Agriculture in Developing Economies

Alain de Janvry, Elisabeth Sadoulet, and Tavneet Suri[1]

March 31, 2015

I. Introduction

This chapter deals with the role of field experiments (FE) in agriculture. We discuss the issues and challenges that we think are specific to agriculture when using field experiments. We then go on to highlight attempts at addressing these challenges and issues where possible.

Agriculture is an important sector of economic activity in developing countries, with chapters devoted to it in the Handbook of Development Economics as well as a separate Handbook of Agricultural Economics. Here, we focus on the role that field experiments can play in analyzing development issues in this sector when they may be valuable and where the particularities of agriculture may play a role in how the field experiments should be designed and implemented. Although there is a vast literature on agriculture, this chapter will not attempt to review it. We will instead highlight some of the contributions that field experiments have made in relation to six problem areas where the specific features of agricultural environments and farmer behavior in these environments complicate field experiments. For each of these areas, we describe in detail the issues and the implications these issues have for the design of field experiments.

Over the last decade, there has been a rapid growth in not just the use of field experiments in development, but more specifically in studying agricultural issues in development. A number of the field experiments conducted to date have been focused on first-order of importance problems such as technology adoption and diffusion, and learning from peers, as well as more broadly, from social networks. Other field experiments have focused on the interaction between finance and agriculture, studying the role of liquidity constraints, savings constraints (and commitment issues), and the role of insurance given the high risk nature of rain-fed agriculture. More recently, field experiments have also ventured into the domain of understanding alternative business models that provide value to farmers (such as studying the importance of crop storage), contracts between farmers and traders, seasonality in labor markets and crop cycles, etc. Though this chapter is not a review of the literature, a number of these recent and ongoing field experiments will be discussed. Supporting the growth of field experiments in social science research, there has been a commitment from donors to better understand some of the outstanding questions in agriculture via the use of field experiments. A good example is the Agricultural Technology Adoption Initiative (ATAI) at the Abdul Latif Jameel Poverty Action Lab (J-PAL) and the Center for Effective Global Action (CEGA) that funds studies on the adoption and impact of technologies in agriculture. In the process to starting ATAI, a white paper on the outstanding issues with technology adoption was written (Jack (2011)) – it provides an excellent review of the state of the arts at the time, though it does not of course include results from the ATAI (or other) studies conducted since.

[1] We are grateful to Shweta Bhogale, Erin Kelley and Gregory Lane for excellent research assistance.

This chapter is structured as follows. In Section II, we provide a conceptual framework to better understand the role field experiments can play in agriculture. In Section III, we discuss each of six areas in agriculture where specific features of agriculture and the environment (both natural as well as economic) that farmers face have implications for the design and implementation of field experiments. The six areas we cover are: (a) dependence on random weather realizations and risk; (b) heterogeneity across space and transaction costs; (c) seasonality and long lags; (d) market failures and non separability; (e) spillovers, externalities and general equilibrium effects; and (f) difficulties of measurement. For each of these areas, we first discuss the important issues to be studied and, second, the implications for the design of field experiments. In Section IV we discuss the potential use of FE to reveal the production function in agriculture. Section VI concludes.

II. Issues in Agriculture and Field Experiments: A Conceptual Framework

Natural phenomena play a large role in agriculture. In most developing economies, agriculture is mainly rain-fed, which implies that there is often only one long growing season per year. In areas of the developing world that are irrigated (principally in Asia), while there are frequently two or three growing seasons per year, the seasonal pattern of production remains sharp as crops grown in different seasons are not the same, but interactions between them are important. Different crops have different growing season lengths and affect soil nutrients and are affected by plant diseases in different ways. All of this leads farmers to decide on an overall production plan for the year. Within one annual production cycle, the farmer will typically produce multiple crops over multiple plots and seasons and frequently also attend to a herd of different animal species. He may also work off farm to supplement this farm income.

In traditional modeling, one thinks of agricultural production as an implicit relationship between a vector of outputs Q, a vector of inputs X chosen by the producer, and a number of fixed factors, Z^q, with output affected by the weather realization W:

$$f\ (Q, X, Z^q, W)\ =\ 0 \tag{1}$$

In a context of perfect markets for output, inputs, credit, and insurance, the behavior of the decision-making unit (the household in most cases) would be to choose inputs to maximize expected profit subjected to some constraints and the production function in (1) above. The reason why it would be expected profit is that the farmer would be optimizing over his expectations of future prices as well as of weather outcomes. This leads to a set of optimal inputs, each of which is in itself a function of the fixed factors, input prices p^x, expected output prices Ep^q, and the distribution function of weather outcomes $g(W)$, in addition to possible constraints. Output, in turn, depends on all the same variables and the weather realization in a particular year.

However, given the more common scenario of imperfect markets for inputs and/or outputs, the quasi-absence of credit and insurance markets for smallholder farmers, and large transactions costs in rural areas, the household will instead maximize its utility over consumption (that includes home-produced and purchased goods and leisure), given some specific household preferences, subject to an internal equilibrium for non-traded goods, a time constraint for own family labor, and a budget constraint over traded goods. This then implies that the optimal input choice depends not just on the production characteristics described above, but also on household preferences, call them Z^c.

To complicate things further, note that these optimal input choices will vary from year to year and from location to location. This is because input decisions are taken throughout the long annual cycle. Expected prices and weather realizations up to the time of input choice and the expectations of weather realizations for the rest of the season till the outputs are produced all vary from year to year as well as from one location to another. Prices all have an explicit spatial dimension, depending on harvests which in turn partly depend on weather realizations. Similarly, weather distributions and their realizations are very local in nature, both for a given year as well as in a given location.

A field experiment will typically introduce a treatment T affecting any of the exogenous decision factors represented in our framework by the fixed factors (property rights and, somewhat abusing the terminology, access to information, extension services, etc.), the constraints (access to credit or insurance), or the prices (subsidies, payment for environmental services, contractual arrangements). To the extent that inputs and outputs are jointly determined, this treatment will usually affect all input decisions taken after the intervention, and then all outputs. Hence, measuring the outcome of a specific intervention cannot be limited to the specific output expected to be directly affected by the intervention. It also requires measuring impact on the many other outcomes that may be linked on the input side (which may well be all inputs given that family labor is an input that spills across all activities). The producer's response to any treatment that aims at inducing some increase in output could produce substitution into specialization at the cost of decreasing other activities and hence a potentially lower welfare effect. Or conversely, a strong shock on any input or output could have a very small effect if the farmer re-optimizes all his choices and spreads the shock over all activities.

Like in any well-balanced randomization, the average treatment effect can be simply computed by the difference in means of the selected outcomes Y. However, given the integrated decision process, it is very likely that the many dimensions of heterogeneity described above will translate into heterogeneous treatment effects. At the extreme, as examples, the promotion of a labor intensive technology may have no effect on households that are labor constrained; information on prices will not affect households that are too far remote from the market and hence have no bargaining power with traders who come to their farm gate; a drought resistant crop variety will have no benefit (or even potentially a negative effect) in a year where there is normal rainfall. Keeping track of these heterogeneities then becomes essential to the conduct of the experiment. The outcome of interest Y could be anyone of these inputs or outputs, or some aggregate value, sales, profit, or a welfare measure related to consumption. The general expression for the conditional impact will thus be:

$$E[Y_{ilt} \big| (T=1), Z_i^q, Z_l^q, Z_i^c, Ep_l^q, p_l^q, p_l^x, g(W_l), W_{lt}\big] - E[Y_{ilt} \big| (T=0), Z_i^q, Z_l^q, Z_i^c, Ep_l^q, p_l^q, p_l^x, g(W_l), W_{lt}\big] \quad (2)$$

where Y_{ilt} is the outcome for farmer i in location l in year t; the Z^q are fixed factors that either vary by individual (i) or location (l); the Z^c are household preferences; Ep_l^q are the farmer's expectations of output prices p_l^q; p_l^x are the input prices; $g(W_l)$ is the distribution of weather for that location and W_{lt} is the weather realization for that location in year t.

The strong inter-relationship between the multiple inputs and outputs of any agricultural activity, and its potential dependence on consumption, forces the researcher to take a broad view on the impact of even a simple treatment *T*. This is because a partial view on, for example, the impact of a treatment on a single output would give a very incomplete measure of overall impact.

Finally, the spatial dispersion of agriculture and the presence of high transaction costs create local economies, with intensive interactions within the localities and relative isolation from the rest of the larger economy. These conditions carry opportunities for high spillovers, externalities, and general equilibrium effects. In the framework developed above for example, information externalities will occur if the treatment diffuses in the locality affecting the underlying exogenous context Z_l^q among the non-treated. An intensive intervention facilitating the accumulation of stocks of harvested products will affect local prices p_l^q and p_l^x, which spread the benefits to all sellers in the community and reduce the benefit of arbitrage for the beneficiaries (Burke, 2014).

This schematic presentation of the agricultural production process highlights some key issues that field experiments will encounter and that we develop below: (a) the critical role of weather in affecting production and the treatment effect, and the difficulty of computing expected impact with only very few realizations over time; (b) the spatial heterogeneity in physical and economic contexts, and the tension between getting precise results on homogenous groups and the scope of external validity; (c) the seasonality and long lags in the production process, that impose high demands on information collected by researchers, but also introduce opportunities for shocks and changes affecting behavioral responses between the intervention and the outcome; (d) market failures and non-separability of household decisions that create a whole additional dimension of heterogeneity that is not easily characterized by simple exogenous factors; (e) social spillovers, environmental externalities, and general equilibrium effects that affect both the impact of the intervention and its measurement; and (f) measurement issues that stem from the necessity of observing the quantities of so many inputs and outputs and their prices to properly measure impact and the channels of causality of even a fairly targeted and simple intervention.

