INTERNATIONAL CONFERENCE OF AGRICULTURAL ECONOMISTS
ICAE
29th | Milan Italy 2015
UNIVERSITÀ DEGLI STUDI DI MILANO AUGUST 8 - 14
AGRICULTURE IN AN INTERCONNECTED WORLD

Targeting, bias, and expected impact of complex innovations on developing-country agriculture: Evidence from Malawi

Beliyou Haile, Carlo Azzarri, Cleo Roberts, David J. Spielman†

Sustainable intensification and climate-smart agriculture initiatives promote complex systems-based innovations to simultaneously improve yields and conserve natural resources. These innovations are usually tested under near-perfect experimental conditions with purposively selected farmers. Using a quasi-experimental approach and geographic information system, we evaluate a systems-based sustainable intensification project in Malawi aiming at improving whole-farm productivity and nutrition through integrated agricultural innovations. We find adopters of these innovations to systematically differ from non-adopters and suggestive evidence of potential systematic targeting of project locations and households. Econometric results using efficient influence function and propensity score matching methods show consistently higher maize yield and value of harvest, on average and across quantiles, for project beneficiaries, compared to that of randomly selected non-beneficiary households in non-target villages. Our findings highlight the need to rethink selection criteria for systems-based innovations, something that could potentially bear severe implications upon scaling up.

Keywords: Experiment, Evaluation, Innovation, Agriculture, Adoption, Policy

JEL codes: C93, D04, O31, Q01, Q16, Q18

†All co-authors are from International Food Policy Research Institute. Corresponding author: Beliyou Haile (b.haile@cgiar.org). We thank Zhe Guo and Eric Haglund for their research assistance and Africa RISING project partners in Malawi for sharing documentation on the Malawi project and for their assistance during site selection. All authors except David J. Spielman are members of the Africa RISING Monitoring and Evaluation team. The views expressed here are those of the authors and do not represent the views of the project's management, partner organizations, or the funding agency.

IAAE

1. Introduction

The livelihoods of rural households in many countries in sub-Saharan Africa (SSA) are based mainly on small-scale crop and livestock production systems. These systems are diverse, ranging from single-season, single cereal crop production to highly integrated tropical systems involving multiple cereal crops, roots and tubers, horticulture crops, and livestock. Common characteristics across these systems include low productivity, limited access to modern inputs and technologies, and vulnerability to weather shocks (Khan et al., 2014; Asfaw et al., 2012; Kamanula et al., 2011; Jayne et al., 2010; Shiferaw et al., 2008). Increases in short-term climate variability and long-term climate change are expected to exacerbate the challenges facing these systems and the people who manage and rely on them.

The adoption of locally appropriate technologies is often viewed as a primary means of improving agricultural productivity and strengthening the resilience of rural households in SSA (NEPAD, 2013; Asfaw et al., 2012; Minde et al., 2008). Innovations can increase the returns to farming, allow vulnerable households to accumulate income and assets, and enhance their ability to cope with weather and price shocks. Public spending on agricultural science is thus a high-return investment (Renkow and Byerlee, 2010; Raitzer and Kelly, 2008; Alston et al., 2000) with strong productivity and poverty impacts, especially when compared to alternative uses of scarce public resources (Fan and Pardey, 1997; Fan, 2000; Fan et al., 2000). Agricultural technologies may also complement other solutions designed to protect vulnerable rural households from shocks. Interventions such as microcredit services, conditional cash transfers, and weather index insurance can contribute to productivity growth or poverty reduction when coupled with the introduction of new technologies (Gilligan et al., 2009; Devereux, 2008; Devereux et al., 2008).

During the second half of the twentieth century, investments have been largely allocated to improving the uptake of improved cultivars and synthetic fertilizers, both of which are inputs which embody modern technology and are used in production in a fairly straightforward manner. More recently, with increased recognition of the threats posed by climate change to agriculture in SSA, emphasis has shifted to the uptake of complex, systems-based agricultural technologies and management practices.[1] This technology class is commonly characterized by

[1] The terms “systems-based technologies” and “systems approaches” are used here as shorthand for a broad class of complex and integrated technologies and practices. Scholars and practitioners familiar with many of the systems-based technologies

its simultaneous contributions to increasing on-farm productivity without comprising either future production capacity, the natural resource base, or the wider environment. While there are several ways to describe this approach, “sustainable intensification” (SI) is its most common labels (Garnett et al., 2013). SI may be viewed as an umbrella that covers the “climate-smart agriculture” (CSA) and/or “new agronomy” suite of solutions designed to increase productivity, conserve natural resources, and build resilience to frequent, complex, or unpredictable climatic shocks (The Montpellier Panel, 2013; Sumberg et al., 2012; Sumberg and Thompson, 2012; Beddington et al., 2012; Pretty et al., 2011). SI draws extensively on systems-based research, or the inter-disciplinary study of how plants, livestock, soil, water, and climate interact at plot, farm, and landscape levels, and how these interactions influence (and are influenced by) social and economic factors such as commodity prices, normative behaviors, institutions, and government interventions. SI technologies that result from this research are generally knowledge-intensive farm management approaches that balance modern inputs and improved cultivars with practices designed to conserve soil fertility, water, and biodiversity at both the farm and landscape levels.

In this paper, we examine the testing of several systems-based technologies in a research-for-development program, the Africa Research in Sustainable Intensification for the Next Generation - Africa RISING (hereafter AR). The main aim of AR is to understand whether systems-based technologies that are tailored to smallholder farmers’ local conditions and introduced in a sequential and participatory manner are more likely to be adopted by targeted beneficiaries and be scalable to similar populations and environmental settings. In Malawi, AR is testing these hypotheses with smallholders using a participatory action research approach described in detail later.

Our aim is to illustrate how the approaches being taken by the project, while possibly useful in strengthening the innovative capabilities of participating farmers and ensuring wide project coverage, cannot be used to say much about the efficacy or cost-effectiveness of SI, systems-based technologies, or participatory approaches to encouraging adoption. Using a quasi-experimental design, we examine the socio-economic characteristics of households targeted by the project to evidence strong sample selection biases. We then use efficient-influence function and propensity score matching methods to provide preliminary evidence of the predicted effects of the project on maize yields and the value of harvest, noting how these predicted

discussed in this paper will recognize the shortcomings of this terminology, thus we encourage its interpretation in the broadest terms only.

effects will likely fall short of the project's SI goals. Finally, we explore the implications of these findings for the project's scalability and demonstrate how household characterization is essential to improving targeting criteria and technology selection and to replicating the project across a broader population of farmers in Malawi and other countries and SI projects.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature to illustrate the complexity of systems-based technologies, the difficulties in evaluating their impact, and the resulting risks of investing in SI programs based on potentially biased evidence. Section 3 outlines the research context, evaluation design, and data used in the analysis. Section 4 discusses our identification strategies and Section 5 presents regression results. Section 6 concludes the paper.

2. Literature review

2.1 System-based technologies

Evidence on the productivity and poverty effects of technological change in developing-country agriculture has been built largely around a single class of innovations that we describe here as discrete, embodied technologies. These are products for which the underlying technological advantage resides in a tangible agricultural input and is directly realized by use of the input, typically within a single season. The most common examples are high-yielding cultivars and synthetic fertilizers. These technologies are often promoted as land scale-neutral packages that governments and markets can deliver with relative ease and farmers can readily integrate into their existing crop management practices. Furthermore, the impacts of these technologies on yields, returns, household food security, or national food staple production are relatively easy to measure, as demonstrated by many (Tilman et al., 2011; Pretty et al., 2011; Fan et al., 2011; Evenson and Gollin, 2003; Fan, 2000; Fan et al., 2000; Fan and Pardey, 1997).

Systems-based technologies, on the other hand, are complex sets of tangible and intangible elements that are combined with scientific guidance to bring about desired outcomes. Some elements may be discrete, and easy to identify, for example, a specific crop variety or synthetic fertilizer, a technique for removing weeds or pruning trees, a recommendation for the spacing of plants, or a timetable for the rotation of crops. But what characterizes the systems-based approach is the way these elements interact in a system to create synergistic effects, productivity and sustainability outcomes that are greater than the sum of their discrete parts. Consequently, descriptions of these technologies often consist of a series of steps to be taken or practices to follow. Any one of these practices might be advisable on its own, but to view the system as a complete technological innovation entails adopting all or most of its various components. Examples include integrated pest management, zero tillage wheat cultivation, and systems of rice intensification (Bennett and Franzel, 2013; Thierfelder et al., 2012; Giller et al., 2011&2009; Uphoff et al., 2010&2008; Senthilkumar et al., 2008; McDonald et al., 2008&2006; Van den Berg and Jiggins, 2007; Feder et al., 2004; Orr and Ritchie, 2004).

A second commonality of systems-based technologies is their high degree of context specificity. Because the innovation exists at the level of a system and not at the level of a particular piece of equipment, cultivar, or chemical product, systems-based technologies require considerable adaptation to accommodate variation in the site at which they are adopted. Indeed, such adaptation is essential, as the optimal way to manage a system in one location at one point in time is unlikely to be the same for a different set of agroecological and socioeconomic conditions. The exact specifications under which a system will function best on a given parcel of land, or the variant of the system that will best satisfy an individual farmer's preferences, are impossible to know a *priori*. Flexibility, though often not made explicit, is as crucial an aspect of any systems-based technology as any more clearly articulated rule or principle. Successful systems-based technologies provide a balance between clear, replicable, and transferable principles, and sufficient room for experimentation and adaptation. They offer precepts without prescriptions.

A third feature of systems-based technologies is that they place greater demands on farmers to learn new skills, revisit longstanding beliefs about agricultural practices, and adopt an experimental and empirically minded approach to farm management. In other words, they require the accumulation of capacity for learning, management, and adaptation. This notion of adaptive capacity is a central element in the study of complex adaptive systems (Hall and Clark, 2010; Nelson et al., 2007). While the changes in cognition and behaviour associated with increases in adaptive capacity are sometimes considered as complementary to or supportive of the technology itself, they are more properly understood as essential components of the technology. A systems-based technology can no more function without a capable farm manager than it can without water or sunlight.

A fourth feature of systems-based technologies is their groundings in the concept of resilience. The term "resilience" denotes the capacity of an ecological system to absorb perturbations, adapt to changes, and continue to function in a stable manner over time (Nelson et al., 2007;

Adger et al., 2011; Holling, 1973). The concept has been used to describe socio-economic systems in a similar manner (Levin et al., 1998; Folke, 2006), and expanded to combine agro-ecological and socio-economic resilience into a single conceptual framework (for example, Van Ginkel, 2014; Tittonell, 2013). In effect, systems-based technologies are applications of this combined resilience framework, or efforts to create sustainable production systems that are sufficiently robust to ensure that households, communities, and landscapes can withstand shocks, adapt to change, and continue to produce efficiently over time in a manner that ensures both food security and sustainable use of scarce natural resources.

All these features impose strict constraints on the probability and rate of adoption as identified by Rogers (2003), namely, difficulty in demonstrating relative advantage over other options; incompatibility with the user's practices, behaviours, norms, or values; complexity in application and understanding by the user; trialability, or the capacity to accumulate experience before committing fully to adoption; and observability by farmer. Or, drawing on Cash et al. (2003), individuals deciding whether to adopt a new technology will respond to information they perceive as being credible (believable), salient (relevant to the decision being made), and legitimate (produced through a process that respects stakeholders divergent values and beliefs), all of which are difficult to ascertain when the technology is disembodied, intangible, or incompletely defined. These two descriptions help frame the issue of why farmers may be dubious about adopting systems-based technologies before they have had a sufficient opportunity to evaluate its performance, adapt it to their own needs, and fit it into their social, economic, and cultural context.

These systems-based technologies are different propositions than many other technologies promoted during the past five decades. Some are even radical changes in how farmers cultivate their crops and manage livestock, water, land, soil, trees, residues, and waste. What to an outsider may seem like innocuous recommendations on planting dates, sowing methods, tillage practices, plant spacing, irrigation timings, or residue disposal are, for many farmers, counter-intuitive to generations of collective experience. They are, in effect, controversial and contentious-a concern recognized by Sumberg et al. (2013). Examples that are already highlighted in the discourse include practices as varied as integrated soil fertility management, integrated pest management, system of rice intensification, organic agriculture, minimum tillage, agroforestry, and conservation agriculture. And while very few of these technologies dispense entirely with improved cultivars or synthetic fertilizers, they do illustrate how the future of agriculture will likely be characterized by elements that are more intangible and disembodied from the technology, which in turn demands greater emphasis on farmer-led leaning and adaptation to their particular context.

