Recent Advances in Impact Analysis Methods for *Ex-post* Impact Assessments of Agricultural Technology: Options for the CGIAR

Alain de Janvry, Andrew Dustan, and Elisabeth Sadoulet
University of California at Berkeley

October 2010

SPIA report version **3.1**

Prepared for the workshop "Increasing the rigor of ex-post impact assessment of agricultural research: A discussion on estimating treatment effects", organized by the CGIAR Standing Panel on Impact Assessment, SPIA, on Saturday October 2, 2010, Berkeley.
We thank Mark Rosenzweig and SPIA members for their comments.
emails: alain@berkeley.edu, adustan@berkeley.edu, esadoulet@berkeley.edu

Executive Summary

This paper analyzes the challenges faced by the CGIAR in evaluating the impact of agricultural technologies and suggests avenues for improving the methodology used in impact analyses. Our focus is on technologies such as crop varieties whose adoption is easily described as a binary choice rather than best practice or policy.

The dynamic nature of technology adoption and diffusion defines some sharply contrasted types of analyses: (i) the "microeconomic" analyses that attempt to measure the impact of adoption on individual adopters, in a context where diffusion is not so widespread that there are still a lot of non-adopters, and general equilibrium effects have not taken place; (ii) as opposed to measuring the aggregate impact of a continuously evolving line of variety improvements. The paper mostly focuses on issues and methods for the microeconomic impact analyses, an area that has witnessed numerous recent methodological developments that are still not widely or appropriately used in the practice of impact assessment of agricultural technology adoption. A short section, however, recaps current practices in aggregate and long-term approaches.

The key quantity that we are seeking to estimate is the *average effect* that adoption has on outcomes for those who have adopted (average treatment effect on the treated or *ATT*). Because of the selection effect, the main challenge is establishing the proper counterfactual group against which to compare adopters. We argue that experiment station as well as on farm trials are not appropriate, because they are unlikely to reflect the conditions faced by actual adopters, nor their behavior in terms of the choice of complementary inputs, for instance. We also question the validity of selection on observables designs (regressions or PSM), because adopters and non-adopters certainly differ not only in their observable characteristics but also in key unobservable characteristics (such as entrepreneurship or ingenuity) that explain both their choice of adopting the new technology and the outcomes of interest. Difference-in-difference methods certainly improve on the single difference methods, although they are also based on the non-trivial assumption that outcomes should be evolving similarly for those who choose to adopt and those who do not.

An additional and closely related issue is that of spillovers from adoption, both on other adopters and on non-adopters. Spillovers complicate the search for counterfactuals, in that true counterfactuals should not be affected by adopters. But spillovers (positive or negative) need also to be accounted for in the impact of adoption. It is also important to understand that the *ATT* varies over time, because (i) adopters change their usage of the new technology as they use it and learn more about it; and (ii) the set of adopters of a new technology almost certainly changes with time.

The broad suggestion made here is that, whenever possible, microeconomic impact analysis should have explicit research designs that allow to estimate the effect of the new technology without relying exclusively on the observable characteristics of potential adopters. While randomization offers a solution to the selection problem, the design needs to insure that the treated producers would be adopters as in a non-experimental set-up and that there is no constraint on their behavior. Hence randomization of technology over plots within a farm is not suitable. Nor are encouragement designs that induce a random sample of normally non-adopters to adopt; this would estimate a local average treatment effect (on those induced to adopt) rather

than the desired ATT. With spillovers, even a scheme of randomization over households within a village will leave many unresolved problems.

Our suggestions are the following:

- Use either natural or randomized experiments where the village or the community is the unit of randomization. Doing so addresses the issue of spillovers not by ignoring them or trying to create an environment in which they do not exist, but rather acknowledging that they are potentially important results of adoption and incorporating them into the measured effect of the new technology.
- If randomization is used, pursue supply-side interventions where the new technology is introduced to entire villages. These villages should not have been exposed to the technology before, and the technology should be sold at its market price, not at a subsidized price. Adopters are then the entire set of farmers who find it optimal to adopt when the technology is available for purchase under real-world conditions.
- Do not limit research designs to RCTs. Natural experiments can yield reliable estimates of impact even in the absence of controlled, explicit randomization. Rollouts of a technology that were arguably random, even if they were not explicitly randomized, can be analyzed in a similar way as RCTs. The assumption of randomness in the rollout cannot be fully tested, so knowledge of the institutional context of the rollout and verification of some statistical properties are necessary. Rollouts allow for the analysis of technologies that have already been diffused and often allow use of very large administrative databases. This is a distinct advantage over RCTs, which are only possible for new technologies. Other kinds of natural experiments may be usable as well, like a geographic discontinuity.
- Leverage public-private-civil society partnerships to perform supply-side interventions, such as with agro-dealers.
- Plan the evaluation before, and conduct it during, diffusion of a new technology. The fact that impact analysis is referred to as *ex-post* should not suggest that they should be planned and performed after the fact.

Illustrations of these principles are given in three sketches of what could plausibly be impact analysis designs: for the genetically improved farm Tilapia, a goat parasite treatment, and drought tolerant varieties.

A last section addresses the analysis of long-term and aggregate effects, with the objective of measuring *ex-post* the aggregate benefit of a technology that has diffused over large areas. The challenge is of course that there is no counterfactual situation that can be observed. Researchers have resorted to different types of analyses. One of them is to focus on smaller units of observation (such as villages) on the presumption that markets are not well integrated so that they each represent a small "economy" and one can rely on econometric analysis of the observations over time to identify the causal effect of an uneven development of technological change on these units. The second type of analysis is to resort to simulation models to extrapolate impacts measured at the micro-level (most often increases in yields) to the level of aggregate effects. This includes the economic surplus method and the CGE simulation models. These are useful simulations but not impact estimations.

1. Introduction

This paper analyzes the challenge faced by the CGIAR in evaluating the impacts of agricultural technologies, as well as suggesting avenues for improving the methodology used in impact analysis. This purpose is shared in part by other efforts within the CGIAR, especially SPIA's recent review of the topic (Walker et al., 2008) and further work by Maredia (2009). Our paper addresses issues similar to those in the above papers, albeit with a different perspective and differing conclusions, while also reviewing recent impact analyses performed by CGIAR research centers and proposing some ways to proceed with future research. As with the aforementioned papers, our focus is the effect of specific technologies on outcomes such as consumption, income, and poverty. This contrasts with another kind of impact analysis which focuses on the effect of CGIAR's *research expenditures* on similar outcomes.[1] We are also not contemplating a related but different question, which is the analysis of the extent or the *determinant of adoption*.

In order to place this paper in the vast literature on the impact of technology adoption, we need to consider the dynamic nature of technology adoption and diffusion, because it defines some sharply contrasting types of analyses. When new technologies are made available, some individuals choose to adopt them and expectedly benefit from this adoption. As time passes these individual adoptions result in diffusion of the technology and its benefits in the population, with the particularity that the nature of the impacts changes fundamentally over time. Very broadly speaking, the benefits of a technology tend to diffuse in the economy to consumers and workers, while only partially remaining with producers. The share of benefits accruing to each set of actors in the economy varies as markets adjust to the effects of the new technology on outputs as well as demand for production inputs. The extent to which producers retain the benefits from adoption, as well as how these benefits vary with time, depends on the specific technology and the good being considered, including the rules of price formation on the corresponding markets.

This creates a sharp contrast between two types of questions and impact analyses:
(i) The first concerns the relatively early adopters. This sort of analysis is typically performed in a context where adoption can be described as a binary decision and there are still a lot of non-adopters. The challenge is to find among those non-adopters some good counterfactual that can be validly compared with the adopters. Much progress has taken place over the last 10 years in this type of impact analysis, with the development of methods extensively based on strategies for constructing a comparative sample of non-adopters. Applications are well-developed in the fields of education and health, from which there is much to learn, but less so in the field of technology. For lack of a better term, we will refer to this type of analysis as being "microeconomic".

(ii) A different question one can ask would be to measure the impact of, say, 20 years of improvement and diffusion of a particular line of technology development. This is a substantially different problem, in the sense that: first, it covers many years, with continuous evolution of the "technology"; second, there is likely no longer a proper counterfactual (still using the technology from 20 years ago) available to compare to the current adopters; and third, many benefits of the

[1] Examples of this type of analysis are Fan et al. (2000), Evenson and Gollin (2003), Raitzer and Kelley (2008), Maredia and Raitzer (2010), and Alston et al. (2000).

technology have diffused to consumers and workers through changes in prices and general equilibrium effects. Rigorous estimation of an impact in this context resorts to the standard econometric techniques that can exploit the progressive and heterogeneous diffusion of the technology over time and space, provided one can identify enough units that can be treated independently. The particular challenges with this method are to address the endogeneity of the diffusion process, and the existence of comparable information over time and space on the outcomes of interest such as profit, income, poverty, etc. A very different hybrid approach to this question combines some estimation of microeconomic impact as defined above with observed patterns of diffusion and a model for changes in prices and general equilibrium effects, and produces simulated aggregate and long-term effects. Because of the extreme difficulty of carrying out a rigorous and credible estimation of long-term aggregate effects of a stream of technological change, this hybrid approach is the most commonly used in impact assessments of technology by the CGIAR.

This paper focuses almost entirely on the microeconomic methods of impact analysis for two reasons. First, this is an area that has witnessed numerous recent methodological developments that are still not widely or appropriately used in the practice of impact assessment of agricultural technology change. Second, even if one is mostly interested in the long-term macroeconomic impact of technology change, these microeconomic estimates serve as an important element of the hybrid approach mentioned above, so it is worthwhile to estimate them correctly. We will, for the sake of completeness, present a short section addressing the aggregate and long-term approaches, although we are not aware of significant recent advances in this type of analysis nor do we have suggestions to make that justify a more extensive treatment of these methods.

Section 2 presents some preliminary reflections on the objectives and challenges in impact analysis. Section 3 sets forth a simple analytical framework for analyzing adoption and its effects at the microeconomic level. Section 4 reviews impact analysis methods used in recent studies and offers critiques of those methods. Section 5 offers suggestions for improving impact studies, while section 6 gives example of applications of these suggestions. Section 7 departs from the focus on microeconomic analysis used in most of this paper and discusses methods for addressing the long-run and aggregate effects of a technology.

2. Impact Analysis: Objectives and Challenges

2.1. Agricultural technologies under consideration

Before defining impact analysis and discussing its implementation, it is worthwhile to consider the different kinds of agricultural technologies that are likely to be evaluated. Each type of technology has unique limitations that must be kept in mind when setting expectations about what we can learn from impact analysis and the challenges that will arise when implementing evaluations.

1. *Yield-increasing and cost-saving technologies*. Examples of yield-increasing technologies are new seed varieties whose main advantage is in output per hectare, fertilizers, and certain new cultivation practices. Cost-saving technologies may also include new seed varieties that require fewer complementary inputs, as well as cultivation practices that produce equal results with less effort. Both yield-increasing and cost-saving technologies reduce the cost per unit of output, with the possible difference that yield-increasing

technologies actually allow for higher gross output if some inputs (especially land) are limited. These technologies are often presented or recommended to producers as packages including a seed variety and the associated best management practice. We focus however here on the "seed" component of the technology, for two reasons. First, much of CGIAR-funded research consists of the development of these new varieties.[2] Second, the choice of how much of complementary inputs to use is itself an endogenous response to the adoption of the new variety, and hence it is an integral part of what determines the impact of adopting a new variety.

In what we defined above as microeconomic impact analysis, we contrast adopters to counterfactual non-adopters, therefore measuring the *marginal* effects of the adoption of a new variety over the variety still in use by those non-adopters. This suggests that the method is best used for relatively large technological jumps that are likely to have a large impact. In the case that the technology being evaluated may have only incremental increases in yield or decreases in production cost when compared to the prevailing variety, estimating the (potentially small) marginal impact of the new technology probably requires large sample sizes, or else lack the power to precisely estimate the effect. This is an important practical consideration.

Measuring the effect of the new variety over the unimproved (i.e., not the next-best) one if there is no counterfactual group using the unimproved seed requires being able to find the counterfactual in the past and controlling for everything else that may have happened over time. We discuss this in section 7, where we present methods to retrospectively estimate the aggregate impact of a lengthy research program that has released many successive outputs, such as those discussed in Byerlee and Traxler (1995) and Morris (2002).

2. *Risk-mitigating technologies*. These technologies might not raise yields in times where conditions are favorable, but they reduce the risk of very bad outcomes when negative shocks occur. Drought- and pest-resistant seed varieties and livestock vaccines are good examples of risk-mitigating technologies.

   Evaluating risk-mitigating technologies is difficult. While adoption may impact *expected* outcomes, these effects may not always be observed. For example, consider a drought-resistant variety that minimizes yield losses in years of low rain but is otherwise the same as other varieties. Adoption increases expected yield, but if the farm survey takes place in a year with good rains, no benefit is observed. If the survey takes place during a drought year, the yield gain is observed, which the researcher might mistakenly generalize as a benefit that is realized annually. A similar problem applies to livestock vaccines, where inoculation could insure against devastating herd losses due to contagious diseases.[3] But if the risk of disease outbreak in the region is relatively low, then even a years-long evaluation of outcomes could find no benefit to vaccination. Little can be done about this problem—if shocks are required for the benefit of the technology to manifest itself, and

[2] We do not specifically address in this paper issues concerning CGIAR research on best practice or policy.

[3] For an example of impact evaluation of livestock vaccines, see Catley et al. (2009).

the shocks do not occur, then there is no way to estimate its impact (absent a well-understood insurance market that prices production risk).

3. *Quality-improving technologies*. These technologies result in outputs that are of higher quality in some respect, even if yield does not improve. Perhaps the best example of such a technology is quality protein maize (QPM). Improved sweet potatoes (see Low et al., 2007) provide another illustration. This class of technologies differs from the previous two in that the main benefits accrue to consumers.

   The impact of quality-improving innovations is difficult to evaluate, in part because the channel of transmission from the availability of the new variety to the manifestation of benefits involves several actors. "Adoption" by consumers requires that producers have themselves adopted and produced the variety so that it is available to consumers, and that consumers have chosen to consume it.

   We can think of two polar cases. The first is when the commodity can be clearly identified and there is an effective demand for what is now a differentiated product. With market forces at work, the commodity will command a higher price than the unimproved variety. An obvious way to estimate the economic valuation of quality gains in a commodity is to see how the price of the improved variety compares to the traditional one. With knowledge of the supply and demand curves, one can calculate welfare gains from its introduction.

   The second case is where market failures may prevent the internalization of quality differences into prices, either because the product is not visually discernible from unimproved varieties and no labeling system exists to differentiate it on local markets, or because the potential beneficiaries do not command an effective demand. The challenge then is to put into place a supply chain and induce consumer demand, before we can even think of a strategy to evaluate the impact of the development of the new variety.[4] In the meantime, however, one can focus on one piece of this chain of causality and measure the impact of the "availability of the new variety" to the consumers. This requires a research approach that focuses on these consumers as the unit of interest, rather than on producers, and the outcome of interest will be measures of nutrition or health, for example, rather than monetary values.[5]

   One example of a quality-improving technology whose benefits accrue primarily to consumers is QPM, which controlled experiments have shown to have nutritional benefits (see Gunaratna et al., 2010). The current state of development of impact analyses is to randomize the supply of QPM to consumers, initially providing food to the children directly (which makes it more of a biological experiment), but now more often supplying

[4] The variety for example needs to be at least as good for the producer as the traditional one, probably even better in some way so that adoption emerges by itself. Consumers have to be aware of and willing to consume the commodity.