By its integration with the life of households, agriculture is clearly more than a sector of economic activity. We saw above that in the context of imperfect markets, consumption and production decisions are strongly connected. An intervention encouraging specific cropping patterns may have as its main objective the family's nutrition and health. Other interventions may have the explicit purpose of reducing the labor burden on children or of affecting the balance of power between genders in the household. Because of its very large dependency on natural resources, agriculture is also strongly related to the environment. Interventions encouraging certain practices may be seeking to enhance the long-term sustainability of resources, such as soil or water conservation. The range of outcomes of interest thus spans a large domain.

Field experiments can be very useful to precisely measure the impact of specific treatments and the channels involved. Other topics are not so easily studied through FEs. Notable is the long-term effect of technological change (because of its general equilibrium effects on prices), agricultural policy interventions (because of lack of degrees of freedom), etc. For this, other approaches are necessary, such as natural experiments capitalizing on the rollout of policies or discontinuities in treatment. Quite often, a natural experiment can be complemented by a FE, for instance to measure a particular impact on behavior or to experiment with the design of a complementary intervention.

III. Why Agriculture is Different: Implications for Design and Implementation of Experiments

a. Dependence on Random Weather Realizations and Risk

As highlighted above, the outcomes of most interventions and farmers decisions in agriculture depend on the specific realizations of rainfall and more generally of weather, especially in rain-fed agriculture. In Sub-Saharan Africa, for instance, 93% of arable land is rain-fed. Ignoring for now the other elements of heterogeneity in the general expression (2) given above, we should consequently think of the outcome Y in such an agricultural field experiment with treatment T and weather realization W as $Y(T, W)$. For any given weather realization, W_{lt}, which varies over both space and time, the conditional average treatment effect of the intervention is:

$$\mathrm{ATE}(W_{lt}) = E[Y_{ilt}|T = 1, W_{lt}] - E[Y_{ilt}|T = 0, W_{lt}]$$

The treatment effect of interest is either the average treatment effect for a given location $\mathrm{ATE_l}$, which integrates $\mathrm{ATE}(W_{lt})$ over the inter-temporal distribution of weather $g_l(W_{lt}\ )$ for that location; or the average treatment effect for the area represented by the selected locations, which is the double integral of $\mathrm{ATE}(\mathrm{W_{lt}})$ over both space and time:

$$ATE = \iint ATE(W_{lt})g(W_{lt})dW_{lt}$$

where $g(.)$ is the overall distribution of W over locations and time. With this simple notation, we can describe how the weather interacting with interventions and farmers' decisions poses issues for field experiments. In particular, we highlight four important issues.

First, in any particular year, the average treatment effect over the space of the experiment is given by the cross-sectional distribution of W_{lt} over locations,

$$ATE_t = \int ATE(W_{lt})g_t(W_{lt}\ )dW_{lt}$$

where g_t is the distribution of weather over locations in year *t*. This is generally of limited interest since it informs neither the meaningful ATE_l that may influence the uptake and actions of agents in location *l*, nor the overall *ATE* of interest to the researcher or policy makers. There are, however, situations where computing $\mathrm{ATE}(W_{lt})$ is feasible and may be useful. This is the case when there is a wide range of random weather shocks over the sample area in this particular year. The precision of the conditional results depends on the density of observations at each weather realization. Using secondary data on time series of weather events in specific locations allows one to compute ATE_l by integrating over the distribution of weather realizations in that location. The near universal availability of gridded databases on rainfall and temperature makes this possible. An illustration of this is the result in Dar et al (2013) and Emerick et al. (2014) on the impact of flood tolerance on rice yields across the number of days of flooding. In this case, the conditional results were relatively easy to obtain given the geographically dense cross-sectional variation in occurrence and duration of flooding across farmers' plots.

Second, due to the length of the production cycle in agriculture, we typically only observe a few weather realizations over time for a particular location as part of a field experiment. This implies that we may not be able to compute conditional ATE_{lt} over the whole distribution of weather for each particular location. This is one of the most serious limitations to measuring impact in agricultural field experiments. As a consequence, we only know limited segments of the ATE function for particular climatic events that happen to have been observed. This will leave considerable ambiguity in the characterization of any conditional outcomes, with potential biases if we try to infer other segments of the distribution of ATE_l from observed events.

Third, weather is a multidimensional event which is difficult to characterize. Multidimensionality makes it hard to know what matters in any given weather realization in terms of affecting the observed outcome. For example, the variables used to measure the stress of low rainfall on rice production in India are potentially the date of the on-set of the monsoon, the cumulative rainfall over different phases of the growing season, and the number of contiguous days without rainfall during the flowering period. We also know that temperature matters, as measured by degree-days, as well as a number of other factors such as wind speed and hours of sunlight.

Fourth, incomplete information about the impact of an intervention due to the role of weather events also has implications for how farmers understand the relation between a given intervention and its outcomes. Their understanding of the value of the intervention is based on only one or a few weather realizations. It is then difficult for farmers themselves to get a precise estimate of the returns to the induced action. For example, Beaman et al. (2013) argue that "if the signal on the profitability of fertilizer is weak relative to the noise resulting from weather variability, it will be hard for farmers to learn about how much --if any-- fertilizer is optimal for them to use on their particular plot of land given other possible constraints they face on inputs (including labor, for example)." Learning-by-doing is thus conditional on the realization of weather outcomes. Communicating with others about the outcome of the intervention is also imperfect because the weather realization that conditioned the outcome is difficult to characterize and inform. For field experiments that give importance to behavioral responses, carefully documenting what farmers may have been able to learn by their own doing and from others is important and yet quite difficult to do.

In the particular case where the intervention is a risk-reducing technology or a weather insurance product, the timing of the weather event relative to the intervention will affect the inference that both the researcher and the farmer can draw from the intervention. A risk-reducing technology typically has a yield penalty in normal years. Similarly, an insurance product has a premium to be paid even in normal years. If a normal year occurs before a shock year, this penalty creates a negative wealth effect for subsequent years. Treatment and control are therefore no longer identical after a shock year has occurred. A series of normal years may completely discourage adoption, altering observed behavior when a bad year occurs. Selection into treatment will keep in the farmers who are more risk averse and less liquidity constrained. By contrast, a bad year occurring before a good year may create a wealth effect that will make subsequent identification of the effect of the intervention very difficult. In Emerick et al. (2014), the technology tested had no yield penalty. Additionally, a bad year occurred first, identifying the shock-coping value of the innovation. In this experiment, the first year wealth effect from a seed minikit was sufficiently small not to have meaningful effects on the second year behavioral response. The second year response under normal weather therefore allowed them to identify the pure risk management response to a reduction in downside risk.

These four issues have the following implications for the design of field experiments:

(i) *Clustering*: To facilitate the computation of the conditional treatment impact, the researcher may want to cluster treatment and control observations to have the same weather outcomes as much as possible. For example, treatment and control can be located close to the same meteorological station from which the weather realization will be observed by the researcher. This, of course, needs to be done without compromising the risk of spillover effects between treatment and control that we discuss in more detail below.

(ii) *Dispersion*: Obtaining several conditional ATEs may require spreading the experiment over wide geographic areas to observe different weather realizations. The more covariate weather events are, the more geographically widespread the experiment should be. However, one drawback of spreading experiments over large areas is the potential heterogeneity in other factors across this wide area as well as the difficulty of managing experiments and collecting data in very distant locations.

(iii) *Duration*: Another option is to run the experiments across multiple time periods to get impacts under different realizations of shocks over time. Again, doing this implies that other factors may interfere with the observed outcomes. One is that we may need to allow for learning effects so as not to confound these with causal outcomes. External conditions will also undoubtedly change, and they can be endogenously affected by the diffusion process itself. Field experiments are notably complex to sustain over time, in particular due to the difficulty of preserving a control group, so allowing for an extended time dimension to observe a wide range of weather realizations may have some serious limitations.

(iv) *Information sharing*: If there is rapid loss of learning when there is a sequence of events that makes the innovation un-necessary, then large experiments with information sharing across distant locations will be necessary to preserve the value of learning. This also implies that there will be a lot of fluctuations in individual behavior over time, with many individuals moving in and out of using a particular technology (similar to Suri, 2011), with implications for power calculations in experimental design.

b. The Spatial Dimension of Agriculture: Heterogeneity and Transaction Costs

One of the main issues surrounding research in agriculture is how important local conditions are in economic decision-making. This is more complicated in agriculture than research on most other sectors for two reasons. First, there are biological or ecological (soil quality for crops) and environmental (climate, altitude) differences across space that make the impacts of farmers decisions (like which technology to use, whether to use fertilizer, how much fertilizer to use, which crops to grow, which livestock to raise) very heterogeneous (Suri, 2011). Second, there are the more standard dimensions of how the structure of the economy and economic policy affect (broadly) transaction costs. Effective producer prices differ widely according to location and market power and likely from year to year. Similarly, factor prices, such as for fertilizer, differ across farmers and may also differ from year to year, especially if they are being imported. The implication is that returns from a particular technology differ widely, making it attractive for adoption for some farmers but not for others, and more broadly affecting farmers' decisions and outcomes. These two dimensions of heterogeneity respectively correspond to the production function and the prices and constraints discussed in the framework above. Furthermore, poorly developed irrigation and farmers dependence on rainfall as their primary source of water implies that not only do soil and climate conditions matter at any given point in time, but they are continually changing and changing differentially across space which has impacts on farmers decisions (Suri, 2011).

This heterogeneity raises issues for the farmers themselves and the extension agents that advise them. First, an understanding of the true returns of the technology when it depends on local environmental and soil conditions is limited. This can be seen in the low level of tailoring in recommendations provided by scientists and extension services (for example, Duflo, Kremer and Robinson (2008)). Clearly, there is a need to develop extensive agricultural R&D systems to build up the knowledge base of agricultural technologies (the production function), and this is an area of potential collaboration between social scientists and scientists, which we discuss in Section IV. Second, the focus on yields is insufficient. It only paints half the picture for the farmer. What matters for the farmer is, of course, profits, which are crucially different from yields and much harder to measure well, as we discuss more below. Any given new technology or investment could (and if it increases yields should) increase labor costs and may well involve other costs (like costs of other complementary inputs, effects on soil quality as the technology may draw different nutrients from the soil, etc.). The costs of accessing the technology might be quite different across space due to variation in access to markets, infrastructure or, more generally, transaction costs in both input and output markets.