2.2 Evidence and evaluation issues

Widespread adoption of systems-based technologies may rely on the accumulation of evaluative evidence on performance at the household and farm levels, particularly with respect to the decisions made by farmers who experiment with, refine and adopt (or not adopt, or dis-adopt) these technologies. While there are sound practical and theoretical reasons to believe that these systems-based innovations can be beneficial, there is still insufficient evidence on social and economic impacts. One reason for the lack of evidence is that systems-based technologies are simply difficult to measure. The technology must be readily identifiable and consistently applied by farmers for its impact to be measured and compared across individuals, farms and households.

But many of these new practices are contingent on high levels of site- and context-specificity, meaning that their successful application relies heavily on farmers' ability to adapt a complex set of cultivation and resource management techniques to the specific attributes of their own farm—its unique soil, water, and cropping conditions—and their own individual capabilities. Farmer adaptation, in turn, renders every technology slightly different when put into practice and leaves the evaluator without a consistent intervention to measure. Arguably, the design and implementation of impact evaluations needs to somehow account for the learning processes that make these complex practices work, particularly when learning is itself a reflection of farmers' unobservable innovative capabilities—aspirations, inquisitiveness, or entrepreneurial tendencies, for example. Evaluation efforts become even more challenging when trying to examine determinants and effects of adoption of technologies by a group of progressive farmers who either self-select or are selected by program implementers using specific criteria that may not always be visible to the evaluator.

Previous studies that attempted to quantify the impact of systems-based technologies find evidence of positive impacts and highlight the difficulties involved in accurately measuring the underlying agronomic and socioeconomic determinants of adoption of these technologies and subsequent effects. Some of the studies include Bennett and Franzel (2013) on conservation agriculture, Erenstein (2010) on minimum tillage systems, Kato et al. (2011) on integrated soil fertility management, Barrett et al. (2002) on system of rice intensification, and Franzel and Scherr (2002) on agroforestry. Most do a decent job in measuring the technology and parsing out its marginal effects on productivity and welfare, although their results still attract criticism and controversy (for example, Sumberg and Thompson, 2012). But such studies are still relatively rare, indicating that empirical evidence remains in short supply on issues such as farm-level costs and benefits, determinants of adoption within heterogeneous populations, and adoption dynamics. There is also limited evidence on the role of wealth, education, market access, information asymmetries, and individual preferences on farmers' willingness to adopt.

In some instances, these complex questions relating to both embodied and systems-based technologies have been pursued in experimental settings that address issues of sample selection bias, confounding variables, and identification of causal relationships (Banerjee and Duflo 2008; Duflo et al., 2008). Of course, the randomized controlled trials favored in these approaches have attracted criticism from economists (Deaton, 2009; Leamer, 2010; Barrett and Carter, 2010). The willingness to explore this approach with systems-based technologies has also attracted strong criticism from agronomists who develop and promote these technologies, and whose focus on farmer adaptation and innovation processes puts them at odds with the more structured designs of randomized controlled trials (RCTs).

Arguably, the resistance to RCTs in the evaluation of systems-based technologies also stems from constraints imposed by the project funding cycle and the wider aid effectiveness debate. When RCTs are used to evaluate the effectiveness of a large-scale rollout, they tend to slow down the capacity of project implementers from reaching large-scale outcome targets which, implicitly, frustrate the project donors who release funding against those targets. This is a likely outgrowth of the global discourse over development aid effectiveness which has boosted demand for tangible, numeric results from development programs. Achieving such results requires rapid rollouts of projects and technologies to (self-selecting) high-probability adopters, short project time horizons, rapid feedback to donor constituencies, and a growing culture of project management that puts numeric accomplishments over deeper understanding of complex development processes. The unintended consequences of this approach include the promotion of solutions—including systems-based technologies—without strong evidence of impact or cost-effectiveness. We explore these issues in the evaluation of a specific systems-based innovations in Malawi described in greater detail below.

Impact evaluation of agricultural projects has been shown to pose several challenges (Lucas and Longhurst, 2010; IEG, 2011). Ideally, an RCT could have been designed to evaluate how AR's approach in Malawi improves whole-farm productivity and other development outcomes.

Unfortunately, project expectations were such that the initial rollout was conducted by targeting non-representative groups of farmers: farmers who the implementers identified as early adopters, positive deviants, and model farmers who, through demonstration activities and peer effects, were expected to encourage wider adoption. Under such circumstances, any attempt to evaluate the project will run into several problems.

First, trying to evaluate impact by comparing participants with non-participants may reflect not only the impact of the innovations but also any innate difference between participants and non-participants (see Banerjee and Duflo, 2009 for general discussion). Self-selected beneficiaries are likely to have a wide range of characteristics, expectations, and perceptions that could determine whether and which type of innovations they adopt, as well as their realization of the potential benefits associated with those innovations. This in turn could overestimate impact on, say, crop yield since beneficiaries could have had a higher yield (than that of non-participants) even without the innovations.

Second, unless intervention sites and households are representative of the target rural population within which the menu of SI innovations are expected to be scaled up, external validity and extrapolation of impact will be compromised. In the case of AR Malawi, innovations are expected to be scaled up to reach a target rural population of approximately 6.6 million. One may observe high adoption of innovations and/or stronger impact among non-randomly selected group of progressive farmers in a high agricultural potential area but adoption and/or impact could be low when attempt is made to promote these innovations in low agricultural potential areas, for example. Third, when an intervention involves both a new technology and an advisory (training/learning) component, as does AR Malawi, an ideal evaluation design should be able to disentangle the technology effects from the learning effects. For the same reasons, an ideal evaluation design should be able to measure the potential spillovers on non-treated individuals who may come into contact with the treated (Angelucci and Di Maro, 2010).

3. Research context, evaluation design, and data

3.1 Research context

Africa RISING is an agricultural research-for-development program that aims to promote sustainable intensification of smallholder farming. It was initiated in 2012 and is being implemented in Malawi, Tanzania, Zambia, Ethiopia, Ghana, and Mali (ILRI, 2012).[2] Smallholder farmers participating in the program are offered with menu of various innovations and management practices, with the mix of innovations and their delivery mechanism varying across time, space, local context, and household typology. The Malawi Africa RISING project operates in Ntcheu and Dedza districts using mainly an approach called "mother and baby trials" (MBTs).

MBTs are adaptive research platforms created to identify and eventually disseminate successful innovations and practices with the active participation of farmers. Lead farmers selected from targeted villages actively participate in interactive, researcher-designed, scientifically replicable demonstration trials of the technologies being tested (the "mother" trials). Then other farmers select from among the technologies being tested in the mother trials the ones that meet their respective needs and replicates them on their farms (the "baby trial"). These farmers are likely to be individuals who meet certain criteria (for example, willingness to devote plots of a certain size to the trails). MBTs have been shown to encourage adoption among self-selecting farmers but still raise questions regarding replicability and cost-effectiveness (Kerr et al., 2007; Joshnson et al., 2003). More importantly, the participatory and experimental design of MBTs make socioeconomic evaluation difficult, both in terms of measuring its impact on participating households and on assessing its scalability to a wider population of farmers.

Figure 1 summarizes innovations applied by AR households at the farm level. The number of innovations ranges from seven (maize, NPK compund, Urea, and 4 types of legumes (from among ground nut, pigeon pea, cow pea, beans, and soya bean) by about 24% of the households to just two (puleses) by about 3% of the households. It is this heterogeneity in the type of innovations and the number of adopters that poses a challenge in identifying and attributing effects to a specific (mix of) innovation. Given the focus on integrated SI innovations, we also present a summary of technologies applied by AR households in at least one single plot in Figure 2.

Ground nut, maize fertilized with NPK and urea, soya bean, and pigeon pea are the four most common (mix of) innovations applied at a plot level. Common are also maize fertilized with NPK and Urea (and compost) and intercropped with a legume (pigeon pea, cow pea, or beans) and a system of "doubled-up" legume, where pigeon pea is intercropped with either soya bean

[2] More information about the program can be found here.

or groundnut or ground nut is intercropped with either soya bean, cow pea, or beans (sometimes fertilized with NPK). Previous studies have shown these integrated innovations to improve soil fertility, yield and nutrition while at the same time reducing fertilizer requirements (Bioversity, 2014; Mhango, 2011; Friesen and Palmer, 2004). The specific varieties of maize and legumes adopted by AR households are summarized in Table A1 in the Appendix.

3.2 Evaluation design

To examine project targeting criteria and address the evaluation challenges discussed in Section 2 to the extent possible, we employed a quasi-experimental evaluation design involving several steps.[3] First, to better understand the spatial pattern and homogeneity of the determinants of agricultural potential and stratify the target area by agro-ecologies, we use high-resolution geographic information system (GIS). After reviewing various biophysical and socio-economic data layers summarized in Table A2 in the Appendix, we find elevation and temperature-adjusted rainfall to adequately capture the variability in the biophysical characteristics of the study area. Using these two variables, the study districts were stratified as shown in Figure A1 in the Appendix.

Second, and after identification of program target villages by project implementers, we sampled four sections that are not the focus of the project (control sections). Control sections were chosen such that they would lie within each homogeneous agricultural potential area as AR target sections, while distant enough from action sections to avoid potential contamination. We then sampled 26 control villages using probability proportional to size. Third, we randomly sampled households both from control villages (hereafter "control" households) and AR target villages who were not directly targeted by the project (hereafter "non-beneficiary" households). While the number of AR beneficiary households was determined by project implementers, the sample size for non-beneficiary and control households was informed by our power calculation based on data from the Third Malawi Integrated Household Survey.[4] During August – October 2013, we collected detailed socioeconomic data as part of the Malawi Africa RISING Baseline Evaluation Survey (hereafter MARBES). Figure 3 summarizes our evaluation design.

[3] "Quasi" because the selection of AR target villages and households was non-random and not affected by our ideal evaluation design.

[4] Data and documentations on the Third Malawi Integrated Household Survey can be found here and additional details about the sample size calculation is available upon request.

While control households allow us construct the missing counterfactual to estimate project effects, non-beneficiary households allow us measure potential project spillovers. To the extent the project improves adoption of SI innovations and productivity also among non-beneficiary households, ignoring spillovers would lead to underestimation of project effect and its policy implications. Given the non-random selection of AR villages, comparison of outcomes for non-beneficiary households (in AR villages) and control households (in non-AR villages) captures not only spillovers but also potential effects of project targeting. Given that our survey was conducted right after AR beneficiaries collected their first harvest as project beneficiaries (year zero), the time lapse might have been relatively short to allow strong spillovers. In which case, non-beneficiary households could serve us within-village control group and comparison of AR beneficiaries with this group could provide evidence on project effects.

3.3 Data

Table 1 shows a summary of the study sample by group and geographic area. A total of 1,149 households were surveys of which 398 were AR beneficiary (all beneficiaries as of June 2013), 207 were non-beneficiary, and 544 were control. This sample came from 26 AR target and 28 control villages. A summary of village-level data shows that AR villages were slightly more remote and had lower (higher) population density (historical average rainfall), relative to control villages (Table 2). Also, AR target villages also have better access to agricultural extension services and farmers' cooperative groups.

A summary of selected household-level socioeconomic variables in Table 3 suggests the existance of systematic difference between the non-randomly selected AR households and the other two (randomly selected) groups. For example, AR beneficiaries exhibit larger household size, higher average adult years of education, and are more likely to have a married and male household head than non-beneficiary and control households. AR households are also better off in terms of different indicators of wealth we considered, including land ownership—total area of parcels operated by the household— and an asset-based wealth index we computed through factorial analysis (principal-component factor method) capturing ownership of household and agricultural durable assets, livestock, and housing characteristics. Ar households are also more likely to have parcels near their residence. There is insignificant differences between the three groups with respect to average travel time to the nearest seed supplier as well as access to basic services.[5] In terms of agricultural input use during October 2012 to May 2013, AR households report using more fertilizer (both chemical and manure) and agricultural labor (person-days per hectare) than the other two groups. They are also more likely to have used improved maize, irrigation, hired labor, and practice intercropping. Value of harvest is higher for AR households, as is yield of maize (Malawi's main staple crop).

A summary of the farming systems of households in Figure 4 shows that AR households are more likely to combine maize with many legumes while control households are more likely to practice mono cropping (of maize) or combine maize with fewer legumes. A plot-level summary of adopted technologies in Figure 5 shows maize-beans intercropping to be the most common technology mix, followed by intercropping of maize with one or more of the other pulses (ground nut, soya bean, cow pea, or pigeon pea). Common is also a system of "doubled-up" legumes where ground nut is intercropped with either soya bean or pigeon pea, especially among AR households.