[5] Note that not all consumers stand to benefit from the enhanced variety: for example, only the under-nourished population will benefit from a nutritionally enhanced variety, so that definition of the proper target population is necessary.

the household with grains, therefore avoiding the issue of uptake at the household level, but maintaining the behavioral components in the use of these grains that affect the impact. A further step would be to offer labeled QPM for *purchase* by households at various prices, in order to estimate QPM's effect when consumers face an adoption decision. Much of the discussion in this paper can be applied to such an exercise by considering consumers as the unit of analysis instead of producers.

4. *Technologies that alter environmental externalities*. New cultivation and livestock management techniques may fall into this category, as well as fertilizers. These are differentiated from technologies that improve or maintain plot-level soil quality in that they prevent negative externalities on neighboring property or public resources, for example through groundwater contamination.

   Potential roadblocks to successful impact analysis for these technologies are fairly obvious. Very little of the effect of the technology can be observed at the level of the adopter. The impacts on public resources can be hard to measure, and such impacts could take a long time to manifest. Yet, without taking into account these external effects, the social value of the technology can be vastly under-estimated. Indeed, it has frequently been found that CGIAR returns from investing in technologies that alter environmental externalities has been low.

2.2. Short-run microeconomic versus long-run and aggregate effects

The goal in performing an impact analysis of a technological innovation or intervention is to estimate the total effect of the new technology on some set of outcome variables, after some amount of diffusion has taken place (Maredia 2009). Maredia lays out the steps that existing impact evaluations have pursued to estimate this total effect. Here we summarize her exposition and maintain her notation. In Maredia's framework, two key quantities must be estimated in order to arrive at the total impact of a new technology: the *extent* of adoption (E_c) and the *average effect* that adoption has on outcomes for those who have adopted (E_s). For example, E_s may be the average increase in annual profits per hectare for a farmer adopting a new variety of maize and E_c may be the total number of hectares planted with the new variety. Or E_s may be the change in poverty headcount for a village that received a technological intervention and E_c the number of villages that received the intervention.

There is an intimate relationship between the process of diffusion and the appropriate average effect that needs to be estimated. Due to this dynamic process, E_s cannot be measured separately from the time and location of the adoption, and most probably not after diffusion has taken place. Thus current approaches of estimating E_s and E_c may be appropriate in some cases, but in others it may not be, for a variety of reasons.

The principal reason that E_s is not static is that general equilibrium effects related to a technology's diffusion change its impact over time. This can be extremely important for an impact analysis and the interpretation of its results. Cochrane (1979) points out that when a new agricultural technology increases output, aggregate supply of the commodity increases and prices (of a good with imperfect tradability) must fall for markets to clear. Early adopters may experience large positive impacts of the technology on outcomes such as income and profit, as their yields increase but there are few enough adopters that prices do not fall much. This is essentially a *short-run effect* of the new technology, because low levels of adoption mean that market prices have not been affected yet.

As more farmers adopt, however, the increased output may drive down economy-wide output prices enough that adoption fails to raise the profits of farmers.[6] Input prices also adjust as the new technology results in different demands for factors of production. The decline in profitability does not indicate that farmers are irrational—no (small) farmer accounts for his own adoption's impact on prices, as they simply maximize profits while taking prices as given. In the end, the majority of benefits from the technology could accrue to consumers, who benefit from lower commodity prices. Thus the *long-run/post-diffusion general equilibrium effect* in which output and input prices adjust at the macro level can be much different than the short-run partial equilibrium effect. In addition, once diffusion has taken place, even absent of general equilibrium effects, it remains difficult to find valid counterfactual non-adopters, since those that remain non-adopters are likely very different from the adopters in meaningful ways.

Keeping in mind the dynamic nature of adoption and the potential for important long-run general equilibrium effects, how should the researcher proceed in estimating the total effect of a new technology?

1. The short-run impact measure is itself interesting. It may prove a useful tool for informing researchers of the value of the technology to at least the first wave of adopters. The average effect of adoption for adopters, $E_{s,}$ is probably the most interesting. But even the extent (or lack of extent) of adoption, E_c, is informative. As we will argue, however, these results may not be extrapolated to infer total impact when adoption expands.
2. There are also cases where one can expect general equilibrium effects to be small. A seed variety that addresses a particular type of disease affecting only a certain region may be extremely important for that region without necessarily affecting the aggregate supply of the commodity in ways that generate changes in prices. Similarly, one may think of a variety that caters to one type of producers (a variety that performs well with limited input use, but would be suboptimal for farmers who have resources to use inputs). Even an important technological change that increases the domestic production of a crop may not induce price changes if the country is open, importing or exporting the commodity. In those cases the aggregate effect of the technology may be measured by the simple product $E_s \times E_c$, provided that the estimated effect size E_s corresponds to the estimated area of adoption E_c. It is still the case that finding counterfactual non-adopters is a major challenge under these circumstances.
3. On the other hand, once the diffusion process is well-advanced and prices have adjusted, it is not clear that a currently measured E_s has relevance for inferring the impact that may have occurred over the past. Nor is it clear how one would be able to observe any non-adopters that would form a valid counterfactual for measuring an impact, however large it remains. We therefore address this type of analysis in a separate section 7.

Estimating the extent of adoption (E_c) can be obtained from an adoption survey that samples the population under consideration. The principal data necessary for estimating E_c is an indicator of

[6] This is known as Cochrane's Technological Treadmill.

whether the household has adopted (if the adoption decision is binary) or a measure of the extent of adoption (if a household's adoption may be incomplete). Measuring the extent of adoption itself may present challenges, especially if it is the case, like in the use of a specific variety, that varieties are crossed, that their names vary over geographical areas, and/or that farmers may not even know the exact varieties they are growing.[8] This is however a fundamental observational problem, not an estimation or measurement problem, and hence we will ignore it as it does not pertain to the questions raised in this paper. Adoption studies abound and generally go far beyond simply estimating E_c, also attempting to study the determinants of adoption. This undertaking is complex and presents its own set of challenges.[9] And while at the end, it may use econometric methods that are similar to those we will be reviewing here, the issues raised for the identification of the determinants of adoption are not the same as those we will raise in measuring the impact of adoption[10]

Estimation of the average impact on adopters from adoption (E_s) is in general difficult and requires careful attention. Thus the remainder of this paper focuses almost entirely on this task and the complications it entails. The key challenges to be addressed are:

1. Estimating effects for the correct population: obtaining the effect of the technology for farmers that *actually adopt.*
2. Establishing causality: isolating differences in observed outcomes that are due to adoption.
3. Accounting for spillovers: including the spillovers from adoption in estimates of a technology's impact.

2.3. Impacts to be considered

The natural place to start when looking for the immediate impacts of a new technology is farm-level restricted profits.[11] It is expected profitability that drives adoption of production technologies and acts as the channel through which adoption increases producer welfare. Compare this to another, apparently simpler, measure of impact for agricultural technologies: yield. This might be an interesting impact to measure, but it does not in itself tell us how much producer welfare was affected by the technology. As Foster and Rosenzweig (2010) state, adoption can be accompanied by input adjustment by farmers, so that the positive impact of yield increases on profits could be mitigated to some extent by increased expenditures on inputs. On

[8] As Gollin (2010) says, "the current generation of improved varieties is not so easily identified. Nowadays, we are often trying to distinguish between one generation of improved varieties and a previous generation. Are farmers growing the "old" hybrid maize, or a "new" hybrid maize? … It is not clear that farmers themselves can accurately tell you what varieties they are growing. Even where they purchase seed, the nature of seed systems in Africa is such that they may not know with any accuracy what variety they are growing".

[9] See, for example, the Agricultural Technology Adoption Initiative (2010).

[10] Hence neither criticism on commonly used methods or suggestions made for designing RCT made in this paper should be directly applied to the estimation of technology adoption.

[11] In the case where farmers engage in substantial subsistence consumption, implied profits can still be measured by treating own-consumption as a sale at local market prices, because this approximates the opportunity cost of own-consumption.

the other hand, a labor-saving innovation might not change yield per hectare but instead give the same amount of output with less work, whether supplied by the farmer or hired labor. Profits account for both changes in revenues from increased output and changes in expenses from input adjustment, and in doing so give us a measure of the first-order microeconomic impact of the new technology.

Impacts on household income, expenditure, and poverty are also important to estimate because they give a measure of the extent to which the technology actually affected household welfare. These impacts may however be mitigated substantially when compared to the effects on farm profits. For example, while a new technology may have large proportional impacts on profitability, farm income from the crop in question might form a sufficiently small portion of total household income that even such large impacts have only marginal effects on household income and little chance of pulling families out of poverty.[12] Thus these measures add substantial information to estimates of profitability and may paint a different picture of the importance of a new technology.

3. Microeconomic impact analysis

3.1. Conceptual framework for adoption

Coherent impact analysis needs to view technology adoption within a conceptual framework that treats potential adopters as agents who make decisions in their own best interest. Foster and Rosenzweig (2010) point out that "adoption and input use are the outcomes of optimizing by heterogeneous agents." This optimization takes place in the presence of constraints on the budget, information, credit access, and the availability of both the technology and other inputs. Viewing adoption through the lens of constrained optimization by rational agents, households should adopt a technology if and only if 1) adoption is actually a choice that can be taken (supply exists and credit constraints do not prevent purchasing the technology) and 2) adoption is expected to be profitable or otherwise advantageous.

A simple model of adoption and its resulting effect on outcomes can illustrate this idea more precisely. What follows is a variant of the well-known Heckman (1979) selection model, where selection into "treatment" (adoption) is made by farmers on the basis of expected profitability. For now, we assume that farmers have access to the technology. There are two sets of variables that determine the expected profitability of adoption for farmer *i* at time *t*: one that is observable by the researcher (Z_{it}) and one that is not (U_{it}). The following rule characterizes the adoption decision:

$$(1) \qquad T_{it}(Z_{it}, U_{it}, \varepsilon_{it}) = \begin{cases} 1 \text{ } if \text{ } E\pi^*(Z_{it}, U_{it}; T_{it} = 1) - E\pi^*(Z_{it}, U_{it}; T_{it} = 0) + \varepsilon_{it} > 0 \\ 0 \text{ } otherwise \end{cases}$$

where T is a binary indicator of adoption,[13] $E\pi^*$ is the maximized value of a restricted[14] general expected profit function, and ε_{it} is an i.i.d. error term. When ε_{it} is zero, adoption takes place if

[12] Furthermore, exit from poverty may take place through a slow accumulation of assets due to increased profits, which would take a long time to manifest and become observable to the researcher.

[13] The adoption decision is modeled as binary here for simplicity, as well as to make it comparable with the familiar Heckman selection model and easily applicable to the propensity score matching methods that have become so important in impact analysis in recent years. The analysis can be extended to continuous adoption choices while

and only if maximized expected profits with the new technology exceed maximized expected profits from non-adoption. Larger variance in ε will cause more farmers to mistakenly adopt or not adopt in spite of expected profitability.[15] Unless ε_{it} is very important, the farmers we observe adopting are in large part those who expected the technology to be profitable.

The outcome variable (for example, household consumption, poverty status, or profits) Y_{it} is a function of observed variables X_{it}, unobserved variables V_{it}, adoption status T_{it}, and an i.i.d. error term η_{it}:

(2) $$Y_{it} = Y_{it}[X_{it}, V_{it}, T_{it}(Z_{it}, U_{it}, \varepsilon_{it}), \eta_{it}]$$

where *X* and *Z* can share elements and *U* and *V* can share elements.

3.2. Estimating the effect of adoption for adopters

The selection model in (1) and (2) shows why it is important for impact analyses to focus on estimating the effect of adoption for those who actually adopt, rather than for the entire population of potential adopters. Using the terminology of treatment effects, our interest is in the average treatment effect on the treated (ATT) rather than the average treatment effect (ATE).[16]

Looking at (1), we see that adopters and non-adopters are fundamentally different in the benefits that they would experience from using the technology. Adopters have characteristics, both observed and unobserved, that make adoption profitable in expectation. On the other hand, non-adopters refrain from using the technology because they expect to lose money by doing so. In this simple model, then, the ATT for profits is positive because it only includes adopters, while the ATE could be positive, negative, or zero because it also includes non-adopters. The outcome of interest is not always profit, but to the extent that profits are positively correlated with such measures as consumption and poverty, this relationship between the ATT and ATE can be expected to hold.

Extending this simple model, though, one can find situations in which the ATE does not necessarily give a lower bound for the (absolute value of the) ATT. When some non-adopters would have higher gains from the technology than the adopters, the ATE can exceed the ATT. Sunding and Zilberman (2001) review the literature on risky technologies and present a model in which technologies may not be fully adopted even if they raise expected profits. A hypothetical scenario in developing countries is that even though a new technology increases expected profits

---

retaining its key results. Adoption of a specific technology, such as a seed variety, can often be thought of as binary, even if the farmer does not fully adopt the technology “package” by making self-selected adjustments to inputs and farm management practices. Management and input use are endogenously adjusted by the farmer in response to seed variety adoption, which is dichotomous unless the new variety is used alongside an old variety on the same plot.

[14] The profit function is restricted because fixed factors such as land are not taken into account.

[15] Here we are setting aside the issue of risk and any other factors that make profit maximization inadequate for characterizing the adoption decision, but one can conceptualize them by replacing the profit function with a utility function.

[16] The average treatment effect is the average effect from adoption for the entire population under consideration, whether or not adoption takes placed and the effect is actually realized.

more for poor farmers than rich ones,[17] it also increases risk, which the poor are unable to insure against through insurance or credit markets. Thus the rich could be more likely to adopt than the poor despite their lower gains from adoption as they can self-insure, so that the ATE on expected profit may exceed the ATT.[18] The ATE could also exceed the ATT if those with the highest returns cannot adopt due to credit constraints or other supply constraints.

Thus while the ATE of a technology is interesting in its own right, it is not useful in the context of an impact analysis where selection into adoption may be important. The ATT—the effect of adoption for adopters—is the necessary quantity to be estimated.

3.3. Selection and the counterfactual

Adopters and non-adopters usually differ in more ways than their potential returns to adoption. They also differ in variables that determine the outcomes of interest. The obvious problem is selection bias: if the unobservable variables in U (which determine adoption) and V (which determine outcome) are correlated, then estimating (1) and (2) will give a biased estimate of the effect of adoption on the outcome. The extent of this bias depends on how important the unobservable variables are in their respective equations, as well as how strongly the unobservables determining adoption are correlated with those determining outcomes.