The limited understanding about how these external factors affect returns to technologies and investments has implications for field experiments. Without such an understanding, it may be hard to rationalize the decisions farmers make. Field experiments should therefore seek to measure conditional impacts, where conditionality is in terms of the dimensions of heterogeneity that matter for the outcomes of interest. Knowledge of conditional impacts will perhaps also allow better customization of recommendations to the specific conditions under which a particular farmer operates, potentially achieving larger benefits from the recommendation than what can be achieved through average treatment recommendations.

As above, if X measures the dimensions of heterogeneity that matter for the outcome Y of interest, conditional impact is measured as $ATE_X = E\left(Y \mid T=1, X\right) - E\left(Y \mid T=0, X\right)$ and the ATE is given by:

$$ATE = \int_{\substack{X \in \text{Experi-} \\ \text{mental Pop}}} ATE_X \, f(X)\, dX \,.$$

This particular ATE has the following limitations: (i) it is not informative for specific individuals within the population, and may even not apply closely to any one in particular if there is considerable heterogeneity in the population; (ii) it may be measured with a large standard error; (iii) it cannot be used out of sample. Measuring conditional impact will require either a large sample that internalizes the relevant dimensions of heterogeneity, with enough power to measure conditional impacts, or a focus on homogenous population subgroups where the average treatment effect is meaningful. A compromise may be a stratification of the population to focus on homogenous subsets that correspond to population clusters of importance. The difficulty for implementation is knowing which dimensions of heterogeneity matter, describing them with selected indicators, and characterizing their distribution in the population. This is particularly challenging if the relevant dimensions of heterogeneity are multidimensional. Once these have been identified, policy recommendations can be customized to corresponding population subgroups, approximating precision farming for economic results with large potential payoffs for farmers. Field experiments have an important role to play in gathering this wealth of information needed to make policy prescriptions in agriculture more useful.

These issues of heterogeneity in costs and benefits have the following specific implications for the design of field experiments in agriculture.

(i) *Heterogeneity of spatial conditions*: First, ideally the field experiment should cover a large heterogeneity of spatial conditions so that the researcher can better understand whether the benefits of the given farmer decision or investment under study varies by soil type or micro-climate. However, this does not come without costs. Such field experiments are extremely costly both in terms of pure budget numbers as well as in terms of the time needed to design, manage, and study an intervention that extends across a large geographical area. This is not just true for the researcher but costs are also higher for the implementers of the intervention. An alternative is to try to work within homogeneous sub-populations, in a way where homogeneity can be measured based on observable characteristics and therefore hopefully lead to findings that can be generalized over these dimensions. Examples of such an approach include Carter et al. (2014) who chose to only work with farmers located along major roads, and Burke (2014) who only worked with the highly selected clients of his partner organization, the One Acre Fund. This of course restricts what can be learned relative to a larger field experiment (and also reduces external validity) but may be more manageable on the cost and partner side.

(ii) *Characterizing heterogeneity*: There has been little work on (and there is therefore a lot of room for) measuring some of these external factors that matter so much. One example would be to try to measure soil quality differences to get a sense of local heterogeneity and how much that matters. We know less about how to measure this, but there are surely lessons to be learnt from agronomists. In a recent study, Fabregas et al. (2015) study soil testing and the demand for it among farmers in Western Kenya. However, there are few such studies that explicitly measure soil quality. In addition, understanding how soil conditions vary over time, with climatic events, and with past use of fertilizers, and whether this affects treatment outcomes will be important. On the climate side, rainfall data is more easily available – across space and over time. These data should be incorporated into analyses where possible to better understand the role of micro-climates in agriculture. The other way to measure this would be to use an observable outcome that predicts soil quality deficiencies, for example Islam (2014) uses the leaf charts in Bangladesh that show how well the plant is taking in nutrients from the soil.

(iii) *Heterogeneity along unobservable dimensions.* Producers are also heterogeneous along unobserved dimensions (such as an intrinsic productivity) and in the effort that they are willing to apply to a new process, and hence may enjoy very different benefits from getting access to new technologies. This may explain either large non-compliance or observed low average impact of experiments that foster adoption. In reality one may be interested in measuring the impact of access to a technology for those that eventually would want to adopt it rather than for a random sample of the population. Experimental designs can be geared to reveal such heterogeneity and/or foster the desired selection of participants to the experiment. For example, Jack (2013) shows that allocation of subsidized tree-planting contracts though an auction led to higher tree survival than allocation through random assignment in an experiment in Malawi. Beaman et al. (2015) show that farmers that took loans from a micro-lender in Mali have a higher return to investment than those that did not borrow (they compare the returns to a randomly assigned cash grant given to a sample of non-borrowers to that of another population not offered loans at all and hence not self-selected out of credit). The theory of such two-step randomizations to reveal and measure unobserved heterogeneity is developed by Chassang et al. (2014) who suggest the use of "selective trials" as a generalization of RCTs. These selective trials include a mechanism that let the agents reveal their own valuation of the proposed technology, and may furthermore incentivize effort in order disentangle the role of the product itself from that of the effort applied to its use. Chassang, Dupas and Snowberg (2013), for example, invited farmers to bid to enhance their chances of winning a new mechanical technology, allowing them to measure the impact of the technology on those most eager to try it. It is, however, important to be sure that the selection mechanism reveals the unobservable characteristics that the researcher is looking for and not some other constraints such as wealth or access to liquidity.

(iv) *Heterogeneity and learning*: A lot of the above issues also complicate commonly studied concepts like learning and diffusion of technology through social networks. If individuals are very different from each other (or their plots are of very different qualities), then there may be limited room for learning. For example, Conley and Udry (2010) find that farmers are more likely to learn from those who are more similar to them or who have a similar experience. Similarly, Tjernstrom (2014) finds that soil quality heterogeneity (at the village level) makes farmers less likely to respond to their peers' experiences of a new technology. The design of field experiments about adoption should therefore pay particular attention to the degree of similarity with specific others in social networks, and how the degree of analogy matters for learning.

c. Seasonality and Long Lags

Seasonal patterns and long lags pose significant barriers to farmers. They also create specific difficulties and impose costs on researchers working on field experiments in agriculture. The seasonality of agriculture for cereal or staple crops production makes economic decisions in agriculture a lot more complicated than for, say, a high-turnover vegetable producer. Farmers make investments before or during their planting season, with no returns (i.e., no incomes) until months later when the crop is harvested. Because agricultural cycles depend on rainfall, there is also a commonality in the cycle across all farmers – they all tend to plant and to harvest around the same time. This means that a lot of their demands tend to be quite correlated, especially with regards to labor. Their sales and purchases of food products are also correlated, with a sharp cycle of prices on local markets if there is imperfect tradability. This means that the farmer's realizations and expectations of output prices are not just indexed by location and year, but also vary within a given year depending on when a farmer chooses to sell.

This long cycle in agricultural production also allows for partial revelations of uncertainties and adjustment of behavior along the year. Once the rains are starting, farmers will update their decisions about the optimal time to plant (Kala, 2014) and what inputs to use. As more of the weather realization is revealed, the farmer will continue to adapt his input choices accordingly. This further complicates the inter-relationship between the multiple inputs and multiple outputs that we described above.

More specifically, these seasonal cycles pose a number of issues for farmers. First farmers face long delays between expenses and revenues. This makes access to credit both all the more important as well as all the more difficult. All the more important because farmers often have to make investments that are costly such as fertilizer purchases long before they earn incomes. All the more difficult to manage as standard microfinance loans with regular payments every week or every month are not adapted since households do not have a regular income flow. For a field experiment, this suggests that making credit available is in many circumstances a necessary condition to experiment with any other intervention that implies significant increase in input use. For this reason, many field experiments have credit as a component. By the same token, it is difficult to identify the role of the other elements of the intervention from that of credit, requiring more complex experimental designs.

Second, farmers are often engaged in multiple activities spread over time, so consideration of labor calendars is an important factor in their decisions. For example, if land needs to be cleared for planting, the demand for labor at planting will be high and will involve family labor, possibly supplemented by hired labor. The same is true for harvesting and processing of the crops – both family and hired labor are used. However, in between these two periods, the main activities are crop management, i.e. activities like weeding that are often largely conducted by household labor, partly because less labor is required and partly because monitoring is important for such activities. Designing labor calendars to smooth demands on family labor time becomes an important issue in crop and technological choices (Fafchamps, 1993). This also implies that there will be associated strong seasonalities in other relevant economic variables like the wage rate. Third, due to the long time lags between actions and outcomes, there may be time inconsistencies in decision-making that can have impacts on productivity, sales, and incomes. Commitment devices become essential to overcome these inconsistencies. In this perspective, Duflo et al. (2011) explore a savings commitment device for fertilizer purchases, Brune et al. (2014) a commitment savings product with crop sales paid directly into a bank account as opposed to cash, and Casaburi and Willis (2015) an insurance scheme in Kenya with premiums paid ex-post after harvest.

Fourth, the timing of the agricultural cycle leads to an accompanying sales and nutrition cycle in a lot of developing economies. There seems to often be a market failure in crop storage in that most of the farmers who sell grain will tend to sell it at the time of harvests. This implies that sale quantities are high and prices low at harvest. Over the subsequent months, the price of food rises so that right before harvest, there is less food available and its price is quite high. There is evidence that this leads to seasonality in consumption – a lot of farmers report having a hungry season before harvests come in where food availability and consumption are lower than at other times of the year. This seasonality in food consumption could also lead to seasonality in nutrition outcomes of farmers and their household members though this has not been well or widely documented in the literature. Some new agricultural technologies try to shorten the growing cycle of the crop with the hope of allowing multiple cropping over the year with a shorter hungry season, e.g., NERICA rice (see Glennerster and Suri, 2015). It remains an open area of research to understanding how farmers should best smooth seasonal fluctuations. One option has been to test ways of improving access to storage (for example, Casaburi (2012) and Burke (2014)). However, storage itself may only be part of the story. The hungry season may force farmers into second best activities such as seasonal migration and participation to the labor market (Fink et al, 2014) that may impact on or take away from the productive activities in agriculture, particularly when technology is labor intensive.

