on loan (1/0)†

11. Housing:

(a) Iron Roof (1/0)†
(b) Solid, non-mud walls (1/0)†
(c) Solid, non-mud floors (1/0)†
(d) Number of rooms†

12. Female Asset Control:

(a) Proportion of assets over which female has (joint or individual control)†
(b) Proportion of harvests over which female has control†

    (c) Proportion of crop income over which female has control†
    (d) Average of dummies for female crop livestock ownership, across livestock types.†

13. Pump Questions:
    (a) Planning to buy a new pump (1/0)†
    (b) Type of pump (by category)†
    (c) Spillover: Total occasions pump was lent out†
    (d) Spillover: Total hours the pump was lent out†
    (e) Spillover: Has ever received cash/in-kind payment for lending the pump (1/0)†
    (f) Spillover: Social pressure to lend pump (1/0)†

14. Farm Attributes and Activities
    (a) Total area under cultivation (acres)†
    (b) Average perceived land fertility (across plots)†
    (c) Total land improvements (count of improvements on all plots)†
    (d) Crops grown (dummies for each specific crop)†
    (e) Mono-crop dummy†
    (f) Mixed-crop dummy†
    (g) Percent of land planted with improved varieties†
    (h) How often land is irrigated (estimated times per month, separately for each season)†

15. Water
    (a) Source of drinking water (category dummies)†
    (b) Number of jerrycans of drinking water per week†
    (c) Total cost of drinking water per week (KSH)†
    (d) Number of times per week water is collected†
    (e) Time spent collecting water†

D Field Notes & Formative Work

Not applicable.

E Sample Design

Discussed in main text above.

F Survey Instruments

Please see the attached endline survey instrument:

`KickStart_EndlineSurveyInstrument_060116_FINAL.docx`.

G Pre-Analysis Plan

Pre-Analysis Plan registered with the AEA RCT Registry, ID Number AEARCTR-0001002.

Link: `https://www.socialscienceregistry.org/trials/1002`

H Sample size and power calculations

The calculation to determine the Minimum Detectable Effect (MDE) is the following

$$\beta_{MDE} = (t_{1-\kappa} + t_{\alpha}) \cdot \sqrt{\frac{\sigma_Y^2}{P(1-P)N}}$$

where $1-\kappa$ is the power, α is the significance, σ_Y^2 is the baseline variance, n_T is the size of the treatment group and n_C is the size of the control group.

Power is set to be 0.8, and significance to be 0.05, treatment group size is 656 and control group size is 339, where Cohorts 1 and 2 are the treatment. Therefore $t_{1-\kappa}$ is 0.84, t_α is 1.65, P is approximately 0.66 and N is 995.

Therefore, considering an analogue to the endline Asset Index (the total value of productive assets, vehicle assets and household durables) which has a standard deviation of 123724.1, the minimum detectable effect is 20617.25, or 0.17 standard deviations. The mean of the baseline total value of non-building assets is 64731.06, meaning that the minimum detectable effect is equivalent to 31.85% of the baseline mean.

I Monitoring Plan

- **Back Checks** These checks consist of revisiting respondents that were earlier surveyed and asking them time-invariant questions from the questionnaire. Responses are matched with original responses to monitor the reliability and quality of the data collected. Back checks will be conducted for a randomly selected 10% of the sample.

- **Random Spot Checks and Field Observations** Field officers are supervised by project leads, who will regularaly sit with field officers to observe the manner in which questions are asked in the field. Specifically, project leads observe if probing occurs during questioning and advise the field officers on when and if necessary or appropriate. Observations will be recorded and feedback will be relayed to field officers on areas that require improvement and acknowledgement of areas that were conducted well. Additionall, senior project management will make random visits to the field (approximately bi-weekly). During the random spot checks, management will visit field officers to confirm that data protocols are being followed.

J Descriptive Statistics

	Cohort 1 Mean	Cohort 2 Mean	Cohort 3 Mean	Pvalue: C3=C1/C2	Pvalue: C1=C2
Asset Index	181,890	216,356	223,746	0.74	0.11
Agriculture Time Index	31.89	38.26	36.30	0.86	0.57
Land Input Index	26,040	27,460	24,264	0.17	0.53
Revenue Index	82,182	91,335	88,670	0.90	0.65
Weighted Consumption Index	20.06	19.12	19.10	1.00	0.91
Food Security (Household) Index	0.04	0.14	0.00	0.06	0.18
Food Security (Children) Index	0.02	0.08	0.00	0.30	0.43
Psych Index	-0.02	0.12	0.00	0.17	0.09
Empowerment Index	0.11	0.08	0.00	0.34	0.67
Domestic Violence Index	1.15	1.33	1.59	0.38	0.53

	Cohort 1 Mean	Cohort 2 Mean	Cohort 3 Mean	PvalC3=C1/C2	PvalC1=C2
Total Irrigated Acres	1.49	1.36	1.16	0.62	0.72
Total Irrigated Harvests	2.49	1.90	2.05	0.46	0.00
Net Irrigated Revenue	17,315	6,584	14,710	0.52	0.38
Family Borrowed (1/0)	0.15	0.12	0.02	0.00	0.22
Friends Borrowed (1/0)	0.16	0.12	0.00	0.00	0.05
Neighbour Borrowed (1/0)	0.27	0.34	0.01	0.00	0.02
Community Group Borrowed (1/0)	0.03	0.02	0.00	0.06	0.36

Season	Rank	Cohort 1	% planting	Cohort 2	% planting	Cohort 3	% planting
Irrigated	1	Kale	44%	Kale	50%	Kale	35%
	2	Tomatoes	27%	Tomatoes	26%	Tomatoes	18%
	3	Misc. Other	16%	Cabbage	23%	Misc. Other	16%
	4	Cabbage	15%	Misc. Other	17%	Cabbage	14%
	5	Onion	11%	Spinach	13%	Spinach	11%
Long rains	1	Maize	67%	Maize	69%	Maize	60%
	2	Beans	45%	Beans	42%	Beans	41%
	3	Misc. Other	13%	Misc. Other	13%	Misc. Other	12%
	4	Potato	6%	Kale	7%	Kale	11%
	5	Kale	6%	Potato	6%	Potato	6%
Short rains	1	Beans	28%	Beans	22%	Beans	29%
	2	Maize	28%	Maize	21%	Maize	27%
	3	Kale	10%	Kale	8%	Kale	10%
	4	Cabbage	5%	Misc. Other	5%	Misc. Other	5%
	5	Potato	5%	Cabbage	4%	Tomatoes	5%

K Do Files

See the attached dofile: `cohortAnalysis_MJFcomments.do` and the following ado files: `stepdown.ado`, `stepdown_PS.ado` and `stepdown_PS_allcohort.ado`.