There are many plausible reasons that U and V should be correlated, relating to farmer and plot characteristics and also to temporal shocks. One example is farmer ability, which cannot be accounted for entirely by observable characteristics such as age and education. All else equal, better farmers probably have higher profits (so ability is in V), while they likely have higher returns to the technology because they are more savvy in implementing it (so ability is in U). In the case of fertilizer adoption, Foster and Rosenzweig (2010) use the example that good soil quality (often unobserved) increases yields regardless of fertilizer use and also increases the return to fertilizer use. Thus soil quality is in U since it affects returns and therefore influences the decision to adopt, and it is also in V since it affects the outcome (yield, income, profit) directly. Even plot-level rainfall shocks can enter both U and V if the technology is adopted after some of the season's rainfall takes place, for example some types of fertilizers or cultivation techniques. None of these examples are trivial—in fact, one or more are likely to apply to most agricultural technologies.

This discussion of the selection problem brings us to a main challenge in impact analysis: establishing the proper counterfactual group against which to compare adopters. In order to estimate the effect of a technology, it is necessary to know what the outcome of the adopting farmers would have been if they would not have adopted. The fact that adoption is the result of optimization creates a potentially severe problem with selection into treatment on the basis of unobservable characteristics, as explained above. Thus two farmers who are observationally equivalent in every way except for adoption (and outcomes) are probably *not* equivalent on unobservables. Because of this, using the observationally identical non-adopter as the "without" comparison for the adopter is inadequate. The inadequacy of "selection on observables"

[17] This could occur when the rich are already using an intermediate technology that is superior to that being used by the poor, so that the marginal gain from adoption is lower.

[18] The Agricultural Technology Adoption Initiative (2010) provides more examples of constraints on adoption.

approaches—those in which equivalence on observable characteristics is assumed to imply equal probability of adoption—will be discussed further in the review of recent impact analyses.

In order to arrive at a reasonable counterfactual group of non-adopters, it is necessary to take seriously the possibility of strong selection on *unobservables* that arises from farmers' profit maximization problem. Doing so demands that the researcher move beyond selection on observables and towards research designs that explicitly establish a plausible counterfactual group for comparison with the adopters. Careful research designs are almost certainly more difficult to plan and implement than the ubiquitous selection on observables evaluations, usually requiring advanced planning before diffusion of the technology and the implementation or identification of some mechanism that influences adoption independent of unobservable characteristics. Suggestions for such designs, as well as examples of projects in which they could be implemented, are provided at the conclusion of this paper.

3.4. *Understanding spillovers from adoption*

Thus far in this discussion of microeconomic effects of adoption, it has been assumed that adoption by a farmer affects only his own outcomes. In reality, adoption of a technology can have local impacts on the outcomes of other adopters as well as non-adopters, even in the absence of economy-wide general equilibrium effects.[19] Households interact in local factor and commodity markets in which prices and quantities can change as a result of adoption by some of the participants. Additional output due to adoption can: increase the demand for labor in the local market, potentially raising wages (if there is no excess labor supply) but almost surely increasing the level of employment, raising income for laborers; increase or decrease demand for other scarce inputs, changing their price locally and thus altering the parameters of farmers' profit maximization problems; and if the local market is not well-integrated to outside markets, lowering prices for all buyers and sellers due to increased local output. Adoption by a farmer might have spillovers for other adopters, such as providing the opportunity to learn from his experience (e.g., Conley and Udry, 2010) thereby increasing the realized return to adoption. There may also be effects for non-adopters beyond changing local prices and wages if they are directly affected by the existence of the technology. For example, a technology with negative environmental externalities could affect nearby households and farms.

The existence of spillovers may be an important consequence of the diffusion of a technology, so spillovers need to be included in estimation of impacts. At first it might appear preferable to estimate the ATT separate of the spillovers; after all, the stated goal of impact evaluation so far has been to find the effect of adoption for adopters. In the presence of spillovers, however, it is necessary to qualify this statement. The quantity we truly want for E_s is $(ATT + average\ spillover)$, because this gives the average effect of the technology when it is taken up by actual adopters.

Spillovers from adoption complicate the necessary task of identifying the no-adoption counterfactual. Even if the counterfactual group is defined perfectly in the absence of spillovers, introducing spillovers between the adopters and counterfactual farmers will invalidate the

[19] There are other interesting spillovers as well, such as economy-wide price effects and the effects of adoption by one farmer on the adoption decision-making of other farmers. Here, we are focusing only on spillovers that affect microeconomic outcomes such as income and profits.

comparison between them. This is the well-known problem of control group contamination. The control group is supposed to represent the outcome in the absence of the technology, but if adoption indirectly changes outcomes for the control farmers, then their outcomes no longer reflect the counterfactual of no-adoption. This can lead to either over- or under-estimation of the technology's impact. If spillovers between adopters and the counterfactual non-adopters are positive, then the technology will appear less effective because the control group will be better off than it would have been in the absence of adoption. If the spillovers are negative, then the technology will appear more effective. Failing to account for spillovers will not lead to an estimate that bounds the true effect. Thus it is necessary to account for spillovers explicitly in the research design. In some cases it may lead to changing the unit of observation from, for example, individual producers to villages.

3.5. Dynamics of adoption

When estimating the average impact of a technology on adopters (E_s), it is important to understand that E_s varies over time, even if we disregard long-run general equilibrium effects. Because of this, the timing of the evaluation has an effect on the technology's estimated impact and needs to be considered carefully. The extent to which the dynamics of adoption matters is dependent upon the characteristics of the technology, particularly its pattern of diffusion and the importance of learning in utilizing the technology effectively. The dynamics of adoption is important even in the short-run where economy-wide prices are unchanged by adoption, as we discuss here.

One reason that E_s varies over time is that adopters change their usage of the new technology as they use it and learn more about it. When adopters learn to use the technology more effectively, its impact on outcomes such as yield should increase. Conley and Udry (2010) provide an example of this in the context of pineapple farmers in Ghana, where farmers change their usage of fertilizer based on the results of previous efforts (both their own and those of others). We should expect the effect of the introduction of pineapple farming on profits and consumption in a farmer's first year of growing pineapple to be different from that in his second or tenth year, as he calibrates the fertilizer usage to his own plot. Thus, even if the set of adopters stays constant across years, E_s should evolve over time.

The set of adopters of a new technology almost certainly changes with time. Sunding and Zilberman (2001) make this point clearly in their review of agricultural technology adoption, which we borrow from and expand upon in this discussion. One reason for the change in adopters is that while some farmers may initially choose not to adopt, as they observe other farmers using the technology they learn a sufficient amount about it (both how to implement it and what they can expect profits to be) that they adopt in later periods. Farmers with large landholdings might adopt first because they can experiment with the technology on a portion of their land to learn about it, then expand its use later, while small-holders adopt later after learning from the large-holders. Another explanation is that high interest rates can make adoption prohibitively expensive. This could be particularly important in developing countries—all else equal, wealthy farmers who either have the most resources to self-finance adoption or can access credit at lower rates will be those able to adopt earliest. Poorer farmers may be able to adopt later, however, as they learn by observing adopters that the technology is sufficiently profitable to justify borrowing, the price of the technology falls, or lenders become more willing to finance the technology cheaply after seeing that it is profitable.

The evolving set of adopters is important because, even if farmer-specific effects of adoption never change, these effects differ by farmer. Large-holders and entrepreneurial farmers, often the first to adopt, could have the highest returns from adoption. Hence the average impact of the technology in the first year, when it is only in use by these effective farmers who are often less constrained in their ability to utilize complementary inputs, could be higher than in subsequent years when less-efficient and more-constrained farmers have adopted.

It is clear that the effect of a technology on adopters evolves over time, both because of within-individual changes in returns and the continuous adoption by new adopters with heterogeneous returns from the new technology. Impact analyses using identical methodologies, but taking place at different lengths of time after introduction, will arrive at different estimates of the technology's average impact, even in the absence of general equilibrium effects. This is because the impact of a technology is not a static measure. Rather, it is a *flow* of impacts that changes continuously. The total realized impact of a technology is the integral of this flow of impacts from the time of its introduction until the present. Estimating a snapshot of the flow using an impact evaluation may provide an adequate approximation of the technology's effects, but it is necessary to consider the dynamics of adoption and how they affect the usefulness of such estimates.[20] We return to this question in section 7.

4. Current Approaches to Microeconomic Impact Analysis: Summary and Critiques

Much of the recent microeconomic impact analysis literature, both within the CGIAR and elsewhere, has drawn on a common set of tools to estimate the effect of technological innovations. In addition to qualitative methods, which will be addressed briefly, the most prevalent methods used are experiment station or on-farm trials, selection on observables designs such as propensity score matching (PSM) and regression adjustment, and difference-in-differences or double difference (DD) analysis (sometimes combined with PSM). Because these methods are ubiquitous, it is worth looking more closely at each of them for the specific objective of measuring the impact of technological change, considering both their strengths and weaknesses.

[20] The snapshot may *not* provide a good estimate of such measures as farm profit if farmers take a loss in the first years of adoption while they adjust their farm management techniques and capital stock to best use the new technology. Negative returns from adoption in one year may be outweighed by subsequent positive returns.

4.1. Qualitative methods
Qualitative methods of technological impact evaluation are those that use tools such as interviews and focus groups, rather than quantitative data, to arrive at conclusions about the impact of a technological innovation.[21] Qualitative studies are useful because they typically elicit information directly from people affected by the new technology about likely pathways through which the technology has impacted them. This gives researchers an idea of which impacts to look for in a quantitative analysis. For example, interviews might suggest that people who adopted also hired many more laborers for the harvest, leading researchers to collect detailed employment data among both agricultural and non-agricultural households. After a causal effect has been established quantitatively at the level of reduced form equations, qualitative methods can be useful to suggest the mechanisms at play as they are not be revealed in the quantitative analysis.

Such methods, however, are insufficient to rigorously characterize the impact of a technology. It goes without saying that interview responses do not always match with the story told by the data. Also, while qualitative results might suggest that some impact is present, they cannot say how large that impact is. When the goal is to quantify impacts with any level of accuracy, qualitative methods must play a secondary role to careful quantitative analysis, informing researchers of the data they should make an effort to collect and likely impacts to focus on as they perform statistical analyses.

4.2. Experiment station and on-farm trials
An approach used widely in *ex-ante* impact estimation of new agricultural technologies, but also in *ex-post* estimation, is to pilot the technology on test plots. Typically, the new technology is employed alongside the traditional technology, either on an experiment station or on real farms in areas where the technology is being or will be used.[22] The average difference in yields between plots using the new and old technologies is taken to be the effect of adoption on yield. This estimate can then be used to make inferences about changes in farm-level profitability due to adoption as well as aggregate changes in output and surplus after diffusion has taken place.

A number of recent impact analysis studies use the results from trials to estimate the effects of a wide range of technologies. For example, Alene et al. (2009) use on-farm trial data from various sources to estimate yield gains from improved varieties of maize in West and Central African countries, then project these gains onto the amount of adopted area to arrive at total yield growth attributable to improved varieties. Laxmi et al. (2007) use both station and on-farm trials in evaluating the impact of zero tillage technology on rice and wheat yields in India, as well as on water use and other outcomes. An analysis by the Asian Development Bank (Operations Evaluation Department 2005) of a project involving the WorldFish Center estimated the impact of genetically improved farmed tilapia on yields with both types of trials. SPIA's recent publication, "Strategic Guidance for *Ex-post* Impact Assessment of Agricultural Research"

---

[21] Good examples of this work can be found in Adato and Meinzen-Dick (2007). We should point out that by "qualitative analysis" we mean methods that are not data-driven in reaching their conclusions. There is some confusion on this point because in a quantitative analysis, binary variables (like poverty status) are often called "qualitative" data. Using binary variables in a quantitative framework does not imply that the analysis is qualitative.
[22] Bellon and Reeves (2002) collect a number of recent papers elaborating on on-farm trial methods and their comparison to experiment station trials.

(Walker et al., 2008), considers the use of experimental plots for *ex-post* impact analysis to be a "good practice" in many cases.

One obvious limitation of experiment station trials is in the set of effects that it can estimate, primarily the change in yield due to the new technology. Impacts on even relatively simple measures such as income and profitability cannot be obtained without making a number of assumptions. First, assumptions must be made about input and output prices. Second, and perhaps more importantly, the experimenter must decide how to alter the other inputs (e.g., fertilizer and labor) in conjunction with using the new technology, which may not resemble farmers' solution to the profit maximization problem. Foster and Rosenzweig (2010) point out that accounting for even small changes in the use of inputs can drastically change the estimated effect of a new technology on profits.

Furthermore, experiment stations may not reflect the actual conditions faced by potential adopters. Plot characteristics such as soil quality and access to irrigation may differ substantially between the station and the farm. If the advantage provided by the new technology varies with such characteristics, then the estimated impact may be over- or understated compared to that experienced by actual farmers. Real-world farms may also be managed differently than experiment station plots in terms of farming techniques, input provision, and skillful implementation of the technology. Each of these factors is likely to affect the returns to the technology.[23]

While on-farm trials may ameliorate some of these problems to some degree, a serious drawback remains: there is typically no reason to be certain that the farmers and farms participating in a trial are representative of those who actually adopt the technology. To illustrate this issue, consider the "ideal" case of a randomized control trial where a set of households is randomly chosen to adopt a new HYV seed and another set is randomly chosen to act as a control, continuing to use the traditional seed variety. Suppose the objective is simply to measure the change in yield due to the HYV seeds, and further suppose that there are two types of farms in each group, those whose soil is well-suited to HYV seeds and those whose soil is not (i.e., with no gains from HYV use). At harvest, the measured change in average yield from HYV seeds can be expressed as $\triangle\bar{Y} = \bar{Y}_T - \bar{Y}_C$, where $\bar{Y}_T$ is the average yield on those farms receiving the HYV seed "treatment" and $\bar{Y}_C$ is the average yield on the control farms. This expression can be decomposed into the change due to HYV seed on farms with suitable (S) and unsuitable (U) soil for HYV: $\triangle\bar{Y} = \bar{Y}_T - \bar{Y}_C = p_S(\bar{Y}_{ST} - \bar{Y}_{SC}) + (1-p)(\bar{Y}_{UT} - \bar{Y}_{UC})$,where p_S is the proportion of households with suitable soil.

Once the technology is actually released, it is unlikely that type-U farmers will adopt the HYV seeds because they offer no advantage over traditional seed. If type-S farmers do adopt (we assume it is profitable to do so), then the change in yield *for adopters* is

[23] Note the contrast between the objectives and methods of the agricultural scientist and the economist. The scientist wants to hold all factors constant between treatment and control plots, including all complementary inputs. The economist, on the other hand, most certainly does not want to hold all else constant, as the endogenous adjustments of inputs and management practices due to adoption are potentially important determinants of differences in outcomes between adopting and non-adopting farmers. Failing to adjust inputs and management practices to the new technology corresponds to measuring a sub-optimal use of the new technology from the farmer's vantage point.