From the experimental and research side, this seasonality is also an issue. First, there is a wide heterogeneity in crops and their seasons. This complicates understanding farmers' decisions as most farmers will invest in a multitude of different crops, with decisions made about one crop affecting others. A technological change in any of these crops (such as adoption of a short duration variety) affects decision-making in all other crops. Although there may be a season for a lot of cereal and staple crops like maize, wheat, and rice, there are other crops which do not follow such straightforward cycles and may still be a very important part of farmers' livelihoods. Tree crops like cocoa and coffee are an example, where the tree needs approximately five years to grow before it is productive and so investments are made years before there is a return to them. Another example is cassava which once planted can be left in the ground for multiple years if needs be – farmers often use this as a backup food crop when harvests of their main food crop are poor. Second, it implies long lags between the implementation of any intervention and the outcomes, extending the time horizon for a research project. If the field experiment involves piloting, this would happen over one agricultural cycle (or year) and then the intervention over the next. This raises a lot of the standard issues with research that spans multiple years. The experiment is inherently more risky – there will be more chances for external factors to derail the experiment (for example, the second round of the Glennerster and Suri NERICA experiment has been on hold due to the Ebola outbreak in Sierra Leone); household attrition will inevitably be higher as the experiment spans more time; longer term effects may be extremely hard to measure.

These issues surrounding seasonality and long time-lags have implications for the design of field experiments:

(i) *Understanding seasonality*. This may mean monitoring the outcomes from a given intervention on a seasonal basis rather than an annual basis. Monitoring things only annually could miss important steps in the process and depend on recall data that are plagued by errors and noise (Beegle et al. 2011). For example, any one intervention may impose short-term seasonal costs, with potential longer-term gains or vice-versa. For any welfare interpretation of such interventions it will be important to measure not just the longer-term annual gains but also the short-term seasonal costs. Adoption of a labor-intensive technology, for example, may reduce seasonal migration, with a seasonal cost that is compensated by a higher annual gain. Similarly data on inputs and behavior should be collected during key periods of the agricultural process. An example is the work of Goldstein and Udry (2008) in Ghana who organized data collection with a round of surveying every six weeks, which meant that enumerators had to reside in the villages they were surveying in.

Interventions in agriculture will likely affect operations throughout the year, and hence measuring their impacts requires yearlong observations. It is important to note here the role that new technologies can play in facilitating this data collection: the wide adoption of cell phones in developing economies makes it possible to collect data on labor use, financial transactions, and product sales and purchases by frequent calls (for example, the data collection in the ongoing work of de Janvry et al. (2015) in Jharkhand). There may also be interesting uses of sensors that give real time data, for example, on moisture conditions. We discuss this in more detail in the section below on measurement.

(ii) *Successive adjustments*. Field experiments that focus on interventions early in the season may need to be adjusted or complemented later on so as to be effective. Measuring their effect may require observation of intermediate events and behavior over long periods of time. NEED EXAMPLE

(iii) *Cost of delays*. Field experiments need to build in the time lags inherent in agriculture. Missing the planting season by just a few weeks for any reason including because of administrative delays in the funding or human subjects approval of a project – an unfortunately frequent occurrence which may not matter for other field experiments – can have the heavy cost of a full year lost in research on agriculture. As the timing of planting varies from year to year, for any intervention that needs to happen before planting (like introduction of a new seed or fertilizer), researchers need to build in enough time to allow for a potentially early planting season that given year. They also need to build in hold-on strategies if the planting season ends up being later than expected. This is particularly difficult when ex-ante contractual arrangements have to be made with field partners and enumerators. Finally, a single year may not be enough time to see effects and so researchers should make sure to build in reasonable time frames for agricultural interventions.

d. Market Failures and Non-Separability

In the developing country context, most agricultural production is done by what is referred to as farm households, indicating that production and consumption decisions are integrated into one single decision-making process. These households cultivate small areas of land and usually provide most of the labor needed on their farms. Farm households subject their production decisions to the maximization of utility, rather than to an objective defined on their production outlays, such as expected profit or a function of the distribution of net profit to take into account risk and seasonal patterns of income. A farm household is said to be non-separable when the household's decisions regarding production (adoption of a technology, use of inputs, choice of activities, desired production levels) are affected by its consumer characteristics (consumption preferences including over leisure, demographic composition, etc.)[2]. Many papers have demonstrated that a separability test typically fails for smallholder farmers across the developing world (for example, Benjamin, 1992, and Jacoby, 1993).

[2] We leave aside the cases where risk attitude and time discount are the only 'preference' elements that enter the production function, since these can easily be included in a production model.

The conceptual framework above (see equation (2)) highlighted how any treatment T introduced to a farmer can affect not just input decisions and hence, outputs, but that the response will also depend on household preferences given the separability failure.

The main symptoms of non-separability are households producing much of what they eat, using exclusively family labor for certain tasks, having excess family labor that cannot find outside employment, etc. They correspond to situations where the household is constrained in the amount of labor or food that it can exchange on the market, or where transaction costs on markets are sufficiently high that the internal equilibrium shadow price of the commodity/factor produced and used by the household makes it suboptimal to either purchase or sell it (Renkow et al., 2004). For example, Fafchamps (1993) shows that farmers select an overall annual cropping pattern that maximally smoothens labor needs throughout the year. Similarly, farmers maintain the cultivation of several varieties of rice with different maturation lengths or planted at different times to spread the harvest time (Glennerster and Suri, 2015). Farmers often choose to maintain the production of their main staple food (including in India where they have access to cheap government subsidized rice), limiting availability of resources for potentially more profitable cash crops.

Since the behavioral models of separable and non-separable households are quite different, we can expect some very fundamental heterogeneity in the uptake of and response to certain agricultural interventions along this dimension. Whether this is important or not of course depends on the intervention. For example, the take up of a new technology may be low because farmers cannot sell any excess output they produce or acquire labor they need for a labor intensive technology (see Beaman et al., 2013). Alternatively, a farmer may not be able to respond optimally to changes in the price of crops because they are constrained to produce what they want to consume and to use labor as available in the family, which might prevent the household from reallocating its land toward more profitable crops (de Janvry, Fafchamps, and Sadoulet, 1991). Like any other form of heterogeneity, this does not invalidate the average uptake or ATE found in the standard analysis of a field experiment. It does, however, raise some specific issues.

First, the dimension of heterogeneity in separability is not directly or easily observable, as opposed to say distance to a city or soil quality or rainfall. Whether a household is separable or not is the result of an equilibrium that depends on its resources and preferences. One can sometimes use proxies for this. For example, if the main constraint is on the labor market side (either on sale or purchase), the ratio of available family labor to farm size may provide some indication of the likelihood that the household may be separable or not (Beaman et al., 2013, show that high return to agriculture in Mali is associated with large household size). However, this is imperfect as it does not properly take into account the quality of the labor force and any alternative occupations that determine the opportunity cost of working on the farm. Furthermore, even these resources are in many ways endogenous. A household may choose to rent land in or out, or choose how many of its labor force is made available for farm work. At the end of the day, being “non-separable” is partially an endogenous choice of the household.

Second, separable and non-separable households not only differ in terms of the intensity of their technology uptake or price response, but the channels that explain the impact of any intervention are also different. Key with separability are transaction costs on markets that make the household into a net seller or a net buyer of a particular food item or of labor. Key with non-separability are resource endowments and internal demand that affect the shadow price of this resource or product. For as long as the intervention will affect the shadow price without making it hit the effective farm-gate market price as seller or buyer, the household will remain in the non-separability status (Sadoulet and de Janvry, 1995).

This issue of separability has implications for the design of field experiments in agriculture:

(i) *Household vs. farm as the unit of analysis*. The researcher needs to define the unit of analysis as the household as a whole, not just the farm operation or a given plot. While the intervention may be primarily targeted at the production process, non-separability implies that the effect on other household decisions will be important in assessing overall outcomes.

(ii) *Data on production and consumption*. A field experiment needs to consider how the intervention will affect both production and consumption decisions. To this end, surveys should try to collect detailed information about farmers' production as well as consumption decisions. This can help establish how an intervention targeted at production might be affected by the farmer's consumption preferences. Similarly, an intervention targeted at consumption, such as a guaranteed employment scheme or a food subsidies program, will have potentially important effects on production decisions.

(iii) *Characterizing household-specific market failures*. Researchers need to characterize the prevailing market failures in areas where they design the field experiment, and, of course, these will vary from household to household. Particular interventions, for example, that reduce transaction costs and link households to value chains, may transform households from the non-separable to the separable status, with potentially large implications for technology adoption and the elasticity of supply response.

(iv) *Using FE to test for separation*. Testing for separability relies on detecting a causal effect of household consumption characteristics on production behavior, and has till now been conducted using instrumental variable techniques. An innovative use of field experiments could be to reveal the non-separability of household behavior by, for example, looking at response to an intervention that affects consumption patterns, such as the effective price of food or the opportunity cost of family labor. These would only affect production decisions if the household were non-separable.

e. Spillovers, Externalities, and General Equilibrium Effects

In rural settings, people are closely linked to each other, implying that interventions on some individuals may generate a wide range of spillover effects, externalities, and possibly general equilibrium effects. While externalities are not a specificity of agriculture (and hence will be found in many other chapters), spillover effects in the diffusion of new technologies, externalities on the environment, and general equilibrium effects are particularly important in agriculture. In the context of agriculture, we can classify these community links as personal ties among farmers or their households, cost transfers among neighbors through the environment, or price changes through general equilibrium effects in local markets.