These summary statistics demonstrate how AR households are systematically different from the wider population of farmers, on average, along the various dimensions we considered. The observed difference between AR households (and villages) and those randomly chosen from the population may also suggest other observed (but not measured) and unobserved differences that could affect technology adoption decisions and subsequent outcomes. This means that estimation of internally valid project effects requires controlling for pre-existing differences by building a valid counterfactual (Duflo et al., 2008; Burtless, 1995). Given that project beneficiaries are not representative of the broader population of farmers, project scale up efforts and strategies will also need to be examined carefully. In the next section, we discuss the identification strategy we employed to quantity early effects using one wave of cross-section data collected about three months after AR households collected their first harvest as a project beneficiary.

4. Identification strategy

To examine differences in expected outcomes between AR beneficiary and control households as well as between the other possible pairs, we employ Cattaneo's (2010) multivalued treatment

[5] Factorial analysis (principal-component method) is used to compute an access to services index covering access to the following services (measured by travel time, using the usual mode of transport) ─ motorable road, all-season road, asphalt road, weekly market place, daily market place, district capital, nearest place with daily bus stop, nearest health care facility, primary school, and secondary school.

effects approach. By allowing multiple pair-wise comparisons, this approach allows us to examine predicted project effects as well as potential spillovers and targeting. For example, we expect a significant difference between non-beneficiary and control households if there are spillovers within AR villages. Also, if AR systematically targeted relatively better-off villages, as the descriptive statistics seem to suggest, significant differences between non-beneficiary (in AR target villages) and control households (in non-AR villages) would be expected, over and above what can be explained by spillovers.

Following Cattaneo et al. (2013), we formally specify a potential-outcome model with three treatment types as shown in Equation 1.

$$y_i = \sum_{\tau=0}^{2} d_i(\tau) y_i(\tau) \qquad 1$$

where τ is an index for treatment type ($\tau = 0$ if treatment type is control, 1 if non-beneficiary, and 2 if AR beneficiary) and i is an index for household (i=1, 2, …, N); y is the observed outcome of interest; $d(\tau)$ is an indicator that equals 1 if treatment type is τ and 0 otherwise; and $y(\tau)$ is the outcome when treatment type is τ. Assuming a linear functional form, the outcome equation can be specified as follows (omitting subscript i).

$$y(\tau) = \boldsymbol{\beta}'_{\tau}\boldsymbol{X} + \epsilon_{\tau} \qquad 2$$

where $\boldsymbol{X}$ is a vector of covariates expected to affect y and ϵ is the error term. For each i, observed vectors $\boldsymbol{W}_i$ {= $(\tau, y(\tau), \boldsymbol{X})'$} and $\boldsymbol{y_i}$ {= $(y_i(0), y_i(1), y_i(2))'$} are assumed to be an independently and identically distributed (iid) draw from $\mathbf{W}$ and $\boldsymbol{y}'$. While our random sampling of non-beneficiary and control households removes some of the systematic correlation between treatment assignment and observed and unobserved characteristics, purposive selection of AR villages and households illustrated above necessitates adjustment for possible pre-treatment differences between the three groups. Formally, the treatment probability can be modelled as follows.

$$\mathrm{d}(\tau) = \begin{cases} 1 \; if \; \boldsymbol{\Gamma}'_{\tau}\boldsymbol{Z} + \varepsilon_{\tau} > 0 \\ 0 \;\; \text{otherwise} \end{cases}, \tau = 0, 1, 2 \qquad 3$$

where $\boldsymbol{Z}$ is a vector of covariates that could affect treatment type and whose elements may overlap with those of $\mathbf{X}$ and ε is the error term. Assuming selection on observables and common support condition, Cattaneo (2010) proposes two estimators of conditional means and quantiles of the potential outcome distributions based on a two-step generalized method of moments – inverse probability weighting (IPW) and efficient-influence function (EIF). The IPW estimator models only the selection probability in Equation 3 while the EIF estimator models both the selection probability and outcome equations by including an augmentation term in the latter to correct for potential misspecifications in the former (Cattaneo, 2010). Both estimators involve estimation of the generalized propensity scores (GPS) in the first stage and inverse probability weighting of observed outcomes in the second stage when recovering parameters of the potential outcomes distributions.[6] The "doubly robust" EIF estimator is shown to produce consistent parameter estimates when either model is correctly specified while IPW estimates will be inconsistent if there is misspecification in the selection equation and the propensity scores thereof are biased (Cattaneo, 2010; Tan, 2010; Zhao, 2004).[7] In this paper, we present EIF estimates and compute standard errors using bootstrapping (Cattaneo et al., 2013).[8]

Assuming the errors in Equation 3 to be iid with logistic distribution, we estimate the following multinomial logit model using maximum likelihood.

$$Prob(T_i = \tau|\boldsymbol{Z_i}) = P_{i\tau} = \frac{exp(\boldsymbol{\Gamma}_\tau' \boldsymbol{Z_i})}{1+\sum_{k=1}^{2} exp(\boldsymbol{\Gamma}_k' \boldsymbol{Z_i})}, \tau = 1, 2 \qquad 4$$

where T measures the treatment status and the probability for the reference category, control group, is given by $\frac{1}{1+\sum_{k=1}^{2} exp(\boldsymbol{\Gamma}_k' \boldsymbol{Z_i})}$. The vector **Z** includes variables that affect treatment probability including household size, sex and age of the household head, average adult education in the household, travel time to the nearest seed supplier, elevation, and asset-based wealth index, along with squared terms of the continuous controls. Since AR focuses on a variety of cultivars and different fertilizer application rates and almost all the study households grew maize during the reference period, we examine project effects on two main variables - the value of harvest and maize yield during October 2012 – May 2013. For the value of harvest, we specify the following fixed effects model.

[6] See Cattaneo (2010) for large sample properties of these estimators and Cattaneo et al. (2013) for implementation details in Stata software.
[7] It is not clear which estimator (IPW or EFI) is more robust when both the outcome and the selection probability models are misspecified (Tan, 2010).
[8] We estimated conditional means and quantiles using IPW estimator and found IPW estimates to be consistently higher than the EIF estimates.

$$HarvestValue_i = \beta + \boldsymbol{\Sigma}' \boldsymbol{Input}_i^{Total} + \boldsymbol{\Pi}' \boldsymbol{X}_i + \epsilon_{\mathrm{i}} \qquad 5$$

where $HarvestValue$ is the value of harvest (in thousands of Malawi Kwacha - MWK); $\boldsymbol{Input}^{Total}$ is a vector of *total* agricultural inputs used during the reference period including operated land area (hectare), agricultural labor (person-days per hectare), and fertilizers (kilograms per hectare); $\boldsymbol{X}$ is a vector of other household-level covariates including household size, age and sex of household head, average adult education, travel time to seed supplier, distance to basic services index, an index for agriculture-related wealth (excluding land), and elevation of the household's residence.[9] For maize yield, we specify the following fixed effects model.

$$MaizeYield_i = \alpha + \boldsymbol{\Gamma}' \boldsymbol{Input}_i^{Maize} + \boldsymbol{\Pi}' \boldsymbol{X}_i + \epsilon_i \qquad 6$$

where $MaizeYield$ is maize yield (ton per hectare), $\boldsymbol{Input}^{Maize}$ is a vector of agricultural inputs applied on *maize plots* including labor (person-days per hectare) and fertilizers (kilograms per hectare) and use of improved maize seed, and other variables are as defined before. After estimating conditional means and quantiles using the EIF estimator, we recover expected average and quantile treatment effects through pair-wise comparison of estimated parameter (Cattaneo et al., 2013). We estimate quartiles to examine potential heterogeneity along the distribution of the two variables. As a robustness check, we also present average treatment effects using propensity score matching using kernel and nearest neighbour matching (Rubin, 1974 &1978; Rosenbaum and Rubin, 1983; Abadie and Imbens, 2012).

5. Results

Before presenting conditional means and quartile estimates, we first check whether each household has a positive probability of being in each treatment (the overlap assumption) and that there is no mass of observations with predicted probabilities too close to zero or one (Khan and Tamer, 2010). Figure 6 shows the estimated densities of the predicted probabilities that each type of household (beneficiary, non-beneficiary, and control) is a control (Panel A), non-beneficiary (Panel B), and AR beneficiary (Panel C). Parameter estimates of the best fit multinomial logit model are reported in Table A3 in the Appendix. The overlap plots (Busso et al. 2013) show that the estimated conditional densities for each treatment type have most of their respective masses in the regions where they overlap. While there is no mass of

[9] Agricultural wealth index is computed using factorial analysis (principal-component method) and includes livestock ownership (in value and tropical livestock units) and ownership of various durable agricultural assets (such as cutlass, ox-plough, shovel, tractor, and ox-ridger).

observations at the extreme tails of these distributions, it is worth noting that for AR *beneficiaries*, there is a relatively high mass of predicted probability of being in the control towards the left tail of the distribution.

On the other hand, and as would be expected, for control households there is a relatively high mass of predicted probability of being a beneficiary towards the left tail of the distribution. Estimated densities for non-beneficiaries are generally between those of beneficiary and control households. It is worth noting that while the results from our site stratification exercise resulted in delineation of broad geographic areas to choose AR sites from, selection of AR target villages and households was also guided by other considerations by project implementers, such as synergies with other (similar) projects run by the same.

A graphic summary of EIF estimates of conditional means and quartiles for the value of harvest and maize yield are shown in Figures 7 and 8, respectively, along with the corresponding 95 percent confidence intervals. For both variables, means and quartiles differ by treatment type, with control and beneficiary households having the smallest and the highest values, respectively, and estimates for non-beneficiaries being in between. Especially for the value of harvest, means and quartiles are estimated imprecisely for non-beneficiary households, given the relatively small sample size. Expected average and quantile effects (of the project and targeting) is then computed through pair-wise comparison of the respective parameters.[10]

In addition to the traditional comparison between project beneficiaries and control households, comparing means and quartiles for beneficiaries and non-beneficiaries, while intricate, could provide some useful insights, given that former group is selected non-randomly and the latter is selected randomly both from within the same villages. If spillovers is expected to be negligible in year zero, given the relatively few months that lapsed between exposure of beneficiary households to AR and date of data collection, non-beneficiary households could potentially serve as a with-in village control, since it is highly likely for this group to not have been contaminated. In this case, observed differences between AR and "non-beneficiary control" households can be explained by either project effect or within village targeting of better-off households (or both) and not of village targeting. If both spillovers and targeting are at play, on the other hand, using non-beneficiary households to build the missing counterfactual

[10] Note that pairwise differences in estimated quantiles will not correspond with the quantiles of the estimated differences unless the rank-preservation assumption holds (Cattaneo et al., 2013). This condition will hold if the value of the outcome variable for unit i that corresponds with the q^{th} percentile when all units receive treatment τ also correspond with the q^{th} percentile when all units receive a different treatment τ'.

for beneficiaries would bias our estimates (upwards due to targeting of better-off households and downwards due to the use of households who indirectly benefited from the project as controls).

A statistically significant differences in, say, means of a variable for beneficiary and control households captures both the predicted effect of the project and targeting of better-off villages and/or households. Similarly, a statistically significant difference in means for non-beneficiary and control households captures potential effects of project targeting (of better-off villages, from which non-beneficiaries are randomly sampled) and potential spillovers. A statistically significant differences in, say, medians for beneficiary and control households tells us the difference between the median of the population potential outcome distribution of the variable if all households become a beneficiary and the median of the population potential outcome distribution if all households are in the control group.

Given the one wave of cross-section data used in this study, we caution the reader to interpret our "treatment effects" as a combined predicted effects of the innovations and project targeting. Pair-wise differences of means and quartiles are reported in Table 4. Differences of estimated means of value of harvest are significant only when we compare beneficiary and control households, worth about 43, 000 MWK. The difference in harvest value between beneficiary and control households is also found to be significant across all quartiles, with the highest difference observed for the median and third quartile. This implies potentially weaker effect of the project for households at the left tail of distribution.

For maize yield, we also find a statistically significant expected average and quantile effects when comparing beneficiary and control households. We find an expected average treatment effect of about 0.8 tonnes per hectare and quintile effects ranging from about 0.3 (first quartile) to 1.1 (third quartile), once again suggesting a relatively smaller effect on households at the left tail of the distribution. We also find a significant difference when comparing non-beneficiary households (in project target villages) with control households, which maybe capturing systematic targeting of areas with higher maize yield potential.

Comparison of outcomes for AR and control households using propensity score matching (PSM) shows expected average treatment effects on value of harvested and maize yield that ranges between 42, 000 and 49, 000 MWK (Table 5, Columns 1-2) and 0.83 and 0.88 tonnes per hectare (Table 5, Columns 7-8), respectively, depending on the estimator.[11] These average treatment effects are reasonably comparable with those from EIF. Comparison of harvest value for AR beneficiary and non-beneficiary households also show a statistically significant (at the 5% level) ATE. PSM estimates of ATE effects on maize yield for the other two pair-wise comparisons (Table 5, Columns 9-12) are all significant, especially when comparing non-beneficiary household (in AR target villages) with control households.