$\triangle \bar{Y}_S = (\bar{Y}_{ST} - \bar{Y}_{SC}) \neq \triangle \bar{Y}$. Note that this is the quantity in which we are interested (E_s) because it corresponds to the average gain realized due to actual adoption rather than the predicted gain for a random, possibly non-adopting, household. In the terminology of treatment effects, the on-farm trial gives the ATE while the quantity of interest is the ATT.

In the example, an on-farm trial would understate the yield gains due to adoption of the new technology because it includes households who would not gain from the new technology and thus not adopt. But the bias need not be downward. If, for example, the households with the highest gains from adoption were also the most credit-constrained and consequently unable to adopt, then the RCT could overstate the real-world ATT. Because the direction of the bias from an on-farm trial (whether randomized or not) is ambiguous, such an exercise is unable to give a firm lower or upper bound on the effects of the new technology under real-world conditions.

Both types of trials are unable to account for the potentially important role of spillovers arising from the introduction of a new technology. This issue will be addressed at length below.

Hence, while experiment station and on-farm trials may be useful in some capacity, especially for *ex-ante* analysis, they do not offer a reliable way to estimate the effects of a technology on simple measures such as yields, much less more complicated outcomes like profits.

4.3. Selection on observables designs

In contrast to experiment station and on-farm trials, selection on observables approaches attempts to recover E_s by observing outcomes after households have already chosen whether or not to adopt the new technology. The main problem faced by such methods is in finding an appropriate group of non-adopters with whom to compare the adopters.[24] Selection on observables designs, whether using regression adjustment or PSM, both attempt to solve this problem by assuming that adoption is "as good as random" after conditioning on some set of observable household, plot, and/or community characteristics. Returning to equations (1) and (2), this implies that after controlling for Z_{it}, U_{it} (the set of unobserved determinants of adoption) is uncorrelated with V_{it} (the unobserved determinants of the outcome variable).

A regression adjustment model (linear regression that controls for observables affecting selection) assumes that a linear combination of the observables is sufficient to control for all factors simultaneously affecting both the adoption decision and the outcome variable.[25] There is often some confusion about how much PSM relaxes this assumption. By matching adopters and non-adopters on the basis of the propensity score generated by a first-stage logit or probit model, PSM basically allows for a somewhat arbitrary *non-linear* combination of the observables to control for factors affecting both adoption and outcomes (Rosenbaum and Rubin, 1984). *It does not alter the basic assumption that the observable explanatory variables are sufficient to characterize all determinants of adoption that also affect the outcome variable.*

A plethora of recent impact analysis papers are willing to make this assumption in a wide range of evaluations. Kumar and Quisumbing (2010) use PSM to study the effect of adopting new fishpond management technologies and vegetable varieties on household-level economic and

[24] See section 3.3.
[25] This assumption is relaxed when using difference-in-differences, explained below.

nutritional outcomes in Bangladesh. Dillon (2008) uses PSM to try to control for endogenous placement of groundwater wells for agricultural irrigation in Northern Mali. Dey et al. (2009) apply PSM to investigate the economic impacts of adopting integrated agriculture-aquaculture systems in Southern Malawi. Kassie et al. (2010) compare adopters and non-adopters of improved groundnut varieties in Uganda in terms of crop income.[26]

If the assumption of selection on observables holds (and the logit or probit functional form adequately approximates the true selection equation) then PSM gives the ATT (E_s), as desired.[27] Selection on observables is clearly a strong assumption in the context of technology adoption and, of course, is fundamentally untestable. Returning to Foster and Rosenzweig's (2010) assertion that adoption (or non-adoption) is a choice that results from optimization, we can reconsider the adoption and outcome equations in the context of PSM to show how PSM can fail to establish a viable comparison group of non-adopters against which to measure changes in outcomes.

To simplify the illustration, we will suppose that there is only one observable factor, x, and one unobservable factor, u, and that these factors affect both the profitability of adoption and the outcome variable. We can also rewrite the change in profitability due to adoption as $B(x_{it}, u_{it})$ so that adoption occurs only if $B(x_i, u_i) + \varepsilon_i > 0$ (suppressing the time subscript). Finally, we assume without loss of generality that $\frac{\partial B}{\partial u} > 0$ and $\frac{\partial Y}{\partial u} > 0$, where Y is the outcome under consideration, such as yield or profit.

Suppose that there are two farmers with equal values of x, but that farmer A adopts and farmer N chooses not to adopt. Then PSM will use farmer N as the counterfactual for A since their observables are the same. But because A adopted while N did not, $B(x_A, u_A) + \varepsilon_A > 0 > B(x_N, u_N) + \varepsilon_N$. Then it must be true that $u_A > u_N$ and/or $\varepsilon_A > \varepsilon_N$. In the former case, the assumption that $\frac{\partial Y}{\partial u} > 0$ implies that if neither farmer had adopted, farmer A would have had higher expected outcome than farmer N. That is, $EY_A(x, u_A, T_A = 0) > EY_N(x, u_N, T_N = 0)$. Hence the non-adopter was not a valid counterfactual for the adopter because they would have had different expected outcomes in the absence of treatment.

Minimal assumptions about the relationship between unobservable attributes, adoption, and outcomes were necessary to arrive at this breakdown in PSM. It was sufficient for $\frac{\partial B}{\partial u}$ and $\frac{\partial Y}{\partial u}$ to be nonzero (i.e. the unobservable factor affects both the profitability of adoption and the outcome variable itself), which is not only plausible but also probable in most cases of technology adoption. Intuitively, the problem is that PSM assumes that observationally similar farmers are on average the same, even when one of them has chosen rationally to adopt and the other has not. It is highly unlikely that this is the case precisely because farmers are choosing to maximize profit or otherwise optimize some outcome. By employing PSM, we virtually *guarantee* that

[26] PSM methods are also common in studies that go beyond analysis of agricultural technologies, for example evaluations by IFPRI of community development projects (Nkonya et al. 2008) and farmer field schools (Davis et al. 2010).

[27] For this to be true, it must also be the case that there are no spillover effects between the adopters and non-adopters. This is discussed below.

there will be an imbalance of unobservables after balancing on observables between adopters and non-adopters. This is the standard selection bias problem often discussed in the context of linear regression models. If linear regression models do not solve the selection bias problem, then PSM does not either. The advantage offered by PSM is that it more flexibly models selection on *observables*; like linear regression, it does not address selection on unobservables.

Existing impact analyses rarely consider the adoption process carefully when applying PSM and do not seriously ask whether the available observational data are sufficient to characterize the adoption decision as a rational business decision in the face of constraints. Ravallion (2005) notes in the context of antipoverty programs that the performance of PSM relies heavily on the adequacy of the data collected in characterizing adoption. In his case, where the rollout of an antipoverty program is still highly incomplete, selection is made from among households that were not offered the option to participate to the program, hence the selection bias does not occur. In the case of technology adoption, it is difficult or even impossible to collect enough data to reasonably predict adoption. For example, Kumar and Quisumbing (2010) predict adoption of new fishpond management technologies and vegetable varieties using farm size, household composition and education level, and whether various shocks were experienced during the study period. It is unlikely that these factors exhaust the true determinants of adoption (keeping in mind the decision as one of profit or utility maximization), such as land quality, farm characteristics, available assets and credit access at baseline, and farmer skill. Indeed, it is unlikely that many of the important determinants of adoption could be collected or quantified even if significant monetary resources were available to the researcher.

It is difficult to imagine that farmers decide whether or not to adopt technologies in a way that is largely random in relation to farm- or household-level outcomes. Yet unless the adoption equation is strongly predictive of the adoption decision, we are left to believe just that. For this reason it is important to know the strength of the adoption equation, for example its pseudo-R^2, in order to know if the observable variables adequately predict adoption. Many impact analysis studies do not report this statistic, while those that do (e.g. Kassie et al., 2010) tend to indicate that the adoption equation is quite weak, leaving much of the decision attributable to unobserved factors.

A final technical note on PSM is in order. Many studies rightfully “trim” the adopting and non-adopting observations in order to ensure overlap of the propensity score between the two groups. Ravallion (2005) makes an important point on this subject: if trimming the dataset results in the dropping of some adopters (i.e., those with the highest probability of adoption), then the resulting estimated impact of adoption is not the true ATT. Adopters with the highest propensity score may be those with the highest gains from adoption, in which case trimming them from the sample means that the benefit of adoption for adopters (the ATT) is understated. Of course this does not suggest that researchers employing PSM should not trim their dataset; the problem is that those with exceptionally high propensity scores simply do not have a valid counterfactual counterpart, even on the basis of observables.

4.4. Difference-in-differences methods

A growing and welcome trend in impact analysis is the use of difference-in-differences methods using panel data. The advantage of using DD is that it allows for the researcher to control for the time-invariant characteristics of individuals or households when comparing adopters and non-adopters of a technology. This weakens the key assumption required for the validity of comparisons between adopters and non-adopters: single-difference (cross-sectional) approaches require that, after controlling for observable characteristics, the two groups would have the same expected outcomes in the absence of adoption. DD methods instead require that after controlling for observables, the *change* in expected outcomes between the pre- and post-adoption surveys would be the same in the absence of adoption.

Due to the increased data requirements of DD analysis, few technology adoption analyses have used it. An early example of using longitudinal data in impact analysis for agricultural technologies is Walker and Kshirsagar (1985), which uses two waves of surveys to study the effects of adopting machine threshing technology in India. Dillon (2008), mentioned above, uses DD along with PSM to investigate the impact of irrigation from wells on agricultural production, household consumption, and nutrition in Northern Mali. Also mentioned above, Kumar and Quisumbing (2010) apply DD with PSM to study the consequences of adopting new fishpond management technologies and vegetable varieties in Bangladesh. Rusike et al. (2010) also use PSM alongside DD in investigating adoption rates (not the impact) of new varieties of cassava in Malawi. Finally, Omilola (2009) attempts to apply DD to a dataset where only one wave of data was collected; the "baseline" data are in fact constructed by asking farmers retrospective survey questions. The goal of that paper is to see if tubewell adoption in Nigeria decreased poverty. DD methods are more widely applied in program evaluation contexts, where evaluation is more likely to be planned in advance and where the intervention may be better-defined than the introduction of a new technology.

While expanded use of DD in impact analysis would be a positive development and certainly no worse than single-difference methods in controlling for selection bias into adoption, DD methods do not eliminate the need to think carefully about the adoption decision and ways in which adopters may differ from non-adopters. It is not *a priori* obvious that outcomes should be evolving similarly for those who choose to adopt and those who do not, even after considering observable characteristics.

For example, consider that more innovative and entrepreneurial farmers may be those who adopt a new technology. Such characteristics will not be recorded on a survey and are not necessarily highly correlated with observable characteristics. If these farmers are in general the most successful farmers in their village, we might expect them to be increasing their yields and profits at a faster rate than non-adopters even in the absence of the new technology, as they continually improve their farming practices and possibly adopt other technologies. Thus DD estimates would falsely attribute these increases to adoption, when in reality they are due to the fact that the yields and profits of more able farmers follow a different trajectory than less able farmers.

As another example of where DD could fail, suppose that adopters of a new technology have plots that are more sensitive to rainfall shocks than non-adopters. If during the follow-up survey there has been a drought that affected the entire sample of farmers, then adopters will have lower yields due to their greater responsiveness to the rainfall shock, but this effect cannot be disentangled from the effect of adopting the new technology.

With these illustrations in mind, it is clear that DD does not solve the potentially serious issues of selection bias involved in technology adoption and that adopters and non-adopters need not follow parallel trends in outcomes in the absence of adoption. One way to test for the validity of the parallel trends assumption is to use multiple years of pre-adoption panel data to see if the two groups are following similar trends prior to adoption. While finding parallel trends does not guarantee that the trends would be the same during the post-adoption period, it does provide a compelling piece of evidence that this may be the case. Of course a long panel dataset requires significantly more data collection than a simple DD approach, so it is may be difficult to implement in many cases.

One lesson from the existing impact analysis studies using DD is that in order for such a study to be convincing, it needs to be planned in advance of the introduction of the new technology so that proper baseline data can be collected. Using retrospective data from questions asked during the post-diffusion survey instead of a proper baseline survey, as in Omilola (2009), likely results in substantial measurement error of both dependent and explanatory variables. Measurement error of explanatory variables is a particularly serious problem in analyses using longitudinal data (such as DD) and can result in estimates that are greatly biased toward zero. Likewise, collecting baseline data after the technology has already been adopted fails to capture the full effect of adoption. This is the problem that Kumar and Quisumbing (2010) face because their baseline surveys occur up to several years after adoption, a problem that they discuss in detail. Many of the adopters in their sample had probably already experienced many of the gains from adoption by the time the baseline data was collected, so the difference between baseline and follow-up survey outcomes probably understated the effects of adoption.

Designing the impact evaluation prior to rolling out a new technology can prevent these problems by planning baseline survey collection ahead of time, as is currently done with program evaluations. Of course, if it is possible to plan a technology rollout and its evaluation, there are probably better evaluation methods available than DD. These will be discussed in section 5.

4.5. Addressing spillovers from adoption

The previous discussion has focused primarily on the limitations of popular econometric methods in establishing the proper counterfactual group against which to compare adopters. An additional and closely related issue faced by most impact analyses is that of spillovers from adoption, both on other adopters and on non-adopters. Section 3.4 addressed the issue of spillovers in detail. Here we will discuss how the existence of spillovers affects current impact analyses and the ways in which it may affect the conclusions in such work. Spillovers are not simply econometric issues that must be addressed with improved methods, but rather fundamental consequences of technology adoption that must be considered carefully.

Most impact analyses (including nearly all that have been mentioned in this section) compare adopters with non-adopters within the same village or set of villages. Even supposing that the researcher successfully creates a valid counterfactual group for the adopters among the non-adopters (i.e. the two groups would have the same outcomes in single-difference models or the same change in outcomes in DD models), the existence of spillovers can result in incorrect estimates of the impact of adoption. Miguel and Kremer (2004) make this point clearly in the context of de-worming drugs. While the drugs indeed had large impacts on rates of illness and other outcomes, failing to account for the fact that treated students ceased infecting untreated students would cause one to conclude erroneously that the drugs had had not effect.

A similar issue likely afflicts existing impact analyses of agricultural technologies. To illustrate using a recent evaluation, consider Omilola (2009). Suppose for simplicity that within a village, people randomly choose whether or not to adopt a new tubewell or pump technology, so that we can ignore selection bias in adoption. If usage of these technologies for irrigation yields a larger harvest for adopters, then there are a number of ways in which this increase could affect non-adopters. First, the larger harvest may increase demand for labor, driving up wages. This would increase income for non-adopters above the case in which nobody adopted. Second, if the market for agricultural products is restricted to a small geographical area, the increased output will drive down prices faced by all producers and thus lower farm income. Third, if there is some form of mutual insurance in the village, income gains accruing to adopters may be shared with non-adopters. Fourth, non-adopters may have an opportunity to make use of the adopters' wells or pumps for their own farms, even though they have not adopted the technology themselves. This could raise output and income for non-adopters. Certainly there could be other channels through which non-adopters are affected by others' adoption as well.