Firstly, a treated farmer will likely inform others in his social network that he is cultivating something new, changing his farming techniques, or adopting a new insurance scheme. This will impact others' behavior in the community through learning from others, imitation effects, and economies of scale. Second, there are large local environmental externalities in agriculture. For example, the effectiveness of biological control methods depends on whether the fields around yours are similarly managed; the cost of underground water depends on the intensity of pumping of farmers extracting water from the same underground basin as you. Finally, as farmers go to sell their products on the local markets, hire labor, and earn some additional income in off-farm activities, prices and wages may change more broadly. These general equilibrium (GE) effects may in turn alter the initial treatment effect, making long-term outcomes possibly very different from short-term effects.

These externalities and GE effects have traditionally challenged field experiments, as they contradict the stable unit treatment value assumption and threaten the internal validity of many standard experimental designs. On the other hand they can be seen as interesting in themselves, and a key outcome of the intervention or the experimental design. In addition, looking back at equation (2) above, externalities and GE effects can affect all aspects of farmers decisions via their impacts on the costs side, via input prices (such as wages) and expected output prices. On the learning side, externalities may play a big role in farmers decisions as farmers' information sets on the benefits and costs of any decision evolve over time.

We separate the discussion on these issues into three separate areas. First, we look at spillover effects from social interactions, in particular at farmer learning and the diffusion of agricultural technologies. In rural settings, individuals are keenly aware of what their neighbors are doing because they can physically observe their actions. In addition, agricultural communities are closely knit social groups. Individuals within these communities share strong social bonds, which can take the form of informal mutual insurance arrangements, joint production, product sharing, or altruistic behavior. As a result of these two factors, there will undoubtedly be direct spillover effects from one treated farmer to other farmers in the community. An externality commonly studied in agriculture is in learning and diffusion, especially around new agricultural technologies. This is of course an important externality, with a large slew of evidence highlighting when and how learning effects may be important (a non-exhaustive list of examples of field experiments in this area includes Hanna et al., 2014; Buck and Alwang, 2011; Ashraf et al., 2009; and Duflo et al., 2011). Ultimately, there is evidence that farmers do interact with each other, especially within close-knit villages and social networks and that these interactions have important implications for their decisions (see, for example, Cai et al., 2014, for the design of a FE to identify the role of social learning in the adoption of a new weather insurance product). However, the precise diffusion patterns we observe in an experimental setting are a construct of the experimental design. Randomly selecting a handful of farmers in the community to receive a new technology allows one to study some partial features of the network effects, but will likely not mirror the natural processes by which technologies are introduced and spread. The extent of diffusion and spillovers will depend on the structure of the community and its surroundings.

The point of entry matters. For example, Emerick (2014) finds that farmer-to-farmer diffusion of seeds in Odisha is constrained by the deep fragmentation of social relations in village communities along caste positions. A socially neutral door-to-door offering at market price secures a much higher level of uptake (40%) than what can be achieved through farmer-to-farmer diffusion (8%). Mobarak and Ben Yishay (2014) analyze the persuasiveness of different agents in communicating information on new technologies in Malawi. In this study, the alternatives are government-employed extension workers, 'lead farmers' who are educated and able to sustain experimentation costs, and ‘peer farmers’ who are more representative of the general population and whose experiences may be more analogous to the average recipient farmer’s own conditions. Farmers find communicators who face agricultural conditions and constraints most comparable to themselves to be the most persuasive. In a novel experiment, Beaman et al. (2014) investigate the effectiveness of various diffusion models in the transmission of new technologies within rural networks in Malawi. They find that the technology diffuses best when farmers interact with multiple individuals in their network who have experienced the new technology. Geographical proximity on the other hand does not seem to fuel diffusion. Duflo et al (ongoing) find little diffusion amongst the control group, but that treatment farmers are more likely to adopt the more treated friends they have, a compounding or re-emphasis effect of the treatment.

Similarly, the complexity/riskiness of the new technology matters. Some practices or technologies may be easier to learn than others. Some field experiments choose to pair new technologies with visits from extension agents who can answer any questions and provide useful information to farmers. Glennerster and Suri (2015) recognize that early adopters generate positive externalities to surrounding farmers by sharing their experiences with the new technology, and providing information about its effectiveness. They designed a field experiment to test this diffusion process, but found little evidence of any diffusion. Heterogeneous village-level unobservables such as soil quality also matter in this process. Tjernström (2014) shows that unobserved heterogeneity in soil quality makes individuals less likely to respond to experiences within their social network. Specifically, she finds that social network effects are weaker in villages where soil quality is more varied. This shows that the extent to which social networks can be relied upon to transmit information depends on the quality of the information environment in which individuals operate. The speed with which a technology diffuses also depends on community level interactions. Moreover, as these spillovers take time to develop, the short and long term effects of the technology will change. Adoption can be slow if farmers expect to learn from others, and wait until others adopt first. Adoption can also be slow if informal networks are already providing agricultural services – for example, Mobaraq and Rosenzweig (2014) show that demand for formal weather insurance products is lower in areas where the caste group is more strongly indemnifying against village-level weather events.

Second, we briefly discuss environmental externalities. There is a whole additional set of externalities that arise in agriculture that come from crop diseases, pests, water and soil conservation investments, etc. Decisions farmers make about any investments that change the use of water and that affect crop diseases and pests can have broader impacts than just on their own farms. Though there are many examples of such practices (such as integrated pest management, water extraction, etc.), there has been little work in field experiments to study these spillovers. Water is a very specific resource in agriculture, partly because of its importance to enhance agricultural productivity and partly because it is a communal resource. There are therefore large externalities imposed by water extraction, both across space as well as on future generations (see Foster and Sekhri, 2007, for an example). In addition the short run benefits and the long run benefits of improved water access are unclear because it is a communal resource across generations (see Hornbeck and Keskin, 2015). There is clearly a lot of room for field experiments to be designed to help better understand the magnitude of these externalities and their implications for agricultural investments.

Third, we discuss market level and GE effects. Agricultural markets are generally small and isolated from larger trade networks due to high transaction costs. Local farmers produce a few main crops, and sell any surplus they have in nearby markets. Labor markets are typically local, and land markets even more so. In addition, given the externalities described above, a number of field experiments are designed as cluster RCTs, i.e., where the level of randomization is a village or a community rather than an individual. These features have important consequences for field experiments. As a result of these shallow markets, general equilibrium effects inside the village are likely to be pervasive. These effects also occur via non-agricultural outcomes. If the farmers treated by any given intervention see a rise in their agricultural income, they will spend more money on other services provided by the community. This will increase revenue flowing to other non-agricultural activities and contribute to market level changes in prices. Evidence for GE effects is given by Jayachandran (2006) who shows that productivity shocks (in particular, rainfall shocks) cause large changes in the district level wage in India. Similarly, Mobaraq and Rosenzweig (2014) document the general equilibrium effects of rainfall insurance on agricultural wages.

The flip side of these isolated markets is that they reveal that transaction costs between markets are high, that there is immense potential for arbitrage, but that there are market failures (potentially credit markets for traders or traders being imperfectly competitive) that restrict these arbitrages from happening. Any given intervention that affects the trading or market environment may well have impacts on prices and hence general equilibrium effects.

Very few studies have tried to measure these effects, often because the interventions have been too small to affect prices or other GE outcomes in any measurable way. However, understanding these effects is crucial when wanting to understand the impacts these interventions would have at scale. Exceptions include Burke (2014) who tries to measure the price impacts of a crop storage intervention; an ongoing field experiment by Glennerster and Suri on NERICA rice in Sierra Leone; on-going experiments by de Janvry et al. on the labor market effects of drought resistant rice seeds in Jharkhand, and the effect of similar technologies on the water market in Bangladesh. In a non-field experiment setting, Evenson and Gollin (2003) showed that prices decreased a lot due to the Green Revolution and this benefited consumers. In general, farmers only benefited from this if there were significant reductions in production costs that were greater than the fall in prices. Similarly, Svensson and Yanagizawa-Drott (2012) study an intervention where information on food prices was given over radio stations in Uganda. Though there were no overall effects on average crop revenue, there were big distributional effects across rural and urban areas.

Finally, there may be strong feedback loops that result in GE effects. The short-term effects of many interventions will differ from their long run impacts. A priori it may be difficult for research to predict what will happen over longer time horizons. As the agricultural and economic systems adapt to a new intervention, many of the original gains from the intervention may be mitigated (or possibly increased). It is extremely important for policy makers to be aware of these effects before they invest in an intervention strategy that they hope will have long run benefits. This scenario is conceptualized in the famous "technological treadmill" (Cochrane, 1993): early adopters might see larger profits, as they are the first to have access to the new technology. As more and more farmers gain access to the new technology, and "get on the treadmill", the initial adopters will see their profits decrease. For example, Burke (2014) finds positive effects of a storage intervention in areas where there was a low density of treatment. However, in areas with higher treatment density these positive effects were significantly reduced due to general equilibrium price changes. As prices decline on local markets, the short-run early adopters' gains typically measured in RCTs will tend to be rapidly dissipated and transferred to consumers on insulated local markets via lower prices, with an ultimately very different incidence of benefits from technological innovations.

All these issues have a variety of implications for the design of field experiments. Considering first the case of externalities or spillovers:

(i) *Designing the FE to avoid or measure spillover effects*. Field experiments need to either be designed to capture the spillovers that arise (from either learning externalities or the more natural crop disease/pest/water externalities described above) or work to design controls that are largely unaffected by these externalities. The latter would likely affect the level of randomization, which will then have both cost and power implications. Randomization should take place at the village level. Careful designs can also be implemented to accurately measure spillovers (e.g. Baird et al., 2012).

(ii) *Informing spillovers*. The field experiment should collect information to measure the effects on surrounding members within the community (even if the unit of analysis is at the village level). This involves measuring effects on farmers that are not treated by the intervention. For example, Glennerster and Suri (2015) designed a field experiment to measure spillovers in the technology, partly because of learning but also because farmers could just share the harvest of the improved rice with their neighbors. They find no evidence of such effects.