6. Conclusion

Given the projected increase in the world's population and the corresponding increase in the demand for food, feed, and biofuel sources, efforts to ensure food security may need to expand beyond discrete, embodied technologies and their package approaches to delivery. This entails focusing on a more integrated—and arguably more complex—approaches to simultaneously improve yields and manage natural resources more sustainably. With such complexity come challenges in accurately measuring and learning about the adoption and effects of these approaches to help inform scaling up efforts as well as designing and targeting of subsequent interventions.

In this paper, we examine the targeting and expected effects of a systems-based sustainable intensification (SI) project in Malawi. This project aims to identify SI best-bet options that fit the needs of resource-poor farmers and hypothesizes that households will be in a better position to innovate, adapt, and adopt when empirical knowledge about SI is translated into action through participatory action research. To provide evidence on the causal effects of the project and potential spillovers, we have employed a quasi-experimental design involving non-randomly selected AR beneficiary households, randomly selected non-beneficiary households (in project target villages), and randomly selected control households (in non-AR villages). We have complemented our evaluation design with detailed and fine-grained geographic information system to better characterize the study area.

Results suggest that AR beneficiary households show significantly different socioeconomic characteristics than randomly selected households, suggesting that systematic targeting plays an important role. Regression results that control for pre-treatment characteristics show higher value of harvest and maize yield both on average and across all quantiles of the potential outcome distributions for project beneficiaries, relative to control households. Results also

[11] Propensity score histogram in Figure 9 generally confirm balanced propensity scores for each pair-wise comparison.

suggest a higher maize yield for non-beneficiary households (in project target villages) compared with control households. This finding could suggest potential systematic targeting of villages with high maize yield potential, especially considering the relatively short time lapse between exposure of AR households to SI innovations and the time of data collection that potentially renders spillovers (from AR households to non-beneficiary households in the same village) negligible. Our findings highlight the need to rethink the criteria for selection of farmers into systems-based innovations, something that could potentially bear severe implications upon scaling up.

While we are unable to disentangle the different effects that are at play and provide robust evidence based on a wave of cross-sectional data, the analyses presented here provided crucial insights about the project and highlighted the need for inventive identification techniques to discern direct project effects from that of targeting and spillovers. We hope to be able to address some of the limitations of this paper using end-line data to be collected from the same households at the end of the first phase of the project.

References

Abadie, A. and Imbens, G., 2012. Matching on the estimated propensity score.

Adger, W., Brown, K., Nelson, D., Berkes, F., Eakin, H., Folke, C., Galvin, K., Gunderson, L., Goulden, M., O'Brien, K., Ruitenbeek, J. And Tompkins, E., 2011. Resilience implications of policy responses to climate change. *Wiley Interdisciplinary Reviews: Climate Change*, 2(5), 757-766.

Alston, J., Chan-Kang, C., Marra, M., Pardey, P. and Wyatt, T., 2000. A meta-analysis of the rates of return to agricultural R&D: Ex pede Hercules? IFPRI Research Report No. 113. Washington, DC: International Food Policy Research Institute.

Angelucci, M. and Di Maro, V., 2010. Program evaluation and spillover effects, impact-evaluation guidelines, Inter-American Development Bank Technical Notes No. IDB-TN-136.

Asfaw, S., Kassie, M., Simtowe, F. and Lipper, L., 2012. Poverty reduction effects of agricultural technology adoption: a micro-evidence from rural Tanzania. *Journal of Development Studies*, 48(9), 1288–1305.

Banerjee, A. and Duflo, E., 2009. The experimental approach to development economics. *Annual Review of Economics*, 1, 151-178.

Barret, C. and Carter, M., 2010. The power and pitfalls of experiments in development economics: some non-random reflections. *Applied Economic Perspectives and Policy*, 32(4), 515–548.

Bioversity. 2014. Legume Intercropping in Malawi.

Busso, M., DiNardo, J. and McCrary, J., 2013. New evidence on the finite sample properties of propensity score reweighting and matching estimators.

Beddington, J., Asaduzzaman, M., Clark, M., Bremauntz, A., Guillou, M., Jahn, M., Lin, E., Mamo, T., Negra, C., Nobre, C., Scholes, R., Sharma, R., Van Bo, N. and Wakhungu, J., 2012. The role for scientists in tackling food insecurity and climate change. *Agriculture and Food Security*, 1(10), 1-9.

Bennett, M. and Franzel., 2013. Can organic and resource-conserving agriculture improve livelihoods? A synthesis. *International Journal of Agricultural Sustainability*, 11(3), 193–215. 193-215.

Burtless, G., 1995. "The case for randomized field trials in economic and policy research." *The Journal of Economic Perspectives*, 9(2), 63-84.

Cash, D., Clark,W., Alcock, F., Dickson, N., Eckley, N., Guston, D., Jager, J. and Mitchell, R. , 2003. Knowledge systems for sustainable development. *Proceedings of the National Academy of Sciences of the United States of America*, 100(14), 8086-8091.

Cattaneo, M., 2010. Efficient Semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics*, 155, 138-154.

Cattaneo, M., Drukker, D. and Holland, A., 2013. Estimation of multivalued treatment effects under conditional independence. *The Stata Journal*, 13(3), 407-450.

Deaton, A., 2009. Instruments of development: randomization in the tropics, and the search for the elusive keys to economic development. Research Program in Development Studies Center for Health and Wellbeing Princeton University.

Devereux, S., 2008. Targeting social safety nets to support agricultural growth in rural areas, In: Convergence between Social Service Provision (SSP) and Productivity Enhancing Investments (PEI) in Development Strategies, University of KwaZulu-Natal, Pietermaritzburg, South Africa.

Devereux, S., Al-Hassan, R., Dorward, A., Guenther, B., Poulton, C. and Sabates-Wheeler, R., 2008. Linking social protection and support to small farmer development. Paper Commissioned by the Food and Agricultural Organization of the United Nations (FAO), Rome.

Duflo, E., Glennerster, R. and Kremer, M., 2008. "Using randomization in development economics research: a toolkit." *Handbook of Development Economics*, 4 (Chapter 61), 3896-3962.

Erenstein, O., 2010. Zero Tillage in the rice–wheat systems of the Indo-Gangetic plains: a review of impacts and sustainability implications, in Spielman, D. and Pandya-Lorch, R. eds., *Proven Successes in Agricultural Development: A Technical Compendium to Millions Fed.* Washington, DC: IFPRI.

Evenson, R. and Gollin, D., 2003. Assessing the impact of the green revolution, 1960 to 2000. *Science*, 300, 758-762.

Fan, M., Shen, J., Yuan, L., Jiang, R., Chen, X., Davies, W. and Zhang, F., 2011. Improving crop productivity and resource use efficiency to ensure food security and environmental quality in China. *Journal of Experimental Botany*, 1-12.

Fan, S., 2000. Research investment and the economic returns to Chinese agricultural research. *Journal of Productivity Analysis*, 14, 163-182.

Fan, S. and Pardey, P., 1997. Research, productivity, and output growth in Chinese agriculture. *Journal of Development Economics*, 53,115-137.

Fan, S., Hazell, P. and Thorat, S., 2000. Government spending, growth and poverty in rural India. *American Journal of Agricultural Economics,* 82(4), 1038-1051.

Feder, G., Murgai, R. and Quizon, J., 2004. The Acquisition and diffusion of knowledge: The case of pest management training in farmer field schools, Indonesia. *Journal of Agricultural Economics*, 55(2), 221-243.

Friesen, D. and Palmer, A., 2004. Integrated approaches to higher maize productivity in the new millennium. 7. Proceedings of the Eastern and Southern Africa Regional Maize Conference; Nairobi, Kenya.

Folke, C. , 2006. Resilience: The emergence of a perspective for social-ecological systems analysis. *Global Environmental Change*, 16, 253-267.

Franzel, S. and Scherr, S., 2002. *Trees on the Farm: Assessing the Adoption Potential of Agroforestry Practices in Africa.* Wallingford, U.K.: CABI Publishing.

Friesen, D and Palmer, A. , eds., 2004. Integrated approaches to higher maize productivity in the new millennium, proceedings of the seventh eastern and southern Africa regional maize conference Nairobi, Kenya 5 – 11.

Garnett, T., Appleby, M., Balmford, A., Bateman, I., Benton,T., Bloomer, P., Burlingame, B., Dawkins, M., Dolan, L., Fraser, D., Herrero, M., Hoffmann, I., Smith, P., Thorton, P., Toulmin, C., Vermeulen, S. and Godfray, H., 2013. Sustainable intensification in agriculture: Premises and policies. *Science*. 341, 33-34.

Giller, K., Corbeels, M., Nyamangara, J., Triomphe, B., Affholder, F., Scopel, E. and Tittonell, P., 2011. A research agenda to explore the role of conservation agriculture in African smallholder farming systems. *Field Crops Research*, 124, 468–472.

Giller, K., Witter, E., Corbeels, M. and Tittonell, P., 2009. Conservation agriculture and smallholder farming in Africa: the heretics' View. *Field Crops Research*, 114, 23–34.

Gilligan, D., Hoddinott, J. and Taffesse, A., 2009. The impact of Ethiopia's productive safety net programme and its linkages. *Journal of Development Studies,* 45(10), 1684-1706.

Hall, A. and Clark, N., 2010. What do complex adaptive systems look like and what are the implications for policy? *Journal of International Development*, 22, 308-324.

Holling, C., 1973. Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics,* 4, 1-23.

IEG (Independent Evaluation Group)., 2011. Impact evaluations in agriculture: an assessment of the evidence. Washington, DC: World Bank.

ILRI (International Livestock Research Institute)., 2012. Africa Research in Sustainable Intensification for the Next Generation (Africa RISING) Program Framework 2012 - 2016.

Imbens, G., 2002. The Role of the propensity score in estimating dose-response functions, *Biometrika*, 87(3), 706-710.

Jayne, T., Mason, N., Myers, R., Ferris, J., Mather, D., Beaver, M., Lenski, N., Chapoto, A. and Boughton, D., 2010. Patterns and trends in food staples markets in eastern and southern Africa: toward the identification of priority investments and strategies for developing markets and promoting smallholder productivity growth. Michigan State University International Development Working Paper No. 104.

Johnson, N., Lilja, N. and Ashby, J., 2003. Measuring the impact of user participation in agricultural and natural resource management research. *Agricultural Systems*, 78, 287–306.

Khan, S., and Tamer, E., 2010. Irregular identification, support conditions, and inverse weight estimation. *Econometrica*, 78(6), 2021–2042.

Khan, Z., Midega, C., Pittchar, J., Murage, A., Birkett, M., Bruce, T. and Pickett, J., 2014. Achieving food security for one million sub-Saharan African poor through push–pull innovation by 2020. *Philosophical Transactions of the Royal Society*, 369, 20120284.

Kamanula, J., Sileshi, G., Belmain, S., Sola, P., Mvumi, B., Nyirenda, G., Nyirenda, S. and Stevenson, P., 2011. Farmers' insect pest management practices and pesticidal plant use in the protection of stored maize and beans in Southern Africa. *International Journal of Pest Management*, 57(1), 41-49.

Kato, E., Nkonya, E. and Place, K., 2011. Heterogeneous treatment effects of integrated soil fertility management on crop productivity: Evidence from Nigeria, IFPRI Discussion Paper 01089.

Kerr, R., Snapp, S., Chirwa, M., Shumba, L. and Msachi, R., 2007. Participatory research on legume diversification with Malawian smallholder farmers for improved human nutrition and soil fertility. *Experimental Agriculture*, 43, 437–453.

Leamer, E., 2010. Tantalus on the road to asymptopia. *Journal of Economic Perspectives*, 24(2), 31–46.

Levin, S., Barrett, S., Aniyar, S., Baumol, W., Bliss, C., Bolin, B., Dasgupta, P., Ehrlich, P., Folke, C., Gren., I., Holling, C., Jansson, A., Jansson, B., Mäler, K., Martin, D., Perrings, C. and Sheshinski, E., 1998. Resilience in natural and socioeconomic systems. *Environment and Development Economics*, null (2), 221-262.

Lucas, H. and Longhurst, R. , 2010. Evaluation: Why, for whom and how? *IDS Bulletin*, 41(6), 28-35.

McDonald, A., Hobbs, P. and Riha, S., 2008. Stubborn facts: Still no evidence that the system of rice intensification out-yields Best Management Practices (BMPs) beyond Madagascar. *Field Crops Research*, 108, 188-191.

McDonald, A., Hobbs, P. and Riha, S., 2006. Does the system of rice intensification outperform conventional best management? A synopsis of the empirical record. *Field Crops Research*, 96, 31-36.