Many of these spillover channels could lead to the author's conclusion that the effects of adopting these irrigation technologies are small, even if the effects are in fact large. When adoption by some also benefits the non-adopters, the gains from the new technology are understated because it appears that the adopters would have been quite well-off even without the new technology. On the other hand, if non-adopters suffer due to adoption by others, as in the case of prices being driven down, the new technology may appear more beneficial than it really is. Indeed, spillovers invalidate the use of the non-adopters as the counterfactual for adopters because they no longer represent what the adopters would experience in the absence of the technology.

The problems posed by spillovers are mitigated if adoption takes place at the village level, similar to the village-level irrigation programs studied in Dillon (2008). In this case there is probably not much spillover from adopters to non-adopters unless there is substantial inter-village interaction, which the author says is not a concern. If we ignore the selection bias issue, then a comparison of adopters to non-adopters is valid, but another problem exists. Random shocks, particularly rainfall, are often clustered at the village level. It is necessary to account for this intra-village correlation in statistical comparisons between adopters and non-adopters. Doing so can raise the standard errors of estimates substantially, especially if the number of clusters (villages) is small. In the case of Dillon's irrigation study (which does not appear to account for clustering in its computation of standard errors), there are only ten villages, a number low enough that precise comparisons between them that also account for spillovers are unlikely.

Laboratory and on-farm trials might seem to offer an advantage by strictly controlling the behavior of the control group to prevent spillovers from adopters. But, aside from the problems with these trials explained above, it is worth noting that we do *not* want to eliminate spillover effects when studying the effects of a new technology. Spillovers, both on adopters and on non-adopters, are a consequence of adoption in the real world and thus it is important to incorporate these spillovers in the calculation of the technology's impact. By ignoring spillovers—by design—laboratory and on-farm trials fail to reflect the true impacts of adoption.

5. Suggested Approaches and Improvements

5.1. General recommendations

Current approaches to microeconomic impact analysis suffer from two main weaknesses: problematic formation of the counterfactual non-adopting group and failure to account for spillovers between adopters and non-adopters. Addressing these key shortcomings is a formidable challenge that cannot be overcome in the space of one paper. Our recommendations below provide a step toward that goal.

It is evident from the above discussion that current approaches rely almost exclusively on some form of selection on observables and that such a strategy will rarely result in convincing results. The broad suggestion made here is that, whenever possible, impact analysis should be based upon micro-studies with explicit research designs that estimate the effect of the new technology without relying exclusively on the observable characteristics of potential adopters. Optimally, such research programs should be planned in advance of the technology's introduction and diffusion. While such programs are not easy or inexpensive to implement, they allow for the application of econometric techniques for which the underlying assumptions are clear and relatively mild. Thus the results stemming from their implementation will be more credible and more likely to withstand scrutiny than those from *ex-post* evaluations relying on strong and usually unreasonable assumptions.

5.2. Approaches to avoid

Recent papers, both academic and policy-oriented, have focused on the potential application of RCTs to impact analysis of agricultural technologies. While the use of RCTs has significant potential to add rigor to future analyses, and indeed much of the rest of this section discusses this possibility, it is important to point out applications of RCTs that would not be fruitful either because they fail to overcome the problems with current methods or because they introduce new issues that undermine their usefulness.

1) There has been much interest in the work of Duflo et al. (2008) in conducting on-farm trials in which participating farms had one small plot randomly allocated to not using fertilizer and two others to using pre-determined amounts of fertilizer. The paper has appeal in part because it uses a simple method of randomization (at the plot level) to estimate the gains from the new technology, removing the possibility of bias from farmers selecting certain types of land into adoption. The objective of this experiment was to demonstrate the value of an appropriate amount of fertilizer, and this method is for that purpose perfectly appropriate.

   Such a method is, however, not advisable for the *ex-post* impact analysis of agricultural technologies. The fundamental problem with applying this approach to *ex-post* analysis[28]

[28] Duflo et al. (2008) is not a paper about impact analysis in the sense that we are discussing here. The objective of this experiment was to demonstrate the value of an appropriate amount of fertilizer, not to measure the impact of fertilizer adoption. Thus the "problem" mentioned is with using this methodology for the task at hand, not with the methods as they are applied in the paper, where they are indeed appropriate.

is that it does not estimate the effect of the technology for actual adopters (ATT), but rather the average effect (ATE) over an arbitrary set of farmers and pieces of land. As a result, there is little or no improvement over the on-farm trials discussed earlier in this paper. Restricting the sample of farmers to those who would normally adopt is not possible unless adoption is totally determined by observable characteristics, in which case an RCT would not be necessary because adopters and non-adopters could simply be compared after diffusion by using PSM. Furthermore, even if the set of adopters were known, the plots of land they would choose for adoption might differ from those selected for planting in the RCT.

If it were certain that real-world adoption would only take place among those farmers in the sample with the highest returns to the technology, then the RCT might give a lower bound on the ATT because it includes farmers with low enough returns to deter adoption. But non-adoption could take place for a number of reasons (see Agricultural Technology Adoption Initiative, 2010) such that the estimated effect of the technology is not a lower bound. For example, if the farmers with the highest returns from adoption are also the most credit-constrained, then the ATE may underestimate the true ATT. Thus the result of a within-farm randomized design fails to bound the effect of the new technology.

An RCT where technology is randomized within a farm may fail to induce farmers to act as they would if they had optimally chosen to adopt the technology. Farmers who have a small test plot assigned to a new technology may have little incentive to take the time to implement the technology carefully. NGO staff or extension agents who advise the farmers do not solve the problem, as this may induce behavior different than that exhibited by actual adopters. Inputs, both variable and fixed, may not be optimally adjusted on the treated plot.[29] Thus even the ATE may not be estimated correctly. Hence this approach has potential issues with both internal and external validity that make it an untenable option for impact evaluation.

2) One strategy that has been mentioned for moving away from on-farm trials to a design that allows farmers to select into adoption is to offer a demand-side intervention that induces some farmers to adopt an available technology, for example an encouragement design that pushes a randomly selected set of farmers to adopt. The basic idea behind encouragement designs is to use receiving the encouragement as an instrumental variable for adoption, then use two-stage least squares to obtain the effect of adoption on the outcome of interest.[30] Other examples of demand-side instruments are random extension of credit to some farmers or randomized price subsidies through coupon distribution.

   Demand-side interventions will not recover the ATT of a new technology. Regressions using an instrumental variable obtain the local average treatment effect (LATE) of adoption, which is the average effect of adoption only for those induced to adopt because

[29] Foster and Rosenzweig (2010) note, too, that changes in inputs allocated specifically to the test plot might be difficult to measure.

[30] Encouragement designs are common in medical research, e.g., Hirano et al. (2000). Bradlow (1998) has a clear explanation of encouragement designs along with an application to marketing.

they received the treatment (Imbens and Angrist 1994).[31] This makes demand-side instruments undesirable for two reasons. First, they do not estimate the impact of the technology for farmers who would have adopted even without the intervention. These inframarginal adopters probably differ substantially from those who only adopt due to receiving the treatment; they may have higher returns to the technology and thus rationally adopt without the treatment, or they may have lower returns but be less constrained in some way that allows them to adopt without having received the treatment. Second, the estimated effect is, by construction, only for farmers who would not adopt in the real world *sans* intervention. We know this because the LATE measures the difference in outcomes between marginal adopters due to treatment and farmers in the control group who are like them in every way except treatment status but who choose not to adopt. Hence, a demand-side instrument does *not* estimate the effect of adoption for any real-world adopters but *does* estimate the returns for some real-world non-adopters. This is a problem if we believe that technology adoption is the result of optimization by farmers with respect to the expected gains from adoption.

To illustrate this point further, we present a graphical representation of a simple encouragement design taking place within a village where the technology has just been introduced and is available to all farmers, and all farmers underestimate the benefits of adoption. Half of the farmers receive extension services explaining the technology's benefits. We assume that farmers are risk-neutral expected profit-maximizers so that they adopt if and only if it increases their expected profits. We also assume adoption is a binary decision (yes or no). Figure 1 plots supply and demand for the technology. For simplicity, supply of the technology (S) is assumed to be perfectly elastic, e.g., offered at a uniform price set by the government.

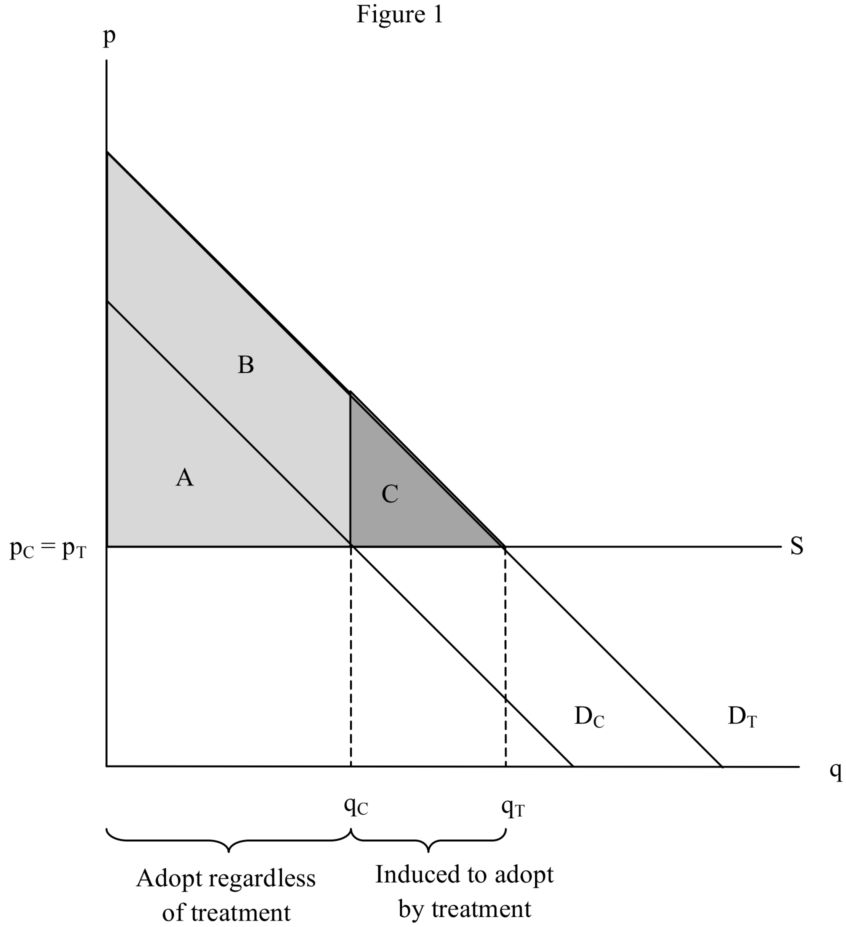
The demand curve for the control group is denoted D_C. If adoption is binary, each point (q, p) on D_C indicates the number of control group farmers (q) for whom expected profit gross of the technology cost is at least p. The encouragement intervention can be thought of as raising the expected benefit of adoption, which we assume (only for convenience) equates expected with actual profitability. Thus treatment shifts the demand curve for the treated group to D_T.

Of the control group, q_C farmers adopt. Of the treatment group, q_T farmers adopt, a higher number than in the treatment group. The important feature of this graph is the location on the demand curve of the farmers induced to adopt by the treatment (between q_C and q_T). These are farmers who have low net profits from adoption. Consider the implications if the researcher's goal is to estimate the true impact of the technology on profitability in the real world (ATT). The correct measure of this is $(A + B)/q_c$, which is the average profit from adoption for farmers who adopt without any demand-side intervention. But the LATE from two-stage least squares is $C/(q_T - q_C)$, the average profit from adoption for those induced to adopt by the intervention. The graph shows clearly that the LATE is

[31] A necessary condition for the LATE to be valid is that treatment not have an effect on the control group. This would be invalid in a demand-side intervention if, for example, the increased demand for a technology increased its price and caused some farmers in the control group to not adopt.

much smaller than the desired ATT because the LATE only measures the effect for the farmers for which benefits are the smallest.

Figure 1



3. Experiments (natural or controlled) in which the randomization occurs at the level of the household or plot are unlikely to result in reliable impact estimates. Randomizing at the plot level gives the problems discussed in point 1 from applying methods like those in Duflo et al. (2008). Randomizing at the household level has other problems, discussed here.

   The most serious issue with randomizing over households (instead of over villages, for example) is that adoption by some in the treated group will probably affect the outcomes of the control group. Such spillovers invalidate comparisons between treated and control groups as a basis for estimating the effect of the new technology. Sections 3.4 and 4.5 explained the problem of spillovers and how it affects current selection on observables approaches. The problems described in the latter section carry over even to otherwise well-planned RCTs where randomization of technology over households is explicit.

   To illustrate, we return to the example of tubewells in section 4.4. Suppose that instead of comparing adopters and non-adopters on observables, there were an RCT that took place at the time the technology was introduced. A group of farmers was selected for treatment in which they were visited by extension agents, told about the new technology, and offered the necessary materials for sale. The control group was not visited. Baseline data was taken before any adoption and then a follow-up took place two years later to see how farm profitability and household consumption changed.

Randomization does not solve any of the spillover problems discussed in the selection on observables case: wage effects, local price effects, mutual insurance effects, and usage of the wells by non-adopters. Any of these could be serious enough to limit the validity of the experiment. Spillovers within the village are a consequence of adoption itself, not the research design being used. They will exist regardless of the researcher's approach.

A further issue with randomization at the household level is that, in many cases, farmers in the control group may gain access to the new technology and choose to adopt it even though it was not offered to them, a problem often referred to as contamination of the control group. This is particularly likely if the new technology is a farming technique rather than a physical input that is purchased. Adoption by control farmers is a problem because the estimated LATE no longer gives the effect for those who were induced to adopt by the offer of the technology compared to those who would have adopted if given the treatment. Instead it measures the effect for those induced to adopt compared to a mixture of adopting and non-adopting households. Dropping the adopting households in the control group from the sample prior to analysis does not fix the problem, because these may have indeed been the farmers that correspond to the adopters in the treated group. In this case, the treated adopters would be compared to control farmers who would never adopt—the wrong counterfactual group for estimating the LATE.

5.3. Specific suggestions

Having cautioned against several potential new approaches for impact analysis, we now present suggestions for future work. The main purpose of these suggestions is to help in clearly identifying a counterfactual group against which to compare adopters, accounting for inevitable spillovers from adoption, and limiting contamination of the control group.

1. *Use natural*[32] *or randomized experiments where the village, community, or other appropriate social grouping is the unit of randomization.* Doing so addresses the issue of spillovers not by ignoring them or trying to create an environment in which they do not exist, but rather acknowledging that they are potentially important results of adoption and incorporating them into the measured effect of the new technology. Randomizing at a level higher than the household has recently become standard practice among development economists when spillovers may be present. In the field of health, for example, Miguel and Kremer (2004) randomize drug treatment at the school level and Cohen and Dupas (2010) randomize mosquito net prices at the clinic level. In education, Muralidharan and Sundararaman (2009) and Kremer et al. (2009) randomize teacher and student incentives (respectively) at the school level.