Considering the case of general equilibrium effects and feedback loops:

(i) *Powering FE to measure GE effects*. Measuring general equilibrium effects is challenging from two perspectives. First, most interventions are small and so the effects they are likely to have on market level prices or other GE outcomes is small, which poses significant power problems in measuring these effects. Researchers need to account for these power issues in their design and understand that even though statistically significant GE effects cannot be detected, this may be entirely due to the low power of the study design to measure these.

(ii) *Using market surveys to inform GE effects*. Second, these effects can be widespread through markets. Field experiments should incorporate the design of market level surveys that go beyond the household level surveys where possible. These surveys can record prices in nearby markets, as well as wage levels. Most importantly, researchers should try to measure welfare effects on consumers via falling prices in local markets. As prices fall in local markets, some net seller farmers who did not adopt the technological innovation may start entering the category of non-separable subsistence farmers.

(iii) *Measuring the cost side of GE effects*. The full set of cost reductions that happen due to any agricultural technology changes are hard to measure. Researchers doing field experiments in this area should collect detailed data on all possible cost reductions of any technological intervention.

(iv) *Using demand-side interventions to preserve early adopter gains*. With shallow local markets, the short and long-term benefits for adopters of any yield increasing or cost reducing innovation can be dramatically different. This stresses the importance of linking more effectively local to global markets for the benefits from technological change to be at least partially retained by farmers. To deliver sustained benefits to farmers, supply-side interventions should thus be complemented with demand-side interventions that effectively deepen local markets.

(v) *Sustaining and scaling-up FE to measure GE effects*. Field experiments should extend over longer periods of time and be scaled up over larger numbers of adopters to see how the profitability of the intervention changes as more and more farmers adopt. It is also important to measure whether the benefits to new technologies wear out, and how often new technologies need to be brought in.

f. Difficulties of Measurement

A field experiment will typically aim at measuring how an intervention on a small number of dimensions affects an agricultural outcome. As mentioned above, inputs and outputs in agriculture interact in complex ways. For example, promoting the production of a particular cash crop will affect the production of all other crops; promoting fertilizer use on a specific crop will affect use of other inputs on this crop and also potentially the whole production process. There is thus no escape to measuring the full set of multiple inputs and outputs, in itself a daunting task due to the extent of the information that needs to be collected and the difficulty of properly characterizing some of the inputs and outputs.

Observing the margins on which the farmer has adjusted to the intervention is important in itself, as it highlights the channels through which the intervention led to the aggregate reduced form effect. For example, Emerick et al. (2014) show how the shock-coping gains from a new risk-reducing rice variety generate further benefits through behavioral responses in risk management. Beaman et al. (2013) show a reduction of the marginal effect of one input through adjustment on another margin.

The next challenge is to aggregate these outcomes into a performance indicator. In the tradition of agricultural economics, this is done by considering restricted profit that measures the return to the fixed factors or the value of all outputs less the cost of all variable inputs, i.e., using the notation from above, $\Pi = pQ - qX$. This raises a new set of challenges: (i) establishing the distinction between fixed and variable factors, (ii) measuring all variable inputs, and (iii) measuring all prices that are needed for the aggregation.

Finally, while a properly designed field experiment will balance any fixed factors over the different treatment arms to allow measuring an average treatment effect over the distribution of these factors, there may be some first-order heterogeneity in the impact of the intervention that ought to be considered. This raises the question of measuring any important fixed factors or characteristics.

We now consider these challenges and the implications they have for the design of field experiments:

(i) *Concepts difficult to inform*. In many cases, farmers do not know the quantities of inputs used or the quantities of outputs produced. This is particularly the case for an output that is at least partially consumed and that is harvested over a period of time. A good example is cassava that stays in the ground throughout the year (and potentially even multiple years) and is only harvested when needed for consumption, or milk that is continuously collected. In this case, enumerators that ask about total production over the last twelve months will not get any accurate answers because this is a concept that farmers do not use when thinking of cassava or milk (Carletto et al. 2015, Zezza et al., 2014). In other cases, the process is so complex that farmers are not even aware of whether what they do is important (for example, Hanna, et al., 2014). This implies that researchers may have to directly observe input use and/or production rather than ask farmers for it. For the case of continuous use, the researcher may need to rely on recalls or else opt for more frequent data collection (see Goldstein and Udry, 1999). The pervasiveness of cell phones in the developing world may make this easier (see de Janvry et al, 2015). This is not without its own problems though, as frequent interviews may encounter issues of survey fatigue and attrition.

(ii) *Quantifying self-provided inputs*. Many inputs are self-provided by the household, for example household labor. There is a debate as to whether these should be considered as fixed or variable factors in the calculation of restricted profit. If the household was completely dedicated to agriculture, household labor could be considered as a fixed factor, but this is generally not the case. A difficult issue is in defining the proper unit in which to measure labor use, because the work is irregular and dispersed. A labor-day is not a concept used by farmers themselves and hence they do not necessarily know how to quantify it. Similarly, detailed accounting of time use has proved very difficult to do (see Jack et al (2015) on how they collected time use and their pilots for various ways of collecting time use). Perhaps new electronic tracking devices can provide opportunities to improve data collection on time use.

(iii) *Measuring prices*. Prices may be observable but often have strong seasonal variations. In a world of well-functioning markets, neither the spatial nor the seasonal variation of prices presents a conceptual measurement problem. Spatial variation would reflect heterogeneity in transaction costs, and seasonal variation the cost of storage. When there are market failures, with strong seasonal variations, the price chosen to value output can make a lot of difference on the presumed profitability of a farmer. Duflo et al. (2008) chose to price maize at the level it reaches just before the next season’s harvest, as most farmers are then net maize buyers and purchase maize at the end of the season after their own stocks have run out. Beaman et al. (2013) chose to value output at producer prices at the time of harvest. They argue that this, “avoids confounding a potential increase in profits from increased output, with the returns to storage”. Burke (2014) tracks the full time path of maize prices between two harvests since he is interested in the potential returns from arbitraging through storage, with farmers going from sellers when prices are low to buyers when prices are high. This is not only an issue for output prices but also for input prices. In general, there is a serious deficit of information on product and factor prices. If the researcher is collecting these as self-reports from the farmers, they need to carefully consider what set of prices they are asking farmers about and whether these are the relevant prices for the problem at hand. Alternatively, Falcao et al. (2014) use traders to obtain high frequency prices from local markets.

(iv) *Pricing family labor*. How to price family labor is another extremely important issue. When family labor is sold on the casual labor market, then the opportunity cost of work on the farm is the wage on that casual labor market at that particular time if work on the farm effectively competes with labor allocation outside the farm (Jack, 2013). However, even in this best-case scenario where there is participation by family labor in the casual labor market, particular on-farm activities may not compete with the outside time allocation. This is the case for tasks done after a regular day of work or during weekends. In this case, the cost of family labor remains a shadow price. If family labor does not work at all outside the household farm, then it is a fixed factor shared by all household activities. The shadow price in each activity (and in agriculture) is the equilibrium price internal to the household and is hence endogenous. This shadow price is not directly observable. It is well known that most family farms are not economically viable when family labor is valued at the opportunity cost on the casual labor market. The continued existence of these farms indicates that this is not the correct valuation of family labor. Valuing it at a shadow price less than the going wage is one of the most difficult issues in the analysis of the competitiveness and survival of the family farm. In field experiments, we are often more interested in the direction of change in profitability due to a particular intervention than in an absolute value. Using a range of labor prices from market price to a fraction of that price and measuring the corresponding range of changes in restricted profits may be a sufficient indicator of impact. However, there is room for a lot of work to better understand agricultural labor markets and the role and productivity of family vs. casual labor in these settings.

(v) *Valuing livestock*. Many outputs are difficult to define and/or measure, such as livestock. Livestock plays an important function as a store of value, a producer of organic fertilizers for crops, as well as being a source of income (milk and meat). For many small farmers, their livestock forms a major component of their asset base. Yet, because of the variety of species and ages, the variation in animal quality and herd dynamic, it is quite difficult to measure the asset value of a head of livestock. In addition, how their value evolves over time is extremely complex as it depends on investments made by the household as well as age effects on productivity. All this is an order of magnitude more complex than crops. Any experiment that will affect livestock, either directly or indirectly should, however, keep track of this value.

(vi) *Measuring soil and seed quality*. There are a number of important fixed factors that matter in agriculture that are hard to measure. Returns to agricultural technologies and inputs are highly dependent on soil quality, which is almost completely unobserved to the researcher. Yet, returns to farmers will vary with the types of soil they cultivate. This will in turn affect the demand for the new technology. An RCT should be sure that soil characteristics are balanced on average across treatment and control groups. A survey can elicit farmers' perceptions of soil quality, but this will be subjective and incomplete. Taking soil samples helps characterize soil quality and potentially discover additional properties of the land not known to the farmer (see Fabregas et al., 2015 and Mahajan et al., 2014). These tests are, however, expensive and time consuming. We still do not know how much relevant variation in soil fertility there is across plots and over time as related to past use and recent weather events. The same applies to the genetic content of seeds in use. DNA testing can be done to reveal the origins of these seeds. They similarly are expensive, and the extent of relevant heterogeneity in genetic content of seeds in use in a particular farm population is not known. Soil and seed quality are potentially important dimensions of heterogeneity that could make a large difference in the customization of plot-level recommendations.