Minde, I., Jayne, T., Crawford, E., Arega, J. and Govereh, J. 2008. Promoting fertilizer use in Africa: current issues and empirical evidence from Malawi, Zambia, and Kenya. ReSAKSS Working Paper No. 13.

Mhango, W., 2011. Nitrogen budgets in legume based cropping systems in northern Malawi, PhD Thesis. Michigan State University.

Nelson, D., Adger, W. and Brown, K., 2007. Adaptation to environmental change: contributions of a resilience framework. *Annual Review of Environmental Resources*, 32, 395-419.

New Partnership for African Development (NEPAD)., 2013. African agriculture, transformation and outlook.

Orr, A. and Ritchie, J., 2004. Learning from failure: smallholder farming systems and IPM in Malawi. *Agricultural Systems*, 79, 31–54.

Pretty, J., Toulmin, C. and Williams, S., 2011. Sustainable intensification in African agriculture. *International Journal of Agricultural Sustainability*, 9(1), 5-24.

Raitzer, D. and Kelley, T., 2008. Benefit-cost meta-analysis of investment in the international agricultural research centres of the CGIAR. *Agricultural Systems*, 96, 108-123.

Renkow, M. and Byerlee, D., 2010. The impacts of CGIAR research: A review of recent evidence. *Food Policy*, 35(5), 391-402.

Rogers, E. , 2003. *Diffusion of innovations*. 5th Edition. New York: Free Press.

Rosenbaum, P. and Rubin, D., 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.

Rubin, D., 1978. Bayesian inference for causal effects: The role of randomization. *Annuals of Statistics*, 6, 34-58.

Rubin, D., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66, 688-701.

Senthilkumar, K., Bindraban, P., Thiyagarajan, T., de Ridder, N. and Giller, K., 2008. Modified rice cultivation in Tamil Nadu, India: Yield gains and farmers' (lack of) acceptance. *Agricultural System*, 98, 82–94.

Shiferaw, B., Kebede, T. and You, L., 2008. Technology adoption under seed access constraints and the economic impacts of improved pigeon pea varieties in Tanzania. *Agricultural Economics*, 39, 309-323.

Sumberg, J. and Thompson, J., 2012. *Contested agronomy: Agricultural research in a changing world.* Oxon, UK: Routledge.

Sumberg, J., Thompson, J. and Woodhouse, P. , 2012. Why agronomy in the developing world has become contentious. *Agriculture and Human Value*, 30, 71-83.

Tan, Z., 2010. Bounded, efficient and doubly robust estimation with inverse weighting. *Biometrika*, 97(3), 661–682.

The Montpellier Panel., 2013. Sustainable intensification: A new paradigm for African agriculture, London.

Thierfelder, C., Cheesman, S. and Rusinamhodzi, L., 2012. A comparative analysis of conservation agriculture systems: Benefits and challenges of rotations and inter-cropping in Zimbabwe. *Field Crops Research*, 137, 237–250.

Tilman, D., Balzer, C., Hill, J. and Befort, B., 2011. Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, 108(50), 20260–20264.

Tittonell, P., 2013. Livelihood strategies, resilience and transformability in African agroecosystems. *Agricultural Systems*, 126, 3-14.

Uphoff, N., Kassam, A. and Stoop, W., 2008. A critical assessment of a desk study comparing crop production systems: The example of the ‘System of Rice Intensification versus ‘Best Management Practice.’ *Field Crops Research*, 108, 109-114.

Uphoff, N., Kassam, A. and Hardwood, R., 2010. SRI as a methodology for raising crop and water productivity: Productive adaptations in rice agronomy and irrigation water management. *Paddy Water Environment*, 9, 3–11.

Van den Berg, H. and Jiggins, J., 2007. Investing in farmers – the impacts of farmer field schools in relation to integrated pest management. *World Development*, 35(4), 663-686.

Van Ginkel, M., Sayer, J., Sinclair, F., Aw-Hassan, A., Bossio, D., Craufurd, P., El Mourid, M., Haddad, N., Hoisington, D., Johnson, N., Velarde, C., Mares,V., Mude, A., Nefzaoui, A., Noble, A., Rao, K., Serraj, R., Tarawali, S., Vodouhe, R. and Ortiz, R., 2013. An integrated agro-ecosystem and livelihood system approach for the poor and vulnerable in dry areas. *Food Security*. 5, 751-767.

Zhao, Z., 2004. Using matching to estimate treatment effects: data requirements, matching metrics, and Monte Carlo evidence. *The Review of Economics and Statistics*, 86(1), 91-107.

Tables

Table 1. Sample size (by treatment status)

					Number of households by treatment status			
District	EPA	Section	Type	No. of	Beneficiary	Non-	Control	All
Dedza	Golomoti	Golomoti Centre	AR target	4	100	24	0	124
Dedza	Linthipe	Mposa	AR target	5	71	81	0	152
Dedza	Lobi	Thete	Non-target	7	0	0	141	141
Dedza	Mtakataka	Mtakataka Center	Non-target	8	0	0	161	161
Ntcheu	Kandeu	Kampanje	AR target	8	135	60	0	195
Ntcheu	Nsipe	Mpamadzi	AR target	9	92	42	0	134
Ntcheu	Nsipe	Mwalaoyera	Non-target	9	0	0	164	164
Ntcheu	Kandeu	Sitolo	Non-target	4	0	0	78	78
		Total		54	398	207	544	1,149

Source. MARBES (2013).

Table 2. Village-level summary of selected variables (by treatment status) (N=54)

Variable	Treatment status: Beneficiary	Treatment status: Control
Length of growing period (Days)	162	161
Slope (Degrees)	1.35	1
Travel time to the nearest town of 20 thousand people	235**	182**
Population density (Number per square kilometre)	198***	332***
Historical annual average rainfall (Millimetres)	937***	919***
Historical annual average temperature (Celsius)	21	20
Elevation (Meters)	897	947
% of villages within tropic warm (semiarid) agro	0.85	0.75
% of villages in medium rainfall-medium elevation stratum	0.65	0.46
% of villages in high rainfall-high elevation stratum	0.19	0.25
Access to basic services index	0.08	-0.04
% of villages w/ extension services	1.0***	0.7***
% of villages w/ farmers cooperatives groups	0.85***	0.46***
% of villages w/ access to improved maize seed	0.46	0.43

Note. * Significant at 10%; ** significant at 5%; *** significant at 1%.
Source. MARBES (2013) and other sources summarized in Table A2 in the Appendix.

Table 3. Household-level summary of selected variables (by treatment status)

Variable	Beneficiary	Non-beneficiary	Control	1 vs 2	1 vs 3	2 vs 3
	(1)	(2)	(3)	(4)	(5)	(6)
Household size	4.97	4.55	4.59	**	***	
Average adult years of education	5.22	4.51	4.72	***	***	
Dependency ratio	1.23	1.28	1.29			
Age of household head(years)	45.7	46.1	45.3			
% with married head	0.75	0.60	0.65	***	***	
% with female head	0.27	0.36	0.34	**	**	
% in the lowest agricultural wealth quintile	0.1	0.3	0.2	***	***	
% in the lowest non-agricultural wealth quintile	0.1	0.2	0.3	***	***	
Tropical livestock units	0.46	0.28	0.21	***	***	**
Per capita land operated(ha)	0.28	0.23	0.23	**	***	
Distance to basic services index	0.0054	0.027	-0.013			
% w/ closest parcel <15 minutes	0.74	0.57	0.54	***	***	
Travel time to seed supplier (Minute)	43.6	41.7	38.9			
Chemical fertilizers used (kg)	119.2	83.8	80.2	***	***	
Agricultural labor used (person-days)	332.4	241.1	228.0	***	***	
% using improved maize variety	0.87	0.63	0.62	***	***	
% using hired labor	0.50	0.31	0.39	***	***	**
% practicing intercropping	0.82	0.86	0.68		***	***
% using irrigation during dry season	0.15	0.11	0.061		***	**
% affected by soil erosion	0.70	0.60	0.60	**	***	
% with access to extension services	0.9	0.4	0.3	***	***	
% using manure	0.68	0.56	0.43	***	***	***
Value of harvest ('000 MWK)	205	150	125	***	***	
Maize yield (ton/ha)	2.4	2.3	1.8		***	***
Observations	405	199	538	604	943	737

Note. Columns 4, 5, and 6 report significance levels from diffrences of means tests between AR beneficiary (Column 1) and non-beneficiary (Column 2) households; AR beneficiary and control (Column 3) households, and non-beneificary and control households, respectively.
* Significant at 10%; ** significant at 5%; *** significant at 1%.
Source. MARBES (2013).

Table 4. Efficient-influence function (EIF) estimates of average and quantile treatment effects

Pairwise comparison	Value of harvest ('000 MWK)			Maize yield (ton/ha)		
	Pairwise contrast	Unadjusted [95% Conf. Interval]		Pairwise contrast	Unadjusted [95% Conf. Interval]	
	Panel A. Average treatment effects					
Beneficiary versus Control	43.24	12.73	73.74	0.83	0.52	1.13
Beneficiary versus Non-beneficiary	28.35	-13.54	70.25	0.39	0.03	0.74
Non-beneficiary versus Control	14.88	-25.52	55.29	0.44	0.16	0.71
	Panel B. Quantile treatment effects (1st quartile)					
Beneficiary versus Control	37.78	25.64	49.91	0.33	0.22	0.44
Beneficiary versus Non-beneficiary	10.61	-6.50	27.73	0.08	-0.08	0.25
Non-beneficiary versus Control	27.16	12.23	42.10	0.24	0.09	0.40
	Panel C. Quantile treatment effects (Median)					
Beneficiary versus Control	45.92	29.08	62.76	0.53	0.35	0.72
Beneficiary versus Non-beneficiary	20.53	-2.98	44.05	0.16	-0.12	0.43
Non-beneficiary versus Control	25.39	1.77	49.01	0.38	0.12	0.63
	Panel D. Quantile treatment effects (3rd quartile)					
Beneficiary versus Control	45.60	19.33	71.87	1.14	0.76	1.53
Beneficiary versus Non-beneficiary	42.80	6.33	79.26	0.40	-0.05	0.85
Non-beneficiary versus Control	2.80	-37.42	43.02	0.75	0.34	1.15

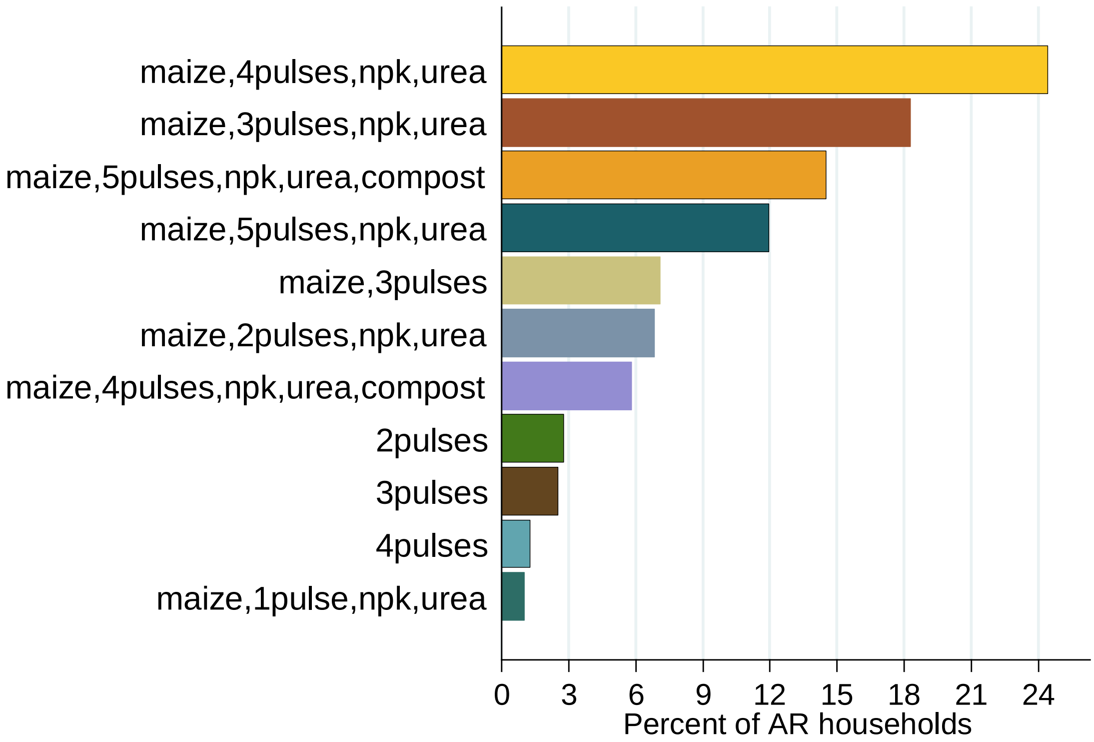
Table 5. Propensity score matching estimates of average treatment effects (ATE)

	Value of harvest ('000 MWK)						Maize yield (ton/ha)					
	Beneficiary versus Control		Beneficiary versus Non-beneficiary		Non-beneficiary versus Control		Beneficiary versus Control		Beneficiary versus Non-beneficiary		Non-beneficiary versus Control	
	Kernel	NN (K=5)	Kernel	NN (K=5)	Kernel	NN (K=5)	Kernel	NN (K=5)	Kernel	NN (K=5)	Kernel	NN (K=5)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
ATE	49.25***	42.36***	35.22**	38.18**	14.62	13.03	0.836***	0.882***	0.256*	0.323**	0.639***	0.696***
	(15.45)	(15.66)	(17.51)	(16.56)	(20.34)	(20.52)	(0.144)	(0.142)	(0.146)	(0.140)	(0.187)	(0.183)
Constant	151.0***	157.9***	169.7***	166.7***	140.7***	142.3***	1.431***	1.384***	2.028***	1.960***	1.562***	1.505***
	(10.92)	(11.07)	(12.38)	(11.71)	(14.38)	(14.51)	(0.102)	(0.100)	(0.103)	(0.0989)	(0.132)	(0.129)
Observations	741	742	770	770	409	409	732	731	764	764	410	409
R-squared	0.014	0.010	0.005	0.007	0.001	0.001	0.044	0.050	0.004	0.007	0.028	0.034

Note. * Significant at 10%; ** significant at 5%; *** significant at 1%. NN (K=5) stands for nearest neighbour matching with K=5. Kernel estimator is based on Epanechnikov kernel. Reported in parenthesis are bootstrapped standard errors (50 repetitions).