[32] By "natural experiment" we mean a situation where assignment of the treatment is as good as random, possibly conditional on some observable variables, but that no specific attempt was made to randomize the treatment as in an RCT. The rollout of a technology over time and space might constitute a natural experiment in some cases.

When randomization takes place at the village level (for example[33]) in an RCT, two-stage least squares can still be performed using households as the unit of analysis, provided that the standard errors are clustered at the village level.[34] Clustering typically makes standard errors much larger and necessitates the inclusion of many villages in the experiment, which increases costs and leads to a much more expansive project. While this is an unfortunate consequence, the alternative of household-level randomization gives a (potentially seriously) biased estimate of the effect of the new technology.

When randomization takes place at the village level and individuals are the unit of observation, regressing the outcome of interest on treatment status gives the average effect of the treatment being offered, regardless of whether the treatment induces adoption. This is known as the intention-to-treat effect (ITT). The ITT is a useful quantity because it tells the average impact of the treatment per household regardless of adoption status, which can be used as a measure of the effectiveness of the *treatment*. The ITT accounts for all effects of the technology, both directly through adoption and through spillover effects on adopters and non-adopters.

The LATE scales the ITT by the fraction of adopters, so it measures the effect of adoption *per adopter*. Because the LATE is just the scaled ITT, it still includes the effect of spillovers. Provided that the treatment induces adoption for all farmers who would adopt under real-world conditions while not inducing adoption for real-world non-adopters, the LATE gives the desired quantity for impact analysis of a technology: the ATT plus the average spillover effect per adopter.

In the case of an RCT, clustered randomization also ameliorates, to some extent, the ethical dilemma presented by Maredia (2009) where some farmers are offered a technology while it is withheld from others in the same village. Withholding treatment from some villages is a less artificial act than it is from households, as the number of treated villages is likely to be determined by budget constraints and treating any of the control villages is infeasible. Of course, data collection must still take place in the control villages, so if the marginal cost of distributing the technology during surveying is low, then the artificial withholding of treatment is still an issue.

A further benefit of including many villages in an evaluation is that the effect of the technology is estimated using several geographically distinct locales. This adds a degree of external validity to the results, as the estimates are based on the results of adoption across locations with potentially heterogeneous effects from adoption. For example, if the new technology is only useful when rainfall is scarce (such as a drought-resistant seed variety), then measuring the effects of adoption only in a village that experiences plentiful rain will not reveal the true returns of adoption. A study covering many villages,

[33] We will use the village as the level randomization in this discussion, for convenience only. The discussion applies to other levels of randomization as appropriate in the context of the technology and location.

[34] See Cameron and Trivedi (2005) for a technical but approachable review of clustered standard errors, which is beyond the scope of this paper.

however, could observe a whole range of rainfall levels and obtain an estimated effect closer to the true quantity.

2. *Use supply-side interventions where the new technology is introduced to entire villages.* Village-level clustering is not sufficient to recover the desired $(ATT + average\ spillover)$ estimate because it does not ensure that adoption due to the treatment corresponds to those who would take up the technology under true market conditions. For this reason, the treatment should simulate the introduction of the new technology on the market as closely as possible. The simplest example of this in the setting of an RCT is to choose a set of villages that do not have the technology and randomly choose a subset of villages in which to sell the technology at the "market price." Adopters are then the entire set of farmers who find it optimal to adopt when the technology is available for purchase, i.e., those who purchase under real-world conditions. The LATE is as desired: the effect of the technology, $(ATT + average\ spillover)$, when the technology is made available for sale.

   It is important to note that not all supply-side interventions yield the correct LATE. If the product is already available in villages and the intervention is a subsidy to sellers that shifts supply of the technology outward, then the analysis suffers from problems similar to those discussed with demand-side treatments. Only the effect on marginal adopters is estimated. Thus it is important that the supply-side intervention be the relaxation of a supply constraint so that villages with no access to the technology are given access at market prices.

3. *Do not limit research designs to RCTs. Natural experiments can yield reliable estimates of impact even in the absence of controlled, explicit randomization*. There are situations in which an RCT is a plausible option that should be pursued. If a new technology must be rolled out over time due to supply constraints, randomizing the villages receiving the technology at each phase of the rollout may be a simple and feasible task that provides precisely the supply-side RCT necessary for estimation of the technology's impact. This can be done by matching pairs of villages based on observables, then randomly drawing treatment within each pair. Care must be taken not to "sabotage" the technology by introducing it to areas in which it is likely to be a failure and lack take-up—such a strategy could damage the reputation of the technology and inhibit its successful expansion.

   In other cases, it is not possible to plan and carry out an RCT. There may still be opportunities for good impact analysis, however, but some creativity is required. Rollouts of a technology that were arguably random, even if they were not explicitly randomized, can be analyzed in a similar way as RCTs. The assumption of randomness in the rollout cannot be fully tested, so it is important that researchers with institutional knowledge of the technology and its rollout process provide guidance on how the rollout occurred. In other cases, the rollout follows explicit rules based on observables that can be used to instrument treatment. Provided that the necessary data are available, treating the rollout process as a natural experiment allows for the analysis of technologies that have already been diffused. It often allows use of very large administrative databases. This is a distinct advantage over RCTs, which are only possible for new technologies.

Other kinds of natural experiments may be usable as well. A geographic discontinuity approach may be possible when a specific area is chosen for diffusion of the technology. Provided that the boundary defining who receives the technology is not physically or politically important such that villages on either side of the boundaries are very different from each other, these two groups of villages can be compared, with lying on the "diffusion side" of the boundary used as the indicator of treatment. The natural experiment is that, since the boundary is assumed to be arbitrary, lying on one side of the boundary or the other is essentially random for villages near it. Regression discontinuity methods might be applied in order to account for differences between the two sides of the boundary due to continuous changes in village characteristics over space.[35] However, here again, we only measure local treatment effects in the proximity of the discontinuity rule.

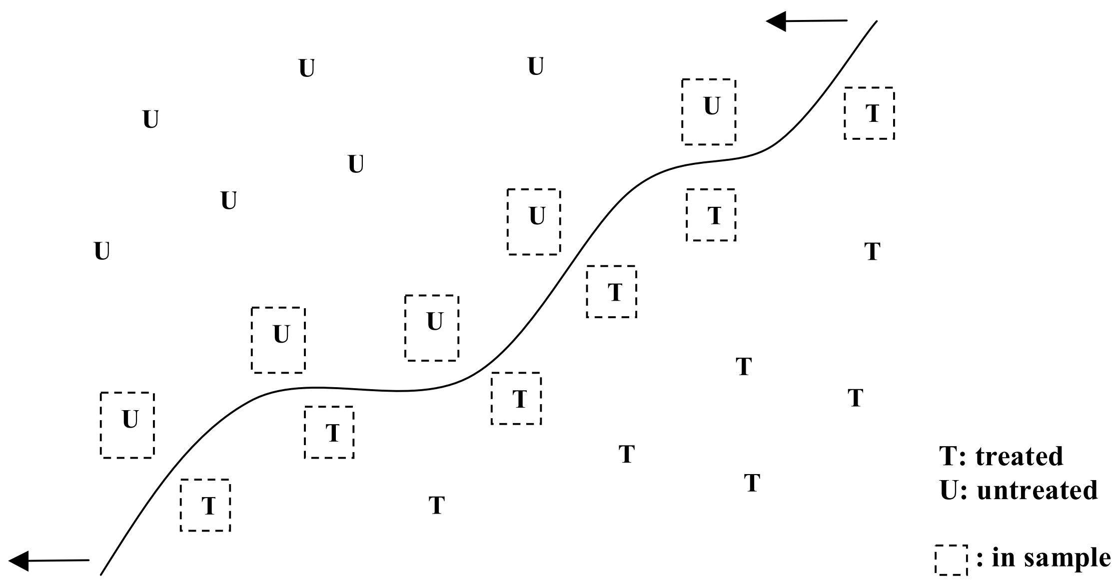
An example in which a boundary discontinuity design might be successful is the introduction of technologies that control *striga hermonthica*, a pest plant that has spread widely in Africa and suppresses yields in affected areas (Berner et al., 1994). Striga infestation spreads from afflicted areas by wind, livestock droppings, and sale of contaminated seed at markets. In an infested area, there may be a geographical frontier between those areas that are infested and those that are not yet, with the frontier advancing from year to year. Suppose that researchers map the infestation frontier and collect baseline data on crop yields (and other outcomes), prior to the introduction of a new striga control technology (whether chemical or biological) for sale to farmers.

Figure 2 gives a stylized illustration of the geographic discontinuity design that could be used in this context. The "T" villages are those affected by striga where the new technology is offered for sale. The "U" villages are still unaffected because the infestation has not reached them yet. The boxed "U" and "T" villages form the treatment and control groups under consideration, because they are close enough to the boundary that they are probably comparable along all dimensions except that striga has not yet reached the "U" villages.

The effect of the technology can be estimated with a difference-in-differences method that uses introduction of the technology as an instrument for adoption, then compares the change in yield for the baseline-infested (treated) areas to the baseline-uninfested areas (control). The intuition is that in the absence of infestation, changes in yield over time between the treatment and control villages would be similar. Then the DD estimator recovers the average change in yield due to adoption, because the uninfested area is an appropriate counterfactual. Note that while no randomization of treatment is necessary for this evaluation, it is necessary to have baseline data for both treatment and control areas, as well as follow-up data after adoption has occurred.

[35] Imbens and Lemieux (2008) provide a useful guide to regression discontinuity designs. An important caveat of all discontinuity designs is that they estimate the LATE for those close to the boundary, which might not be applicable far from the boundary.

Figure 2



boundary of infestation

Needless to say, the list of possible research designs is longer than RCTs, arguably random rollouts and geographic discontinuities. The key is to remain creative in thinking of sources of random variation in supply of the new technology. These unique research designs will be case-specific and require intimate knowledge of the technology's introduction and subsequent diffusion. Brainstorming involving economists as well as staff involved closely in the technology's development and release will provide a good approach to find these opportunities.

Whether using PSM survives as a viable strategy is dependent on the details of the technology diffusion process. The only situation in which PSM is obviously suitable is when availability of the technology in villages is as good as random after conditioning on observable characteristics of the village. This does not seem likely, but if the researcher can justify such an assumption, PSM could be a useful approach.

4. *Leverage public-private-civil society partnerships to perform supply-side interventions*. The likely best-case RCT is random introduction of the new technology into villages. A potentially attractive means for doing this is to pursue partnerships between the originator of the technology and those entities that are already on the ground distributing the technology, whether private dealers or NGOs. In the case of seeds, partnerships with local agro-dealers could provide a fruitful collaboration for introducing new varieties to selected areas. Proximity agro-dealers already provide agricultural inputs in many villages so they are a natural channel for distributing new varieties to farmers.

An RCT including agro-dealers in the supply chain would need to do two things with respect to distribution of the new seed variety. First, it would have to make the new variety available to a random subset of dealers. Second, it would have to provide adequate incentives for dealers to actually buy the seeds and offer them for sale to farmers in their territories. The former should be simple in most cases, as agro-dealers already form part of the supply chain. The latter might be accomplished by subsidizing the wholesale price of seeds for dealers.

An advantage of releasing the new technology through agro-dealers is that farmers are already used to buying from them and can have some confidence that supply will remain available in the future (as opposed to obviously randomized interventions where it is unclear whether the new technology will continue to be offered for sale in future years and growing seasons). As well, prices are set according to market forces because the dealer has a profit motive. This makes observed adoption more reliable than interventions that push new technologies by methods outside of the traditional supply chain. After all, once the rollout is complete, regardless of the method, it is likely that dealers will be the only sustainable suppliers. Incorporating them into the process at the evaluation phase is a natural way to accurately estimate the effects of the new technology.

5. *Plan the evaluation before, and conduct it during, diffusion of a new technology.* The fact that impact analysis is referred to as *ex-post* should not suggest that they should be planned and performed after the fact, a point made clearly by Maredia (2009). Baseline surveys that accurately reflect pre-adoption outcomes must be undertaken prior to diffusion, and these may take some time to carry out. It goes without saying that an RCT requires much advanced planning prior to rollout of the technology, but even if the rollout is not explicitly randomized, any follow-up surveys taking place during the rollout must be ready for administration.

   Failing to plan the evaluation ahead of time has multiple negative consequences. The first is that it may result in a lack of appropriate baseline data for pre-adoption characteristics of farmers and villages. The second is that it may cause the researcher to miss a chance to implement a clear research design. The last is more subtle. It may be that there is a push to evaluate technologies that are perceived to already have been successful. By planning and executing impact analyses even for projects that are not perceived to be successes, it is possible to obtain a better picture of the returns to the entire portfolio of projects.

6. Examples of Approaches to Evaluation

The goal of this section is to give concrete examples of impact analyses that could be undertaken using the methods discussed in this paper. Each uses a new technology that has been developed in collaboration with a CGIAR research center and proposes a way to evaluate its impact as it is rolled out. It should be noted that these proposals are for new technologies that have not yet been completely diffused. This is not a coincidence: the best applications are those for which evaluations take place as the technology is rolling out.

6.1. Example 1: genetically improved farmed tilapia (GIFT)

The Worldfish Center, along with other organizations, has spent decades using selective breeding to develop "better" farmed tilapia. The fish "seed" are distributed through public-private partnerships in many countries, mostly in Southeast Asia. Adoption has been most successful in the Philippines and Thailand. Distribution in Bangladesh has been logistically difficult, resulting in very little adoption. Adoption in Vietnam has had some success, but there appears to be room for more growth.

Previous impact evaluation, summarized in an Asian Development Bank report (Operations Evaluation Department 2005), has mostly relied on experiment station and on-farm evaluations of differences in yield (weight at harvest) and surveys of fish farmers for information on profitability. The report states that the yield numbers are controversial, highly debated, and not carefully peer-reviewed.

Because the ADB report mentions explicitly that there have been supply-side constraints in Bangladesh, an RCT there may be appropriate for assessing the impact of GIFT on such outcomes as yield, income, farm profits, and poverty. While the estimated impacts would be specific to Bangladesh, the results may be applicable to some extent to countries where diffusion has already been successful. The first step of the evaluation would be to identify a random subset of villages or communities that are engaged in tilapia farming and randomize them into treatment and control groups. Then a baseline survey of household and farm characteristics would be conducted.

In the case that public-private partnerships are viable in Bangladesh, aquacultural supply dealers in or near villages selected for treatment would be offered the GIFT seed and monetary incentives to sell the new variety. This would both make supply available to dealers and ensure that they actually offer the GIFT for sale so that farmers have a chance to adopt it. If private partnerships are infeasible but NGO or governmental involvement is strong, these groups could offer extension services that sell GIFT in the treated villages. Control villages would experience no changes. After a period of time long enough for farmers to purchase, use, and realize the benefits and costs of the new variety, a follow-up survey would be conducted.

For the empirical analysis, the dependent variable would be the change in outcome (yield, income, profit, consumption, etc.) between baseline and follow-up surveys. The variable of interest, adoption of the GIFT variety, would be instrumented by a variable equal to 1 if the village was offered GIFT seed and 0 if the village was a control. The estimated coefficient on adoption would then give the effect of adoption on adopters (ATT), plus any spillovers induced by adoption.