(vii) *Observing technological change*. Many field experiments in agriculture have the objective of measuring the adoption, diffusion, and impact of technological change. This is for instance the objective of the ATAI project. There is much interest in the question for public investment and donor accountability in investing in agricultural research, and to uncover barriers to adoption if we believe that there are many good technologies that lie idle. Observing technological change is however difficult. New seeds come under the form of a rapid succession of new releases. Each new vintage only makes a marginal improvement over the previous installment. As a sequence, there may be large gains over time; but each release only makes a deceptive marginal contribution. So, what do we use as a counterfactual? Only once in a while do we observe truly transformative technologies, such as IR6 rice and semi-dwarf wheat cultivars that made the Green Revolution. When "nature does make jumps", measuring the return to investment in research would require using as a counterfactual the original traditional varieties, often long gone or only cultivated under marginal conditions. Disappointing results in an impact analysis of the latest varietal release thus requires caution in design.

(viii) *Are double blind trials useful in agriculture?* A recent measurement issue has been raised in thinking about whether the equivalent of double blind trials is useful for FE in agriculture. Researchers have begun to think about behavioral responses to an intervention or a technology in agriculture as separate from the true yield or genetic returns to the technology (see Bulte et al. (2014) for an example). This distinction between genetics and behavior is however, at best, murky. A lot of technologies are only beneficial when they are accompanied by certain practices. These practices typically involve a change in farmer behavior and they are part of the technological package. In addition, any technology that increases yields for a farmer will automatically change some aspects of farmer behavior, for example labor since there will be more to be harvested and therefore more harvest labor will be required. This is a change in behavior that should be bundled with the technology. The distinction between a behavioral response on the part of the farmer and the technology impact is clearer in the case of Emerick et al. (2014). They find that the technology helps reduce the yield losses due to flooding, which includes some element of farmer behavior as recommended by the extension agent. Farmers additionally respond the following year to this change in the risk profile of their outcomes. The second year behavioral response is thus a lower bound to the role of behavior in contributing to the yield gain of the new technology. In general, we believe that there is little scope for blind tests in agricultural FE.

(ix) *The plot as a unit of analysis.* Finally, keeping track of plot-level panels is a challenge. The characterization of production with homogenous conditions is typically done at the plot level. Flooding for instance is very much a plot-level event. If the farm is fragmented, as typical in smallholder farming where land has been inherited and divided across family members over generations, a particular household will own several plots with very different features, the very reason why there has been fragmentation (Foster, 2014). Working with plot-level panels is thus a natural unit of measurement. The concept of a plot is, however, a challenge in measurement. The definition of a plot is a fluid concept that changes over time. Farmers do not necessarily know how much fertilizer and other inputs they have applied to each plot, while they know how much they have purchased overall. Production from various plots may be combined for threshing, making identification of output per plot difficult. In addition, a farmer may change the crop mix on a plot over time invalidating the previous year's definition of a given plot, making it impossible to match plots over time, even if one had good GPS maps of plots (which in of itself is a challenge given the expense both cost and time wise to do this). Finally, the plot is rarely a unit of decision-making, implying that obtaining information on the rest of the farm operation remains necessary.

IV. Discussion: Using FE to reveal the production function in agriculture

Agriculture is very different from other economic sectors. Unlike the case of manufacturing, where the production function is a construct of the enterprise, we know little about the true structure of the production function in agriculture as it is largely a product of nature. We should just marvel at the fact that a teaspoon of productive soil typically contains more than one million species of bacteria that are interacting in producing soil fertility. It is unclear what the full list of inputs into this production function may be – land, labor, climate, soil quality are only a small subset of the elements that could possibly matter and each of these cannot be represented by a single scalar measure. Consider for example the fact that the timing of input use is very important for the ultimate outcome. This implies that the combination of quantity and time of inputs gives us an almost infinite list of entries. We also know little about how these inputs may interact with each other in the production function for output: are there complementarities between labor and soil quality, almost surely; are there complementarities between labor and climate, again, almost surely. What form these complementarities take is not well understood. Some factors typically contribute positively to output, and others negatively. For this reason, the production function is often specified as the product of two sub-functions: a classical production function with the contribution made by factors of production, and a damage function, where damage is done by factors such as pests and weather (Lichtenberg and Zilberman, 1986).

Because of its link to nature, the production function of agriculture may have more in common with the production function for health than with a standard manufacturing process. In addition, we know little about what farmers themselves know about this production function or how they approximate what this function may be in making decisions. They may have a very limited or even error laden estimate of what the production function looks like, which implies that they are continuously learning themselves and making decisions based on incomplete information. And yet, as T.W. Schultz (1964) taught us, knowing the production function is essential to assess the marginal product of factors and make optimum decisions in factor use.

There are three levels at which research can be done to reveal the production function. The first is Agricultural Experiment Station (AES) research. Agronomic research has been a pioneer in using statistical experimentation, typically randomization in a Latin Square design analyzed in an ANOVA framework. It was famously introduced into agronomic research by R.A. Fischer at the Rothamsted Experimental Station in Great Britain. Greco-Latin Square designs can be used to test two-by-two combinations of experimental treatments, such as seeds and fertilizer doses. A high yielding variety seed and a traditional seed can thus be tested against various levels of fertilizer use. The problems with AES experiments are that: (1) the conditions under which the experiment is conducted are generally not reported. For example, the levels of irrigation water and pesticides applied in the seed-fertilizer experiment. This limits the external validity of the results obtained. In this case, it comes from the fact that the particular segment of the production function that has been revealed is not clearly specified. (2) Expectedly, with yield the reported agronomic outcome, the conditions under which the experiment is conducted (such as water, pesticides, and labor practices) are for maximum yield. This does not correspond to what farmers will subsequently do in their farms, where their objective functions will be profit maximization or utility maximization if there is risk aversion. Comparing AES experimental yields with yields in farmers' fields not surprisingly tend to find that the former are larger than the latter.

The second level of analysis is in demonstration plots set up by extension agents in farmers' fields. The farmer is coached by the agent in applying a number of production practices to a new technology. The rest of the production decisions are left to the farmer, and fixed factors and climatic events are left to nature. Advantage of the approach is that it brings the technology close to potential adopters, with their own production conditions. Problem is that the demonstration plot is generally not accompanied by a counterfactual technology (Hancock, 1992). Everyone is left to compare the treatment outcome to what he does on his own farm, which is for each farmer the next best technology that typically differs from farmer to farmer. Neither is there usually learning in demonstrating the response function of the new technology. We believe that there is room for greater value added in using the demonstration plot approach to identify the production function. Farmer selection should be formalized so we know what types are making decisions. Every farmer should be requested to define a counterfactual technology, and cultivate it in an adjacent plot. Information on impact measured by difference should be diffused across the farm population so there is an opportunity to learn about the response function.

The third is FE in farmers' fields through randomized control trials. This is the typical seed minikits approach that Dar et al. (2013) for example followed in testing the flood tolerance value of SwarnaSub1 rice. Minikits are distributed randomly in treatment villages, farmers plant these seeds in a plot of land of their choice, they apply their own self-selected cultivation practices in accordance with their objective function, particular climatic events occur, and a yield or any other outcome is observed. If there is a cross-sectional range of climatic events (such as days of flood duration), we can observe the yield advantage of Sub1 over the farmer's counterfactual seed over that range of events. Large sample properties create internal validity for the results over the conditions that prevail in the villages selected for the experiment. Advantage is that the outcome corresponds to what farmers will be doing with the technology for their own purpose. Problem is that we only learn about an average treatment effect and some heterogeneity, not the production function. Because it is difficult to characterize the conditions and events under which the measurement has been made, it is difficult for the farmer to learn from the observed outcomes and it is difficult to use the information to help others learn from these farmers' outcomes.

We see two innovative roles that FE can play in research on agriculture. First, experiments can be designed to reveal the production function. We can gradually uncover the input responses, ideally over a wide range of differing environmental and climate conditions and begin to better understand what agricultural production functions may look like. This will of course require an almost heroic effort and may lack some of the glamour that drives economists. A second role that RCTs can play which may be equally important in the short run is to better explore and understand what is the implicit production function that farmers are using in making their own decisions. Here, agronomy meets the social sciences. This will give us a window into better understanding farmer epistemology and behavior, farmer constraints, and how best can policy respond to inform agents, alter behavior, and relax constraints.

V. Conclusion

Field experiments can be a powerful tool for project design and policy recommendations in agriculture. However, the specificities of the production process in agriculture and the agency and social relations in which it is embedded make the design of field experiments particularly difficult in addressing questions of first order importance. These difficulties can, however, be gradually overcome, in particular as we become more expert at designing field experiments and better equipped at collecting and matching data. The large gaps in knowing how to use field experiments in agriculture to address important development issues create a promising research agenda.

References

Ashraf, Nava, Xavier Giné, and Dean Karlan. 2009. "Finding Missing Markets (and a Disturbing Epilogue): Evidence from an Export Crop Adoption and Marketing Intervention in Kenya." *American Journal of Agricultural Economics* 91(4): 973-990.

Baird, Sarah, Aislinn Bohren, Craig McIntosh, Berk Özler. 2012. "Designing Experiments to Measure Spillover and Threshold Effects." Policy Research Working Paper Series 6824, The World Bank.

Beaman, Lori, Ariel Ben Yishay, Paul Fatch, Jeremy Magruder, Mushfiq Mobarak. 2014. "Making Networks Work for Policy: Evidence from Agricultural Technology Adoption in Malawi." Working Paper, Northwestern University.

Beaman, Lori, Dean Karlan, Bram Thuysbaert, and Christopher Udry. 2013. "Profitability of Fertilizer: Experimental Evidence from Female Rice Farmers in Mali." *American Economic Review* 103(3): 381-86.

Beaman, Lori, Dean Karlan, Bram Thuysbaert, and Christopher Udry. 2015. "Selection into Credit Markets: Evidence from Agriculture in Mali." Working Paper, Yale University.

Beegle, Kathleen, Calogero Carletto, and Kristen Himelein. 2011. "Reliability of Recall in Agricultural Data." Research Working Paper WPS 5671, Washington, DC: World Bank Group.

Benjamin, Dwayne. 1992. "Household Composition, Labor Markets, and Labor Demand: Testing for Separation in Agricultural Household Models." *Econometrica* 60(2): 287-322.