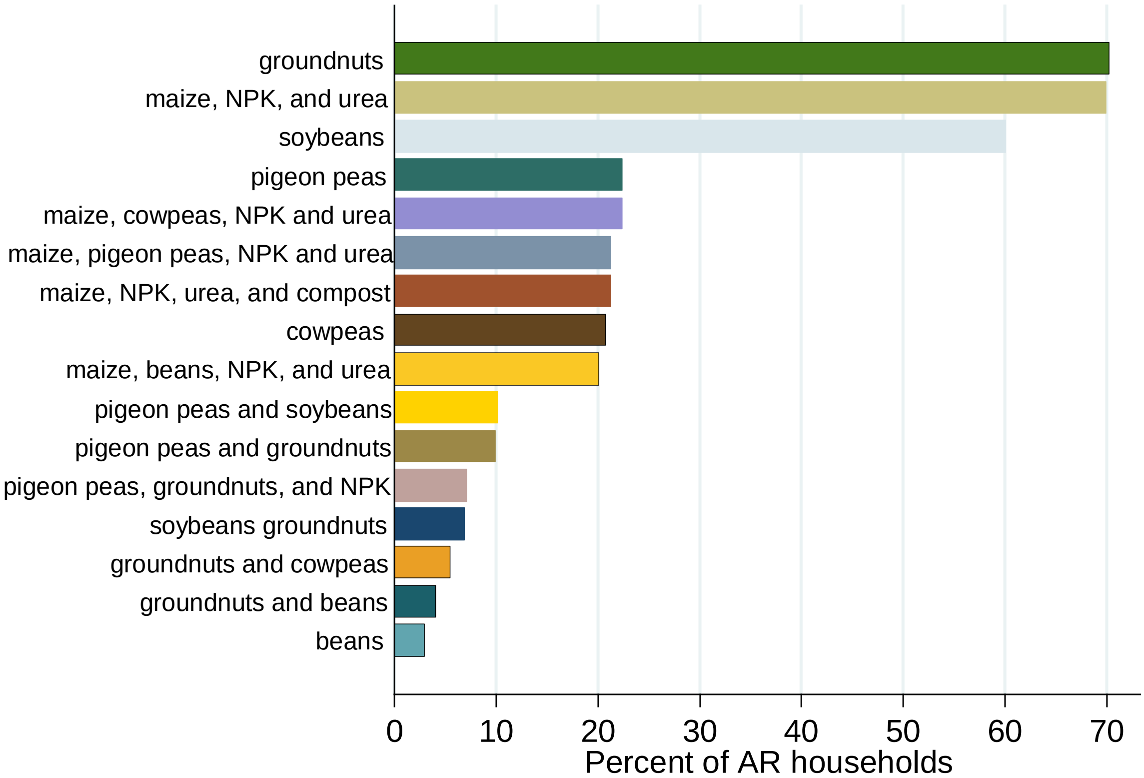
Figures

Figure 1. Technologies applied by AR households at the farm level (N=393)



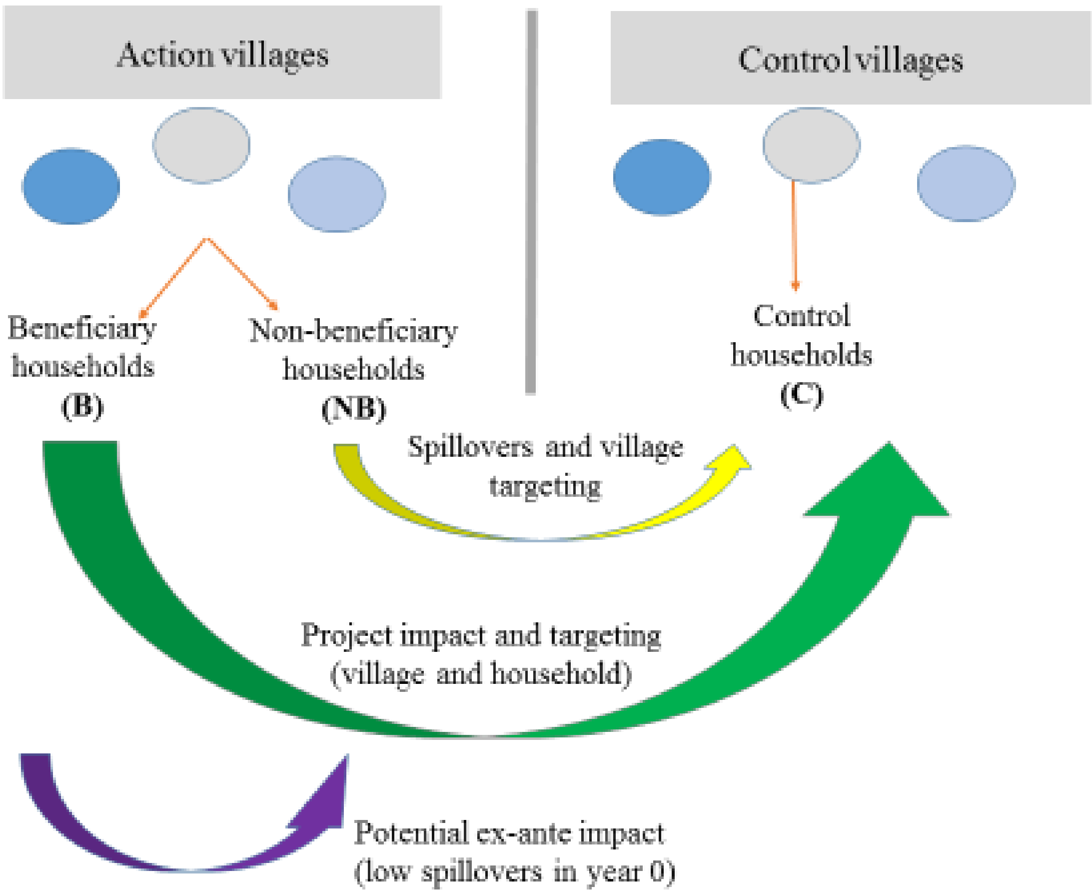
Note: Technology mixes adopted by less than 1% of AR households not shown
Source: Africa RISING Malawi

Figure 2. Technologies applied by AR households in at least one single plot (N=393)



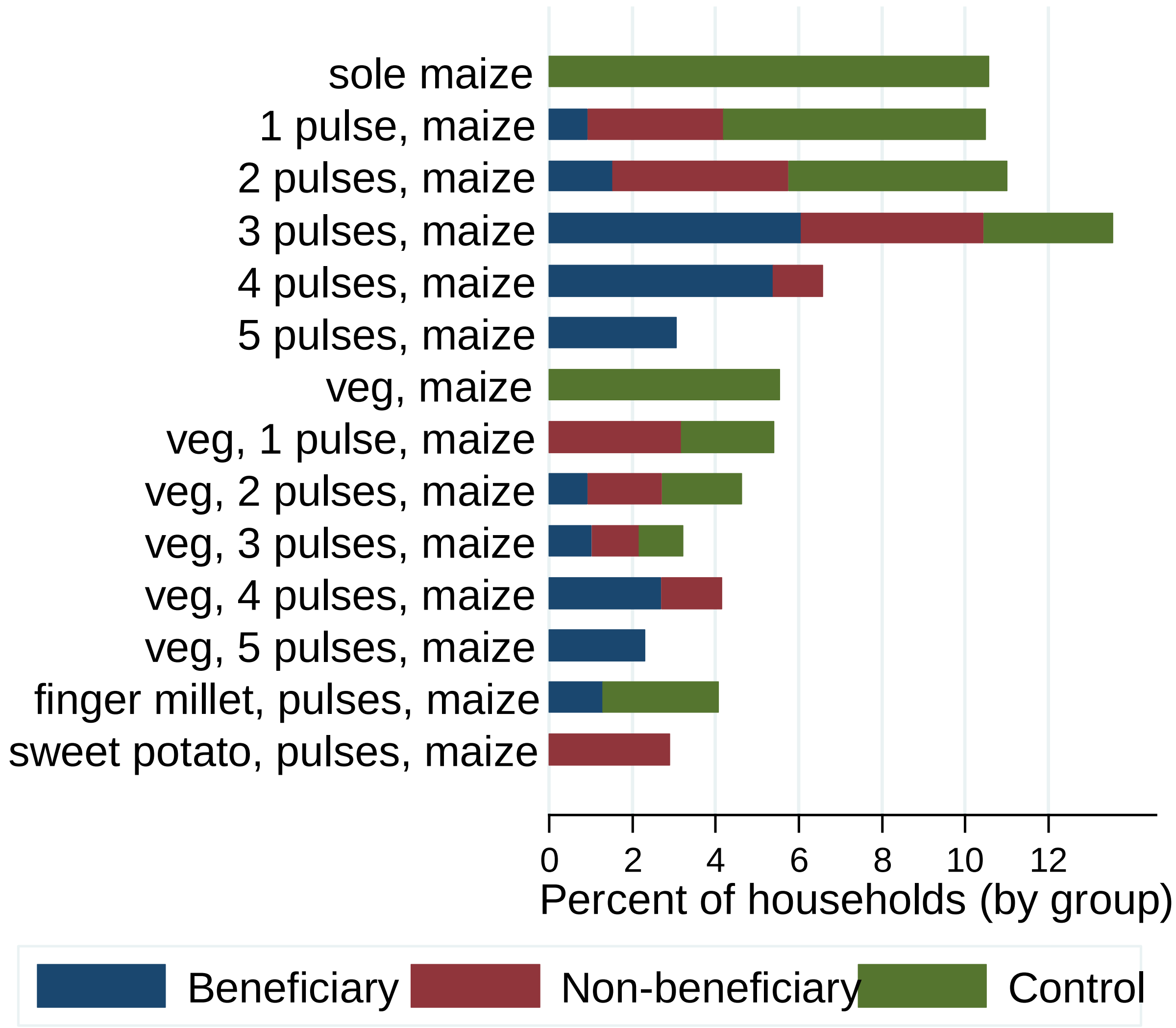
Note: Technology mixes adopted by less than 2% of AR households not shown
Source: Africa RISING Malawi