It is not clear (to the authors) how much learning-by-doing there is in tilapia farming or how fast adoption would take place. If these are thought to be important factors, then follow-up surveys could take place over multiple years to estimate the path of adoption and outcomes over time, but only if the control group did not obtain the GIFT technology in the meantime.

6.2. Example 2: goat parasite treatment

In the past decade, the Australian Centre for International Agricultural Research (ACIAR) and International Livestock Research Institute (ILRI), among others, undertook a program to educate

Filipino farmers about the serious problem of parasites in goats, and to teach them strategies for preventing mortality from parasites. This is documented in a report by ACIAR (Montes et al., 2008). After researching optimal procedures and technologies for preventing and fighting parasites, they performed outreach activities in the form of intensive "farmer livestock schools" funded by national and local governments. Farmers invested their time and effort into these classes as well as purchasing de-worming drugs and investing in on-farm improvements. The focus was on non-chemical means of control, partly because drug-resistant worms are becoming a serious problem.

Previous impact analysis has attempted to estimate the ATT through a case study with very few animals. This ATT does not necessarily reflect the impact of the methods because it is not clear that it represents the outcomes under actual ranching behaviors. These benefits were then projected onto the population by using the estimated adoption rate in the regions included in the program.

The technology being evaluated is a package of livestock management techniques taught by the schools. A simple impact evaluation of this technology would select a group of villages or communities where goats are raised, then randomly offer field school classes to a subset of them. This should be done only in parts of the country that have not yet been exposed to these schools, as the program initially took place only in two regions but would be suitable for other areas of the country as well. After the classes conclude and sufficient time has passed for farmers to fully implement their new techniques, follow-up data could be collected on farm-level outcomes.

An analysis that randomly offers classes in new regions would have two effects of interest. First is the reduced form effectiveness of the schools themselves in changing outcomes. This can be obtained by regressing outcomes of interest (profitability, herd mortality, etc.) on the presence of a field school. The second effect of interest is from an IV regression using the school's presence as an instrument for farmers' adoption of the new techniques.

An obvious concern with this approach is the potential for spillover of knowledge between communities where the schools are offered and those where there are no schools. The seriousness of this problem depends on how geographically distant the communities are and how much they interact. It does not seem that the spread of parasites between different farmers' herds is an important issue. Even in the presence of spillovers of this sort, an RCT would still be useful. If there is no spillover, then the IV regression yields the pure ATT of the new techniques, while if there are spillovers, the IV regression gives the ATT plus the impact of the spillovers from adoption.

6.3. Example: drought tolerant maize varieties

The Drought Tolerant Maize for Africa (DTMA) Project is a major ongoing effort by CIMMYT and IITA to introduce drought tolerant (DT) seed varieties throughout Africa.[36] The DTMA claims that the gains from introduction of DT varieties will lead to yield advantages of up to 34% over improved but non-DT varieties (La Rovere et al., 2010), with up to 50% advantages during drought seasons. This is an *ex-ante* estimate from field trials, so it is important to evaluate such claims under real-world adoption as the program progresses and diffusion takes place.

[36] For more information about DTMA, see http://dtma.cimmyt.org.

A straightforward RCT supplying new seed varieties to a random subset of agro-dealers could be used for such an evaluation. Indeed, given that the project is still at a relatively early stage and supply is still severely constrained, this could prove to be an excellent candidate for an RCT evaluation. There is an additional approach that may be useful, though, and we illustrate it here as an example of using a natural experiment for the evaluation of technologies.

Drought tolerance is a risk-reducing technology designed to limit yield losses in times of drought rather than increase yields in times of adequate rainfall. As discussed in section 2.1, evaluating this technology requires that drought actually occurs for some of the farmers involved in the evaluation. Our suggested approach will be to compare DT adopters in areas with equal levels of drought risk, but for which some experienced drought and others did not. A difference-in-differences estimator that controls for *ex-ante* drought risk can be used to obtain estimates of the yield effects (and other outcomes) for actual adopters. In this case, the randomization required for identification in a natural experiment comes from rainfall shocks.

The first step of this evaluation would be to perform a baseline survey of areas where adoption of DT varieties is likely to be high when it becomes available on the market. This would be followed by a follow-up survey of the same farmers after sufficient time had passed for substantial adoption to occur. Researchers would then identify the subsample of farmers who had adopted new DT varieties made available by DTMA. The sample should include farmers within a similar agroecological zone (although the evaluation could include comparisons within several zones) but who are spread far enough apart that there will be variation in realized rainfall and drought between farmers. The sample would then be further limited to adopting farmers who experienced drought at baseline. This is an important requirement for identifying the effect of the DT varieties.

The next step would be to compile fine-grained historical rainfall data for the areas in the sample, as is now extensively available. This data would be used to estimate the drought risk for each farmer in the dataset.

The econometric strategy is to compare adopters with similar levels of drought risk but for whom realized drought outcomes differed. This can be done either with OLS or with PSM. Indeed, this is a good example of the proper use of PSM, since matching between drought and non-drought farmers would take place on the estimated probability of drought. The key assumption for this analysis to be valid is that risk of drought is random after conditioning on our constructed measure of predicted drought. After matching farmers on the basis of drought risk, the difference-in-difference estimator can be used to see how the change in yield over baseline differed by drought status. The average of this difference is the yield advantage of the technology under drought conditions.

This estimate can then be multiplied by the average probability of drought to obtain the expected annual yield gain from adoption *for adopters*. A similar method could then be used to estimate impacts for profits, income, and poverty status, provided that sufficient data were collected. Adoption effects could then be disaggregated on the basis of drought risk. The effect for farmers with relatively low drought risk might be compared to those with high risk to see if the realized gains during drought seasons are the same. Stratifying the matching on other household characteristics such as gender of household head or education level could give effects for different subpopulations, giving a richer view of the distribution of the technology's impacts.

There are important caveats to this research approach. Adopting the technology and lowering drought vulnerability for maize may lead farmers to re-optimize their production plan and farm activities, possibly planting more of other, riskier crops since their maize production risk has fallen. Since both the adopting farmers experiencing drought and not experiencing drought will have engaged equally in this re-optimization, such an effect would not be identified by looking at differences between the two groups of adopters. Similarly, if the DT variety has a different yield even in times of good weather, this change cannot be identified separately from the time trend in yields for adopters. The estimated adoption effect here would only tell us how much adoption of DT varieties alters the drought-no drought spread in outcomes. Changes in production behaviors and good-weather yield could be examined casually by comparing changes in crop composition and labor supply for adopters with those of non-adopters using difference-in-differences, but the validity of this analysis would depend on the assumption that in the absence of DT adoption, the adopters and non-adopters would have changed these choice variables in the same way and that yield trends for the non-DT variety were similar between adopters and non-adopters.

The key aspect of this example that makes a natural experiment feasible is that rainfall and drought are basically random after conditioning on the past history of rainfall. The fact that the technology's benefits are activated randomly allows us to compare adopters to other adopters. The key, here, is that we are not matching on the basis of a choice variable, but rather on an arguably exogenous variable, drought risk.

7. Long-term and aggregate effects

7.1. *The challenge of estimating long-term and aggregate effects*

The econometric methods so far discussed can be used to establish the impact of technology adoption on outcomes at the producer level. These outcomes can be of different types (yield on the plot, production on the farm, welfare of the household, or labor demand), but in all cases we are measuring impact on the units of observation that were used in the statistical analysis.[37] These units could possibly be a village, but will never be very big by the simple fact that statistical analysis requires among other things a very large number of observations. In addition, such impact can only be measured when the diffusion of the technology is not so complete that one cannot find counterfactuals unaffected by the technology. So impact analyses can only be done before sector-wide or economy-wide effects have taken place if the implementation or even the data collection require some design.

On the other hand, we are often interested in measuring *ex-post* the aggregate benefit of a technology that has diffused over large areas. In this case, the overall impact of the technology should capture the changes that occurred in aggregate supply, demand, and price in the sector. And if the sector is large enough, with spillovers on the input markets, employment, and income effects, the impact analysis should also include general equilibrium effects. There is of course no counterfactual situation that can be observed, and hence researchers will resort to different

[37] With similar methods one can measure the impact at the consumer level, on health for example.

types of analyses. One of them is to focus on smaller units of observation (such as villages) on the presumption that markets are not well-integrated so that they each represent a small "economy" and rely on econometric analysis of the observations over time to identify the causal effect of an uneven development of technological change on these units. The second type of analysis is to resort to simulation models to extrapolate impacts measured at the micro-level (most often increases in yields) to the level of aggregate effects.

7.2. *Estimating the effects of technological change with long panel data*

This is best illustrated by Foster and Rosenzweig's body of work on the effects of the Green Revolution in India (Foster and Rosenzweig, 1996, 2003, 2004). Starting in the mid-1960's with the introduction of new hybrid seed varieties, the Green Revolution led to significant improvement in crop yields over a long period of time. An important aspect of the Green Revolution experience is that it progressed in different parts of the country at a different pace, creating the opportunity of analyzing its effects in a panel setup. A simplified model that captures the essence of the methodology for measuring the impact of yield improvement on household or village level outcomes is written as:

$$Y_{ivt} = \beta yield_{vt} + X_{ivt}\gamma + \mu_i + \nu_t + \varepsilon_{ivt}$$
$$Y_{vt} = \beta yield_{vt} + X_{vt}\gamma + \mu_v + \nu_t + \varepsilon_{vt}$$

where Y_{ivt} (Y_{vt}) are outcomes of interest at the household (village) level, *yield* is an index of yield at the village level, *X* are control variables, μ_v (μ_i) are fixed household (village) effects, ν_t fixed time effects, and ε error terms. The yield index is a Laspeyres index of village level yield on irrigated HYV crops. Foster and Rosenzweig use a panel of about 4,000 households from 250 villages with 3 rounds of observations in 1971, 1982, and 1999.

With spatial and time fixed effects, identification of the impacts of an increase in yield comes from the differential changes in yields across villages. In some specification, the time fixed effect is state specific, focusing the identification to the differential pace of yield improvements across villages within a state. Outcomes of interest are for example agricultural income, non-farm income, or total income at the household level; and rural wage, non-farm employment, total income, or poverty at the village level.

A couple of issues are worth noting:

a) The yield index used in the equation is affected by more than technological change. This is because it results not only from technological change but also from changes in productive inputs such as fertilizer, labor, or education, all possibly influenced by technological change, but also by other factors. So there may be some concern that changes in the yield index at the village level capture those other factors correlated but not due to technology change. To circumvent this problem, Foster and Rosenzweig (1996 and 2003) recover a district-year specific technology factor from estimation of a farm level profit function, and then instrument *yield* with this technology factor and some village characteristics.

b) The estimation measures the effects of relative changes in yields across villages, but not the nationwide or state level aggregate effects of technology change. For example, the extent to which the Green Revolution brought about a large decline in prices over all India, and by this reduced poverty, is captured by the time fixed effects and not attributed to the changes in yield.

Still, this approach goes a long way toward getting at the aggregate and long-term impacts of a large technological change such as that embedded in the Green Revolution. Note, however, that the analysis is very demanding in terms of data, as it requires a large enough sample of observations representative of the aggregate area of concern, over a long period of time that covers the technology diffusion period, and detailed farm data that allow extracting the role of technological change from observed yields.

7.3. *Extrapolating micro results with partial equilibrium simulation models*

A second type of analysis uses simulation models. These simulations translate or extrapolate the (estimated/measured) microeconomic effects into some aggregate number, based on assumptions about the sector or the economy at large. They rely on functional form assumptions, assumptions on supply and demand elasticities, and strong assumptions on the functioning of the markets. Hence, in no way can these models be considered to estimate an aggregate impact in the same sense as econometric methods. On the other hand, they are a powerful way of translating micro-level estimations into plausible orders of magnitude of macro effects, if done with care and with enough sensitivity analysis.

The most common of these simulation models is the economic surplus approach, based on a partial equilibrium model of the sector in which the technological change has occurred. The idea is that technological change induces a shift in the supply curve, which in turn induces a decline in price and a new equilibrium on the market. The economic surplus is calculated as:

$$ES = pqk\left(1 + \frac{1}{2}\frac{k}{\varepsilon^s + \varepsilon^d}\right),$$

where p and q are the initial price and quantity of the commodity of interest, k is the proportional shift in the supply curve induced by technology change, and ε^s and ε^d are the supply and demand elasticities.

The key input to this simulation is of course the k factor, i.e., the direct effect of technological change on the supply curve. The factors that allow the extrapolation from k to the value of the economic surplus are the observed total output and price, and the assumed elasticities (usually drawn from some other studies).

This k factor is determined by the combination of changes in yield and in costs. Pictured on a supply curve, changes in yield are horizontal shifts, while cost reductions are vertical shifts, which can be made equivalent to proportional yield changes with the supply elasticity. Most studies focus on increases in yields. Changes in yields themselves are drawn from either field trials or observational differences in yields between crop varieties. Neither one is very satisfactory. What is needed is an estimated supply shift or yield increase that can be causally attributed to technological change. Hence the challenge is nothing less than what has been extensively described in the previous section on estimating impact of technology on yield.

This method has been extensively used in *ex-post* studies, computing time series of economic surplus based on observed output and prices, and assumed elasticities, and then aggregating the results over years to compute the aggregate effect. In many of these studies, the principal effort has been to measure the area (and output) affected by the specific technological change of interest, notably the development of certain varieties by the CGIAR (Byerlee and Traxler, 1995). Some studies run sensitivity analyses to the assumed elasticities. The method is also used in *ex-ante* studies, based on field trials for specific varieties (see Falck-Zepeda et al., 2007)

7.4. Poverty simulation

One more step has also been taken in applying a "poverty elasticity" to the calculated aggregate production increase or welfare effect to obtain an impact on poverty (Fan et al., 2005; Alene et al., 2009). The key question of course is how the poverty elasticity has been estimated, whether or not it applies to this specific context, etc. Obviously, the poverty effects of an increase in aggregate welfare depends on who benefited from the increase in economic surplus among large producers, small producers, urban consumers, etc., and what were their initial poverty levels. So it seems a bit of a stretch to conduct a simulation exercise by applying a poverty elasticity estimated in a different context to an aggregate increase in production. Poverty impact should be addressed with rigorous econometric analyses like those discussed in section 7.2.

In conclusion, the validity and usefulness of these simulations is nothing less than the validity of the elements that enter into the simulations. It is therefore critically important that the *k* factor be rigorously estimated, that sensitivity analyses be made on the elasticities in the economic surplus simulations, and that it be clear that these are useful simulations but not impact estimations.

7.5. *Computing aggregate impacts with general equilibrium simulation models*

Going beyond the sector model discussed in the previous section, researchers have used general equilibrium models for cases where the change in technology is sufficiently large that it induces effects on trade, and on output and input markets that, in turn, may induce spillover effects on other sectors (Arndt et al., 1999; Dorosh and Thurlow, 2009; Diao et al. 2010). The most commonly used type is the CGE, a general equilibrium model in which supply and demand on all markets are balanced with endogenous prices, although a few studies use the SAM multiplier approach, a Keynesian demand-driven general equilibrium model with excess supply in all markets.