Brune, Lasse, Xavier Giné, Jessica Goldberg, and Dean Yang. 2014. "Facilitation Savings for Agriculture: Field Experimental Evidence from Malawi." NBER Working Paper No. 20946.

Buck, Stephen, and Jeffrey Alwang. 2011. "Agricultural Extension, Trust, and Learning: Results from Economic Experiments in Ecuador." *Agricultural Economics* 42(6): 685-99.

Bulte, Erwin, Salvatore Di Falco, Xavier Vollenweider, and Menale Kassie. 2015. "Disentangling the Benefits of Innovations: Behavioral Responses Versus Technological Change in Tanzania." Working Paper.

Burke, Marshall. 2014. "Selling Low and Buying High: An Arbitrage Puzzle in Kenyan Villages." Working Paper, University of California at Berkeley.

Cai, Jing, Alain de Janvry and Elizabeth Sadoulet. 2014. "Social Networks and the Decision to Insure." *American Economic Journal: Applied Economics*, forthcoming.

Carletto, Gero, Dean Jolliffe, and Raka Banerjee. 2015. "From Tragedy to Renaissance: Improving Agricultural Data for Better Policies." Policy Research Working Paper WPS 7150, Washington, DC: World Bank Group.

Carter, Michael, Rachid Laajaj, and Dean Yang. 2014. "Subsidies and the Persistence of Technology Adoption: Field Experimental Evidence from Mozambique." NBER Working Paper No. 20465.

Casaburi, Lorenzo. 2012. 3ie Grant Report.

Casaburi, Lorenzo, and Jack Willis. 2015. "Interlinking Product and Insurance Markets: Experimental Evidence from Contract Farming in Kenya." Working Paper, Stanford University.

Chassang, Sylvain, Gerard Padró i Miquel, and Erik Snowberg. 2012. "Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments" *American Economic Review* 102(4): 1279–1309.

Cochrane, Willard. 1993. The Development of American Agriculture: A Historical Analysis. Minneapolis, MN: University of Minnesota Press.

Conley, Timothy, and Christopher Udry. 2010. “Learning about a New Technology: Pineapple in Ghana." *American Economic Review* 100(1): 35-69.

Dar, Manzoor, Alain de Janvry, Kyle Emerick, David Raitzer, and Elisabeth Sadoulet. 2013. “Flood-tolerant rice reduces yield variability and raises expected yield, differentially benefitting socially disadvantaged groups.” Scientific Reports 3, Article number 3315, November 22.

de Janvry, Alain, Marcel Fafchamps, and Elisabeth Sadoulet. 1991. “Peasant Household Behavior with Missing Markets: Some Paradoxes Explained.” *Economic Journal* 101(409): 1400-17.

de Janvry, Alain, Elisabeth Sadoulet, Kyle Emerick, and Manzoor Dar. 2015. “The Impact of Drought-Tolerant Rice on Local Labor Markets in India”. Working Paper, University of California Berkeley.

Duflo, Esther, and Rohini Pande. 2007. “Dams.” *Quarterly Journal of Economics* 122(2): 601-46.

Duflo, Esther, Michael Kremer, and Jonathan Robinson. 2008. “How High are Rates of Return to Fertilizer? Evidence from Field Experiments in Kenya.” *American Economic Review, Papers and Proceedings* 98(2): 482-88.

Duflo, Esther, Jonathan Robinson, and Michael Kremer. 2011. “Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya.” *American Economic Review* 101(6): 2350–90.

Duflo, Esther, Daniel Keniston and Tavneet Suri. Ongoing project, Rwanda.

Dupas, Pascaline, Sylvain Chassang, and Erik Snowberg. 2013. “Selective Trials for Agricultural Technology Evaluation and Adoption: A Pilot.” ATAI Pilot Project, CEGA, University of California Berkeley.

Emerick, Kyle, Alain de Janvry, Elisabeth Sadoulet, and Manzoor Dar. 2014. “Risk and the Modernization of Agriculture.” Working Paper, University of California at Berkeley.

Emerick, Kyle. 2014. “The Efficiency of Trading in Social Networks: Experimental Measures from India.” Working Paper, Tufts University.

Evenson, Robert, and Douglas Gollin. 2003. “Assessing the Impact of the Green Revolution, 1960-2000.” *Science* 300(758): 1078710.

Fabregas, Raissa, Michael Kremer, Martins Odendo, Jonathan Robinson, and Frank Schilbach. 2015. “Evaluating Agricultural Information Creation and Dissemination in Western Kenya.” Working Paper, Harvard University.

Fafchamps, Marcel. 1993. “Sequential Labor Decisions under Uncertainty: An Estimable Household Model of West-African Farmers.” *Econometrica* 61 (5): 1173-1197.

Falcao, Lauren, Paul Gertler, and Craig McIntosh. 2014. E-Warehousing for Smallholder Farmers. ATAI Project, JPAL.

Fink, Gunther, Kelsey Jack and Felix Masiye. 2014. “Seasonal Credit Constraints and Agricultural Labor Supply: Evidence from Zambia.” NBER Working Paper 20218.

Foster, Andrew, and Sheetal Sekhri, 2007. “Can Expansion of Markets for Groundwater Decelerate the Depletion of Groundwater Resource in Rural India?,” Brown University Working Paper.

Foster, Andrew. 2014. “Spillovers, Coordination Failures and Land Fragmentation.” Working Paper, Brown University.

Glennerster, Rachel, and Tavneet Suri. 2015. “Measuring the Effects of NERICA, Short Duration Rice, on Harvest Prices.” ATAI Project, MIT.

Goldstein, Markus, and Christopher Udry. 1999. “Agricultural Innovation and Resource Management in Ghana.”

Goldstein, Markus, and Christopher Udry. 2008. “The Profits of Power: Land Rights and Agricultural Investment in Ghana.” *Journal of Political Economy* 116(6): 981-1022.

Hancock, John. 1992. “Extension Education: Conducting Effective Agricultural Demonstrations.” Cooperative Extension Service, College of Agriculture, University of Kentucky.

Hanna, Rema, Sendhil Mullainathan, and Josh Schwartzstein. 2014. “Learning by Noticing: Theory and Evidence through a Field Experiment.” *Quarterly Journal of Economics* 129(3): 1311-53.

Hornbeck, Richard, and Pinar Keskin. 2015. “Does Agriculture Generate Local Economic Spillovers? Short Run and Long Run Evidence from the Ogallala Aquifer.” *American Economic Journal: Applied Economics*, forthcoming.

Islam, Mahnaz. 2014. “Practice Does Not Make Perfect: Understanding Fertilizer Mismanagement in Bangladesh through Leaf Color Charts.” Working Paper, Harvard University.

Jack, Kelsey. 2011. “Market Inefficiencies and the Adoption of Agricultural Technologies in Developing Countries.” White paper prepared for the Agricultural Technology Adoption Initiative (ATAI), JPAL (MIT) / CEGA (Berkeley).

Jack, Kelsey. 2013. “Private Information and the Allocation of Land Use Subsidies in Malawi.” *American Economic Journal: Applied Economics* 5(3): 113-35

Jack, William, Joost De Laat, Michael Kremer and Tavneet Suri. 2015. “Joint Liability, Asset Collateralization, and Credit Access: Evidence from Rainwater Harvesting Tanks in Kenya.” Working Paper.

Jacoby, Hanan. 1993. “Shadow Wages and Peasant Family Labor Supply: An Econometric Application to the Peruvian Sierra.” *Review of Economic Studies* 60: 903-921.

Jayachandran, Seema. 2006. “Selling Labor Low: Wage Responses to Productivity Shocks in Developing Countries.” *Journal of Political Economy* 114(3): 538 – 575.

Kala, Namrata. 2014. “Ambiguity Aversion and Learning in a Changing World: The Potential Effects of Climate Change from Indian Agriculture.” Working Paper, Yale University.

Lichtenberg, Erik, and David Zilberman. 1986. “The Econometrics of Damage Control: Why Specification Matters.” *American Journal of Agricultural Economics* 68(2): 261-73.

Mahajan, Aprajit, Enrique Seira, and Xavier Gine. 2014. “A Multiple Interventions Approach to Increasing Technology Adoption with a View Towards Scaling-Up: Evidence from Mexico.” BASIS Proposal.

Mobarak, Mushfiq, and Ariel Ben Yishay. 2014. “Social Learning and Communication.” Working Paper, Yale University.

Mobarak, Mushfiq, and Mark Rosenzweig. 2014. “Risk, Insurance and Wages in General Equilibrium.” Working Paper, Yale University.

Mobarak, Mushfiq, and Mark Rosenzweig. 2014. “Selling Formal Insurance to the Informally Insured.” Working paper, Yale University.

Renkow, Mitch, Daniel Hallstrom, and Daniel Karanja. 2004. “Rural Infrastructure, Transactions Costs and Market Participation in Kenya.” *Journal of Development Economics* 73(1): 349-367.

Sadoulet, Elisabeth, and Alain de Janvry. 1995. *Quantitative Development Policy Analysis*. Baltimore: The Johns Hopkins University Press.

Schultz, T.W. 1964. *Transforming Traditional Agriculture*. New Haven: Yale University Press.

Suri, Tavneet. 2011. “Selection and Comparative Advantage in Technology Adoption.” *Econometrica* 79(1): 159–209.

Svensson, Jakob, and David Yanagizawa-Drott. 2012. “Estimating Impact in Partial vs General Equilibrium: A Cautionary Tale from a Natural Experiment in Uganda.” Working Paper, Stockholm University.

Tjernström, Emilia. 2014. "Signals, Similarity and Seeds: Social Learning in the Presence of Imperfect Information and Heterogeneity." Working Paper, University of California at Davis.

Zezza, Alberto, Giovanni Federighi, Kalilou Adamou, and Pierre Hiernaux. 2014. "Milking the Data: Measuring Income from Milk Production in Extensive Livestock Systems: Experimental Evidence from Niger." Research working paper WPS 7114. Washington, DC: World Bank Group.