Figure 3. AR evaluation design †



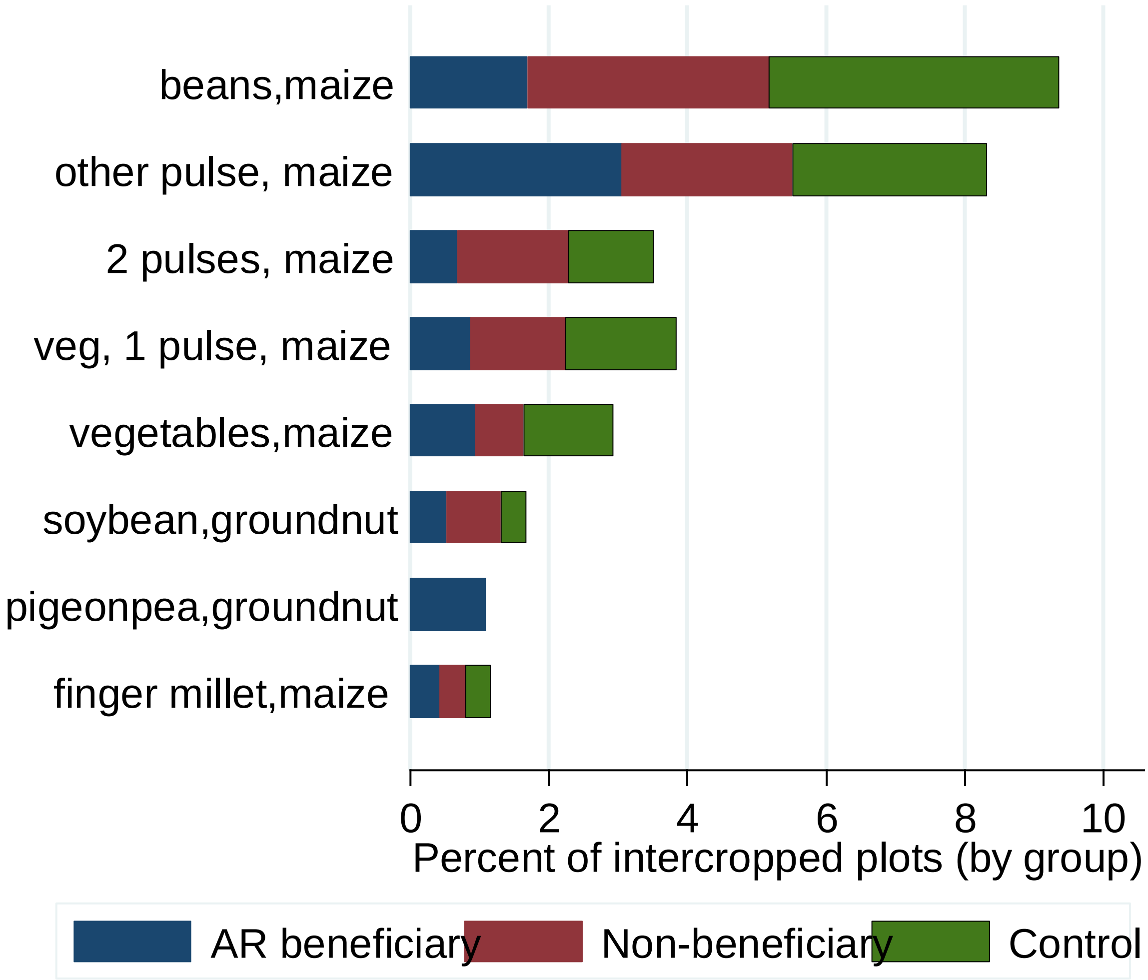
† Different colors denote homogeneous agricultural potential areas.

Figure 4. Household farming systems (by treatment status)



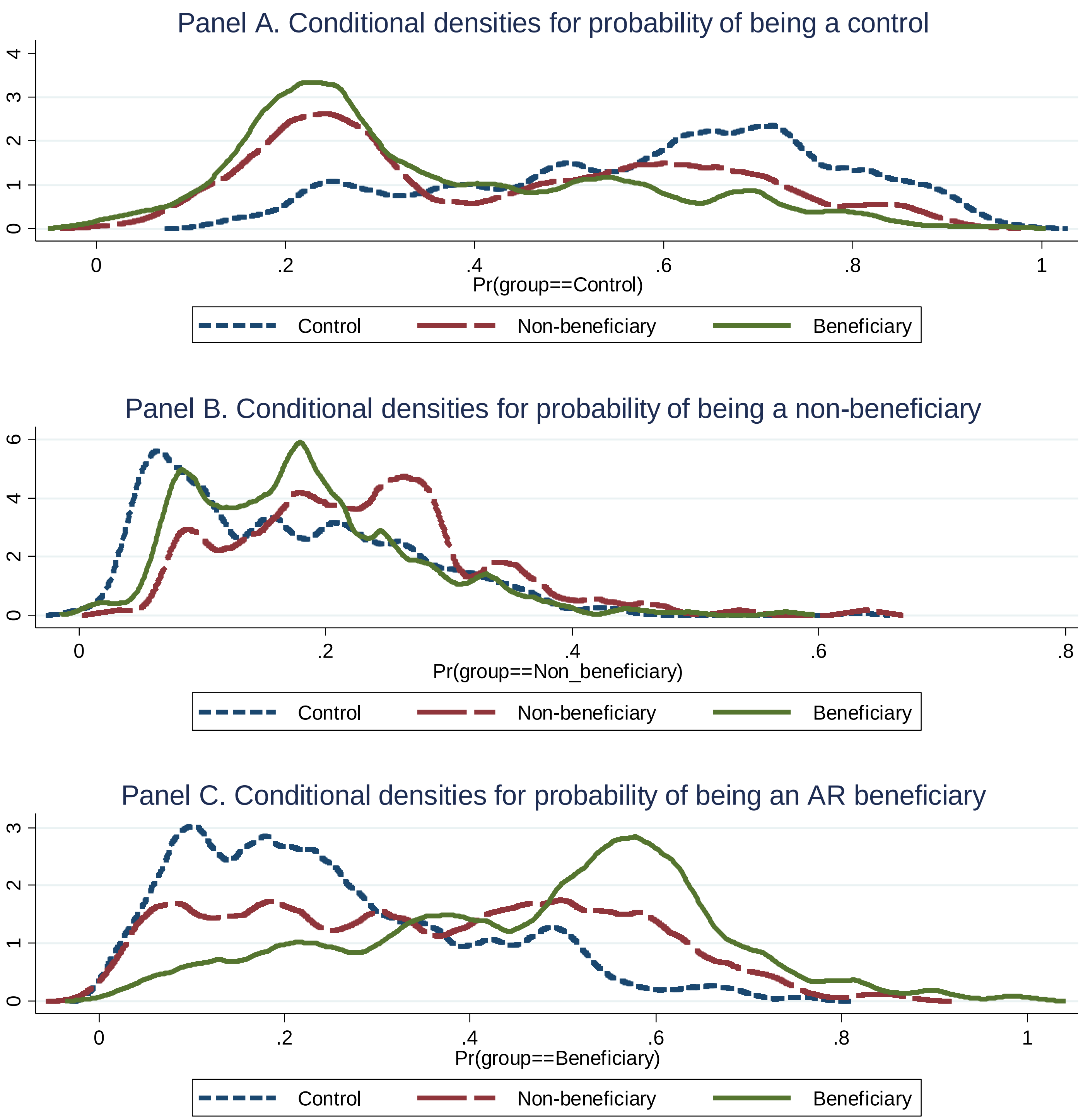
Note: Innovation mixes adopted by less than 2% of households not shown.
Veg = Vegetables
Source: MARBES, 2013

Figure 5. Incidence of Intercropping (by treatment status)



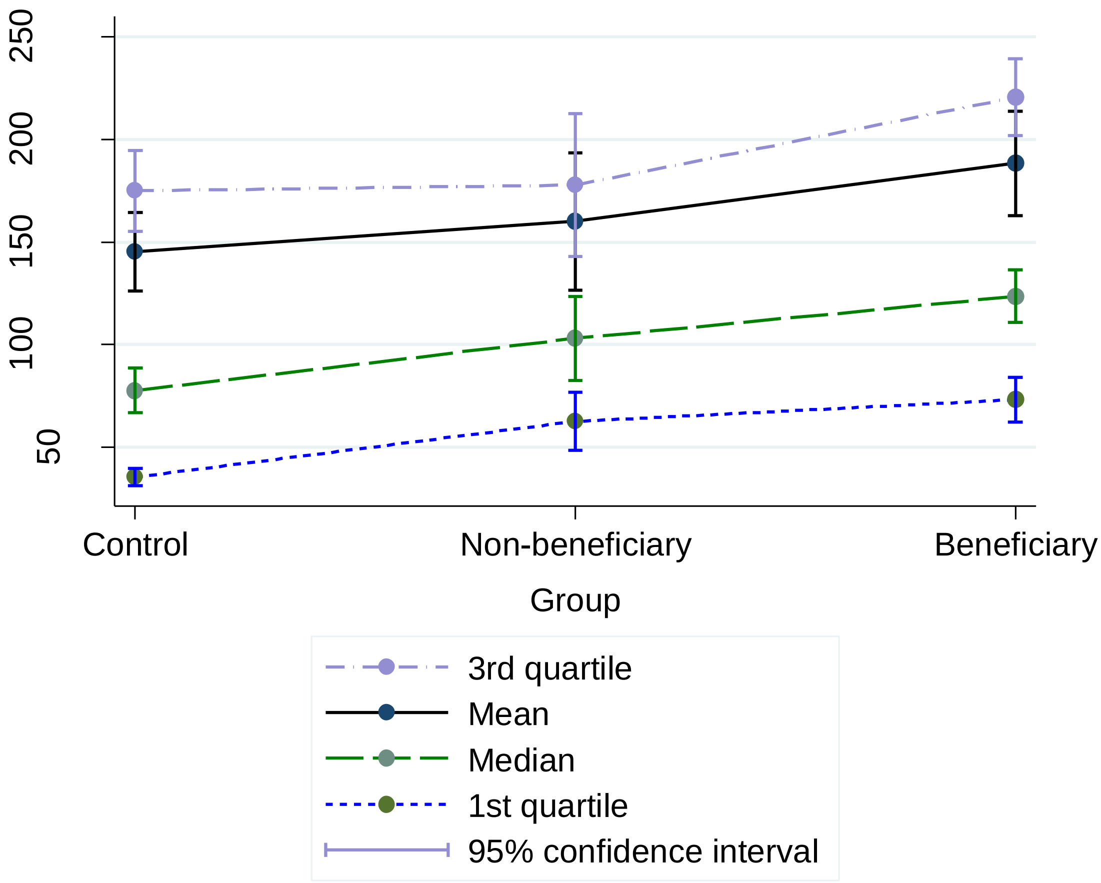
Note: Innovation mixes adopted by less than 1% of households not shown.
Veg = Vegetables
Source: MARBES, 2013

Figure 6. Conditional densities of generalized propensity score estimates[12]



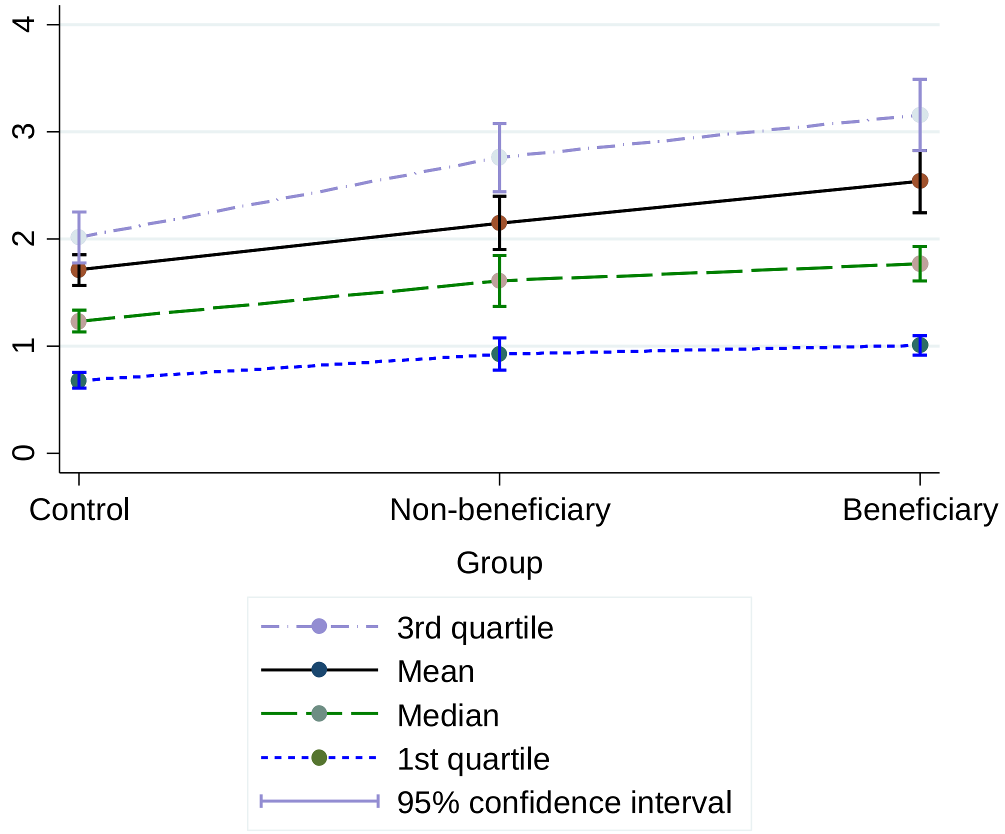
[12] Note: The visibility of the density function in Figure 6 diminishes significantly when the document is transformed into a portable document format (PDF). Graphs of density functions in non-PDF formats are available upon request.