CGEs essentially represent a system of markets that reach equilibrium through prices. Very broadly speaking, they consist in: (i) supply functions for each sector of the economy, derived from production models that are usually some combination of a CES aggregate in primary factors of production and Leontief technology for intermediate inputs, (ii) demand functions emanating from households (using some standard demand system), and (iii) markets that balance either with flexible prices or with quantity adjustments according to tradability. CGEs were initially developed for, and remain strongest in, the modeling of trade. Foreign goods and domestic goods are imperfect substitutes (using CES and CET functions), which also lead to specific demands for imports and supplies for exports in response to the relative prices of foreign and domestic goods. The rest of the model includes government and institutions that tax or transfer mostly with fixed shares (for example return to labor or to capital in each sector is allocated in fixed proportions across households, etc.). Sectors thus compete through their interactions on the input and factor markets. The data needed for such models are essentially a static picture of all the flows in one particular year (the SAM or social accounting matrix derived from national income accounts and an input-output matrix) from which all shares are derived and four sets of elasticities: (i) elasticity of substitution between factors in the production function,

(ii) demand elasticities (or more specifically parameters of a consistent demand system) for households, (iii) elasticity of substitution between imported and domestic goods, and (iv) elasticity of transformation between commodities for the domestic markets and exports for each sector. CGEs can have different levels of disaggregation (number of sectors, number of household types), and because elasticities are rarely estimated and most often "guessed", they face a difficult tradeoff between gaining details in shares and having to rely on an increasing number of assumed elasticities. Most CGEs are static models, although some have introduced updatings of capital stocks, labor supply, and possibly technological factors that make them sequentially dynamic, with however an additional set of assumptions needed for these updatings. In some models, there is an attempt at capturing the particularity of self-consumption (Ardnt et al., 1999), or to link micro-simulations applying the results obtained in CGEs to individual households and by this obtaining a better measure for poverty effects (Diao et al. 2010).

The advantages of CGEs are their consistency framework that forces markets and budgets to balance and their anchoring of all simulated effects on the observed initial relative sizes of sectors and commodities. However, the extent of what is assumed in these models (in terms of the way markets function, the representation of the agents that make choices, the existence of transactions costs and constraints, heterogeneity across producers, etc.) make them better instruments for discussion of alternative broad policy choices than for actual predictions of the effect of specific shocks like a technological change. A frequently cited article that uses archetype CGEs for the purpose of illustrating the different channels by which an agricultural technological change affect the rest of the economy (de Janvry and Sadoulet, 2002) is just that, an instrument meant to exhibit the different channels that are incorporated in a CGE and to show how their relative importance varies with the relative structural features of economies, the degree of openness of the economy, the substitutability between foreign and domestic goods, and the functioning of markets. Stark contrasts can then be shown to exist in the impacts of a same technological change across archetypes that represent a South-Asian or a Sub-Saharan economy. Sensitivity analysis shows qualitative results to be robust, but quantitative results cannot be taken too seriously. In the application that is the focus of this paper, namely measuring the impact of a technological change, what CGE simulations can do is basically help track, under a set of strong assumptions, the economy-wide effects of that particular change.

References

Adato, M. and R. Meinzen-Dick. 2007. *Agricultural Research, Livelihoods, and Poverty: Studies of Economic and Social Impacts in Six Countries*. Baltimore, MD: Johns Hopkins University Press.

Agricultural Technology Adoption Initiative. 2010. "Barriers to the Adoption of Agricultural Technologies in Developing Countries." Draft white paper.

Alene, A. D., Menkir, A., Ajala, S. O., Badu-Apraku, B., Olanrewaju, A. S., Manyong, V. M., and A. Ndiaye. 2009. "The Economic and Poverty Impacts of Maize Research in West and Central Africa." *Agricultural Economics*. 40(5): 535-550.

Alston, J. M., Chan-Kang, C., Marra, M. C., Pardey, P. G., and T.J. Wyatt. 2000. "A Meta-Analysis of Rates of Return to Agricultural R&D". IFPRI Research Report no. 113, IFPRI, Washington, D.C.

Arndt, C., Jensen, H., Robinson, S., and F. Tarp. 1999. "Marketing Margins and Agricultural Technology in Mozambique." IFPRI TMD Discussion Paper no. 43, IFPRI, Washington, D.C.

Bellon, M. R., and J. Reeves, eds. 2002. *Quantitative Analysis of Data from Participatory Methods in Plant Breeding*. Mexico, DF: CIMMYT.

Berner, D. K., Cardwell, K. F., Faturoti, B. O., Ikie, F. O., and O. A. Williams. "Relative Roles of Wind, Crop Seeds, and Cattle in Dispersal of Striga spp." *Plant Disease*. 78(4): 402-406.

Bradlow, E. T. 1998. "Encouragement Designs: An Approach to Self-Selected Samples in an Experimental Design." *Marketing Letters*. 9(4): 383-391.

Byerlee, D., and G. Traxler. 1995. "National and International Wheat Improvement Research in the Post-Green Revolution Period: Evolution of Impacts." *American Journal of Agricultural Economics*. 77: 268-278.

Cameron, A. C., and P. K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. New York, NY: Cambridge University Press.

Catley, A., Abebe, D., Admassu, B., Bekele, G., Abera, B., Eshete, G., Rufael, T., and T. Haile. 2009. "Impact of Drought-Related Vaccination on Livestock Mortality in Pastoralist Areas of Ethiopia." *Disasters.* 33: 665-685.

Cochrane, W. W. 1979. *The Development of American Agriculture: A Historical Analysis*. Minneapolis, MN: University of Minnesota Press.

Cohen, J., and P. Dupas. 2010. "Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment." *Quarterly Journal of Economics*. 75(1): 1-45.

Conley, T. G., and C. R. Udry. 2010. "Learning about a New Technology: Pineapple in Ghana." *American Economic Review*." 100(1): 35-69.

Davis, K., Nkonya, E., Kato, E., Mekonnen, D. A., Odendo, M., Miiro, R., and J. Nkuba. 2010. "Impact of Farmer Field Schools on Agricultural Productivity and Poverty in East Africa." IFPRI Discussion Paper no. 00992, IFPRI, Washington, D.C.

de Janvry, A., and E. Sadoulet. 2002. "World Poverty and the Role of Agricultural Technology: Direct and Indirect Effects." *Journal of Development Studies*. 38(4): 1-26.

Dey, M. M., Paraguas, F. J., Kambewa, P., and D. E. Pemsl. 2010. "The Impact of Integrated Aquaculture-Agriculture on Small-Scale Farms in Southern Malawi." *Agricultural Economics*. 41: 67-79.

Diao, X., Nwafor, M., Alpuerto, V., Akramov, K., and S. Salau. 2010. "Agricultural Growth and Investment Options for Poverty Reduction in Nigeria." IFPRI Discussion Paper no. 00954, IFPRI, Washington, D.C.

Dillon, A. (2008). "Access to Irrigation and the Escape from Poverty: Evidence from Northern Mali." IFPRI Discussion Paper no. 00782, IFPRI, Washington, D.C.

Duflo, E., Kremer, M., and J. Robinson. 2008. "How High are Rates of Return to Fertilizer?" *American Economic Review.* 98(2): 482–88.

Dorosh, P., and J. Thurlow. 2009. "Implications of Accelerated Agricultural Growth on Household Incomes and Poverty in Ethiopia: A General Equilibrium Analysis." IFPRI Discussion Paper No. ESSP2 002, IFPRI, Washington, D.C.

Evenson, R. E., and D. Gollin, eds. 2003. *Crop Variety Improvement and Its Effect on Productivity: The Impact of International Research*. Wallingford, UK: CAB International.

Falck-Zepeda, J., Horna, D., and M. Smale. 2007. "The Economic Impact and the Distribution of Benefits and Risk from the Adoption of Insect Resistant (Bt) Cotton in West Africa." IFPRI Discussion Paper no. 00718, IFPRI, Washington, D.C.

Fan, S., Chan-Kang, C., Qian, K., and K. Krishnaiah. 2005. "National and International Agricultural Research and Rural Poverty: the Case of Rice Research in India and China." *Agricultural Economics*, 33(s3): 369-379.

Foster, A. D., and M. R. Rosenzweig. 1996. "Technical Change and Human-Capital Returns and Investments: Evidence from the Green Revolution." *American Economic Review*, 86(4): 931-953.

Foster, A. D., and M. R. Rosenzweig. 2003. "Agricultural Productivity Growth, Rural Economic Diversity, and Economic Reforms: India, 1970-2000." Mimeo.

Foster, A. D., and M. R. Rosenzweig. 2004. "Agricultural Productivity Growth, Rural Economic Diversity, and Economic Reforms: India, 1970–2000." *Economic Development and Cultural Change*. 52(3): 509-42.

Foster, A. D., and M. R. Rosenzweig. 2010. "Microeconomics of Technology Adoption." Economic Growth Center Discussion Paper no. 984, Yale University, New Haven.

Gollin, Douglas. 2010. "Agricultural Productivity, Economic Growth, and Food Security." UC Berkeley Conference on Agriculture for Development-Revisited.

Gunaratna, N. S., Groote, H. D., Nestel, P., Pixley, K. V., and G. P. McCabe. 2010. "A Meta-Analysis of Community-based Studies on Quality Protein Maize." *Food Policy.* 35(3): 202-210.

Heckman, J. J. 1979. "Sample Bias as a Specification Error." *Econometrica*. 47(1): 153-161.

Hirano, K. and G. W. Imbens. 2000. "Assessing the Effect of an Influenza Vaccine in an Encouragement Design." *Biostatistics*. 1(1): 69-88.

Imbens, G. W., and Angrist, J. D. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica*. 62(2): 467-475.

Imbens, G. W., and T. Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics.* 142: 615-635.

Kassie, M., Shiferaw, B., and G. Muricho. 2010. "Adoption and Impact of Improved Groundnut Varieties on Rural Poverty: Evidence from Rural Uganda." Environment for Development Discussion Paper no. 10-11. Environment for Development, Washington, D.C.

Kremer, M., Miguel, E., and Thornton, R. 2009. "Incentives to Learn." *Review of Economics and Statistics*. 91(3): 2009.

Kumar, N., and A. R. Quisumbing. 2010. "Access, Adoption, and Diffusion: Understanding the Long-term Impacts of Improved Vegetable and Fish Technologies in Bangladesh." IFPRI Discussion Paper no. 00995, IFPRI, Washington, D.C.

La Rovere, R., Kostandini, G., Abdoulaye, T., Dixon, J., Mwangi, W., Guo, Z., and M. Bänziger. 2010. "Potential Impact of Investments in Drought Tolerant Maize in Africa." Report, CIMMYT, Mexico, D.F.

Laxmi, V., Erenstein, O., & R. K. Gupta. 2007. "CIMMYT. Assessing the Impact of Natural Resource Management Research: the Case of Zero Tillage in India's Rice–Wheat Systems." In H. Waibel and D. Zilberman, *International Research on Natural Resource Management: Advances in Impact Assessment* (pp. 68-90). Oxfordshire, United Kingdom: CAB International.

Low, J. W., Arimond, M., Osman, N., Cunguara, B., Zano, F., and D. Tschirley. 2007. "A Food-Based Approach Introducing Orange-Fleshed Sweet Potatoes Increased Vitamin A Intake and

Serum Retinol Concentrations in Young Children in Rural Mozambique." *Journal of Nutrition*. 137: 1320-1327.

Maredia, M. K. 2009. "Improving the Proof: Evolution of and Emerging Trends in Impact Assessment Methods and Approaches in Agricultural Development." IFPRI Discussion Paper no. 00929, IFPRI, Washington, D.C.

Maredia, M. K., and D. A. Raitzer. 2010. "Estimating Overall Returns to International Agricultural Research in Africa through Benefit-Cost Analysis: a 'Best-Evidence' approach." *Agricultural Economics*. 41(1): 81-100.

Miguel, E., and M. Kremer. 2004. "Worms: Identifying the Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica*. 72(1): 159-217.

Montes, N., Zapata Jr., N., Alo, A., and J. Mullen. 2008. "Management of Internal Parasites in Goats in the Philippines." Report, Australian Centre for International Agricultural Research, Canberra, Australia.

Morris, M. L. 2002. "Impacts of International Maize Breeding Research in Developing Countries, 1966-98." Report, CIMMYT, Mexico City, Mexico.

Muralidharan, K., and V. Sundararaman. 2009. "Teacher Performance Pay: Experimental Evidence from India." Working Paper no. 15323, NBER, Cambridge, MA.

Nkonya, E., Phillip, D., Mogues, T., Pender, J., Yahaya, M. K., Adebowale, G., and T. Arokoyo. 2008. "From the Ground Up: Impacts of a Pro-Poor Community-Driven Development Project in Nigeria." IFPRI Discussion Paper no. 00756, IFPRI, Washington, D.C.

Omilola, B. 2009. "Estimating the Impact of Agricultural Technology on Poverty Reduction in Rural Nigeria." IFPRI Discussion Paper no. 00901, IFPRI, Washington, D.C.

Operations Evaluation Department, Asian Development Bank. 2005. "An Impact Evaluation of the Development of Genetically Improved Farmed Tilapia and their Dissemination in Selected Countries." Report, Asian Development Bank, Manila, Philippines.

Raitzer, D. A. and Kelley, T. G. "Benefit-Cost Meta-Analysis of Investment in the International Agricultural Research Centers of the CGIAR." *Agricultural Systems*. 96: 108-123.

Ravallion, M. 2005. "Evaluating Anti-Poverty Programs." In J. Strauss & T. P. Schultz, eds., *Handbook of Development Economics* (Vol. 4, pp. 1-90). Amsterdam, The Netherlands: North-Holland.

Rosenbaum, P. R., and D. B. Rubin. 1984. "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score." *Journal of the American Statistical Association*. 79(387): 516-524.

Rusike, J., Mahungu, N. M., Jumbo, S., Sandifolo, V. S., and G. Malindi. 2010. "Estimating Impact of Cassava Research for Development Approach on Productivity, Uptake and Food Security in Malawi." *Food Policy*. 35(2): 98-111.

Sunding, D., and D. Zilberman. 2001. "The Agricultural Innovation Process: Research and Techonlogy Adoption in a Changing Agricultural Sector." In B. Gardner and G. Rausser, *Handbook of Agricultural Economics* (Vol. 1, pp. 207-261). Amsterdam, The Netherlands: North-Holland.

Walker, T. S., and K. G. Ksirsagar. 1985. "The Village Impact of Machine Threshing and Implications for Technology Development in the Semi-Arid Tropics of Peninsular India." *Journal of Development Studies*. 21(2): 215-231.

Walker, T., Maredia, M., Kelley, T., La Rovere, R., Templeton, D., Thiele, G., and B. Douthwaite. 2008. "Strategic Guidance for Ex Post Impact Assessment of Agricultural Research." Report, CGIAR Science Council, CGIAR, Rome, Italy.