Figure 7. Efficient-influence function (EIF) estimates of means and quartiles of harvest value



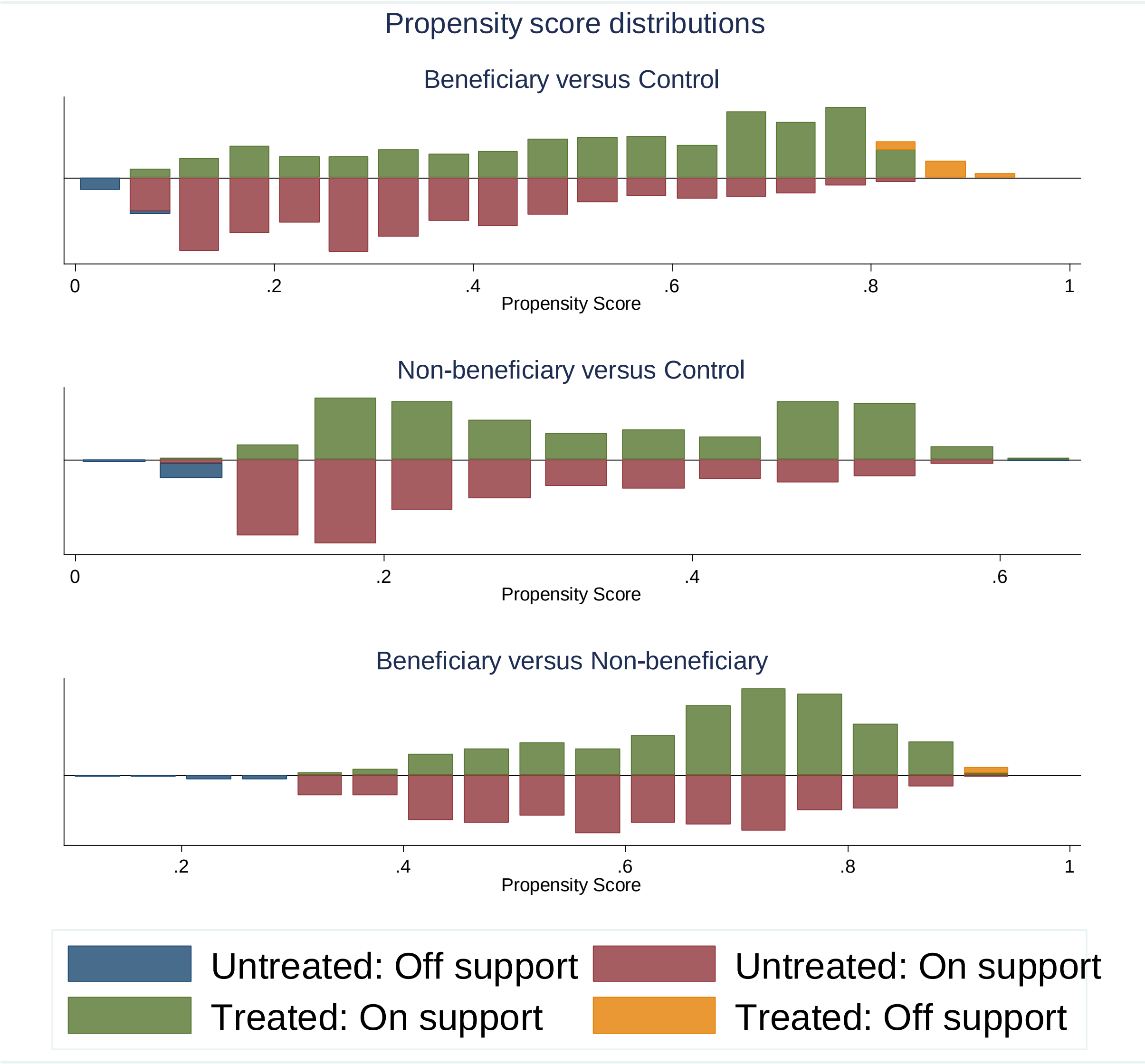
Note: Confidence intervals are based on bootstrapped standard errors (2000 repetitions).

Figure 8. Efficient-influence function (EIF) estimates of means and quartiles of maize yield



Note: Confidence intervals are based on bootstrapped standard errors (2000 repetitions).

Figure 9. Conditional densities of propensity score estimates



APPENDIX

Tables

Table A1. AR Malawi innovations (2012/13 cropping season)

District	EPA	Innovations				
		Maize	Cowpea	Pigeon pea	Groundnut	Soybean
Dedza	Linthipe	PAN 53	Sudan 1	Mwaiwathu alimi	CG7	Makwacha
		DKC 9089	IT82E-16	Sauma	Nsinjiro	Nasoko
	Golomoti	SC403	Nkanakaufiti	Sauma	Nsinjiro	Nasoko
		DKC 8053	Nkanakaufiti	Mwaiwathu alimi	CG7	Makwacha
Ntcheu	Kandeu	SC627	Sudan 1	Mwaiwathu alimi	Nsinjiro	Makwacha
		DKC 8053	IT82E-16	Sauma	CG7	Nasoko
	Nsipe	SC627	Sudan 1	Mwaiwathu alimi	CG7	Makwacha
		DKC 8053	IT82E-16	Mwaiwathu alimi	Nsinjiro	Nasoko

Source: Africa RISING Malawi.

Table A2. Candidate data layers considered for site characterization

Data layer	Spatial resolution	Year	Source
Population density	1 km^2	2000	CIESIN
Agro-Ecological	~10km^2		IIASA
Precipitation	50 km^2	long term (> 50 years) average	CRU
	1 km^2	long term (> 50 years) average	WorldClim
	100 km^2	long term (> 50 years) average	NASA POWER
	50km^2	long term (> 50 years) average	GPCC
	1km^2	long term (1976-2008) average	Interpolated from national
Elevation	1 km^2		USGS
Slope	1 km^2		USGS
Farming systems	shape file		John Dixon (2012 version)
Market access	1 km^2	2000	HarvestChoice
Length of growth	~10km^2	long term (> 50 years) average	IIASA
Maize harvested	~10km^2	2000	HarvestChoice

Table A3. Maximum likelihood estimates from the best fit multinomial logit specification

	Non-beneficiary	Beneficiary
Household size (hhsize)	0.488	0.089
	[0.384]	[0.298]
Age of household head(years) (hh_age)	0.09	0.105*
	[0.055]	[0.048]
Avg. adult yrs of education (educ_ave)	0.296	0.126
	[0.288]	[0.229]
Elevation of HH's residence (meters) (alt)	0.019***	0.018***
	[0.004]	[0.004]
Average travel time to seed supplier (time_2_supplier)	0.012	0.022
	[0.021]	[0.016]
Distance to basic services index (distanceindex)	1.083	1.345
	[0.906]	[0.746]
Total wealth index (wealth)	1.165	2.063**
	[0.952]	[0.699]
c.hhsize#c.hhsize	-0.016	-0.016
	[0.019]	[0.016]
c.hhsize#c.hh_age	0	-0.001
	[0.003]	[0.003]
c.hhsize#c.educ_ave	-0.056**	-0.034*
	[0.021]	[0.017]
c.hhsize#c.alt	0	0
	[0.000]	[0.000]
c.hhsize#c.time_2_supplier	-0.001	0
	[0.001]	[0.001]
c.hhsize#c.distanceindex	0.114*	0
	[0.055]	[0.048]
c.hhsize#c.wealth	-0.038	-0.048
	[0.068]	[0.050]
c.hh_age#c.hh_age	0	-0.001*
	[0.000]	[0.000]
c.hh_age#c.educ_ave	0	0.001
	[0.003]	[0.002]
c.hh_age#c.alt	0	0
	[0.000]	[0.000]
c.hh_age#c.time_2_supplier	-0.000*	0
	[0.000]	[0.000]
c.hh_age#c.distanceindex	0.01	0.019**
	[0.007]	[0.006]
c.hh_age#c.wealth	0.005	0.004
	[0.009]	[0.008]
c.educ_ave#c.educ_ave	-0.001	0.001
	[0.012]	[0.010]

Reported in brackets are standard errors and the control group is the reference category.

Table A3. Maximum likelihood estimates from the best fit multinomial logit specification (Cont'd)

	Non-beneficiary	Beneficiary
	[0.000]	[0.000]
c.educ_ave#c.educ_ave	-0.001	0.001
	[0.012]	[0.010]
c.educ_ave#c.alt	0	0
	[0.000]	[0.000]
c.educ_ave#c.time_2_supplier	0	0
	[0.001]	[0.001]
c.educ_ave#c.distanceindex	0.001	0.06
	[0.047]	[0.040]
c.educ_ave#c.wealth	0.024	-0.073
	[0.046]	[0.039]
c.alt#c.alt	-0.000***	-0.000***
	[0.000]	[0.000]
c.alt#c.time_2_supplier	0	0
	[0.000]	[0.000]
c.alt#c.distanceindex	-0.002***	-0.003***
	[0.001]	[0.001]
c.alt#c.wealth	-0.001**	-0.001**
	[0.001]	[0.000]
c.time_2_supplier#c.time_2_supplier	0	0
	[0.000]	[0.000]
Constant	-12.204***	-8.571**
	[3.470]	[2.648]
Pseudo R-squared	0.15	
Model chi-square	339	
Observations	1134	

Reported in brackets are standard errors and the control group is the reference category.

Table A4. Ordinary least squares estimates from the best fit specifications for the outcome equation

Variable	Value of harvest ('000 MWK)	Maize yield (ton/ha)
Household size (hhsize)	35.720**	0.347**
	[11.725]	[0.110]
Age of household head(years)	0.677	0
	[0.391]	[0.004]
Avg. adult years of education	6.645**	0.055**
	[2.177]	[0.021]
Area of parcels operated(ha)	105.577***	-0.502***
	[9.630]	[0.090]
Agr. wealth (excluding land)	5.405	-0.007
	[5.754]	[0.054]
Total agr. labor used (person-days/ha)	-0.040*	0.001***
	[0.017]	[0.000]
Total fertilizer used(kg/ha)	0.200***	0.003***
	[0.038]	[0.000]
c.hhsize#c.hhsize	-3.119**	-0.035***
	[1.120]	[0.011]
Constant	-103.852**	0.507
	[39.122]	[0.367]
R-squared	0.16	0.24
Observations	1133	1130

Reported in brackets are standard errors

Table A5. Probit estimates of the selection equation (for propensity score matching)

Variable	Beneficiary versus Control	Non-beneficiary versus Control	Beneficiary versus Non-beneficiary
Household size	0.177	0.122	0.031
	[0.097]	[0.132]	[0.135]
Household size^2	-0.020*	-0.012	-0.002
	[0.009]	[0.012]	[0.012]
Age of household head(years)	0.055**	0.025	0.024
	[0.017]	[0.019]	[0.023]
Age of household head(years)^2	-0.001***	0	0
	[0.000]	[0.000]	[0.000]
Avg. adult years of education	0.074	-0.058	0.171
	[0.055]	[0.059]	[0.088]
Avg. adult years of education^2	-0.009	0.002	-0.013
	[0.005]	[0.005]	[0.008]
% with female head	-0.151	-0.011	-0.095
	[0.132]	[0.134]	[0.126]
Average travel time to seed supplier	-0.001	0.002	0.001
	[0.003]	[0.004]	[0.004]
Average travel time to seed supplier^2	0	0	0
	[0.000]	[0.000]	[0.000]
Distance to basic services index	-0.049	-0.115	0.055
	[0.093]	[0.081]	[0.105]
Distance to basic services index^2	-0.062	-0.047	-0.026
	[0.041]	[0.035]	[0.052]
Asset-based wealth index	0.327***	0.056	0.202*
	[0.089]	[0.102]	[0.087]
Asset-based wealth index^2	-0.028	-0.007	-0.019
	[0.030]	[0.040]	[0.030]
Elevation (meters)	0.016***	0.014***	-0.002
	[0.002]	[0.002]	[0.002]
Elevation (meters)^2	-0.000***	-0.000***	0
	[0.000]	[0.000]	[0.000]
Constant	-7.472***	-6.910***	0.78
	[0.816]	[1.032]	[1.116]
Pseudo R-squared	0.17	0.08	0.08
Model chi-square	213	115	94
Observations	929	744	595

Reported in brackets are standard errors

Figure A1. Site stratification