FONDATION POUR LES ÉTUDES ET RECHERCHES SUR LE DÉVELOPPEMENT INTERNATIONAL

Challenges for Impact Evaluation of Agricultural Projects

Elisabeth Sadoulet
Alain de Janvry

Elisabeth Sadoulet is Professor of Agricultural and Resource Economics at UC Berkeley. Her research interests focus on agricultural technologies, microcredit, conditional cash transfers and property rights.

Alain de Janvry is Professor of Agricultural & Resource Economics at the UC Berkeley. He has conducted field research in Latin America, sub-Saharan Africa, Middle East, and in the Indian subcontinent, focusing, among other topics, on rural development and technological innovations in agriculture.

This policy brief is based on the presentation made at the annual meeting of the Agricultural and Applied Economics Association in Pittsburgh, USA, July 26, 2011.

.../...

policy brief 29
August 2011
note brève

Focus of this presentation: measuring the impact of technology adoption

The objective of this presentation is to discuss whether, in the wake of the World Development Report 2008's recommendation to increase investment in agriculture, impact evaluation can help define priorities for agricultural investment, identify its effectiveness, and assist in improving implementation?

Most important in terms of investment in agriculture are sources of total factor productivity gains. For this, the adoption of technological and institutional innovations is key. Let me consequently discuss how to measure the impact of the adoption of technological innovations on welfare or other developmental outcomes of interest.

There is currently a lot of interest in assessing the ex-post impact of agricultural technology adoption, in particular among donors to the Consultative Group in International Agricultural Research (CGIAR). Measuring this impact is essential to use a results-based approach to investing in the generation and adoption of agricultural technology, as well as to achieve accountability with donors. Yet, this is much more difficult to do rigorously than could be expected by donors.

So, let me discuss what are some of the main challenges to impact evaluation of technology adoption, and propose solutions to the difficulties encountered. I will discuss six challenges, two having to do with measurement and four with identification.

Measurement problem #1: Most technologies progress through incremental steps

The problem: Technological innovations in agriculture are rarely one-time changes. Contrary to belief, there is not one miracle Nerica seed (HYV rice for Africa), but a continuum of varieties from traditional to "new", in constant evolution. The same applies to soil conservation or water harvesting techniques where improvements come by small incremental steps. So it is very difficult to define "the" technology to focus on. And each successive incremental step may have a very small impact on outcomes.
There are some cases of quantum changes, such as the introduction of hybrid seeds in a context where farmers would normally conserve their own seeds, adoption of fertilizer use, investment in irrigation, and adoption of flood resistant new rice varieties conveyed by the transfer of a single gene.

These quantum changes are easier to measure, but most likely very partial, since most of the better aspects of technology may come in the subsequent vintages or after local adaptation has taken place. This was the case with the "Green Revolution" where there was an initial quantum jump, associated with the release of specific HYVs, but continuous improvements afterwards.

Thus, "adoption" is ill defined, the "technology" keep on changing, and the adopters and non-adopters keep on adjusting, and the last thing you want to do is freeze the producers in a disequilibrium state with the early seeds or the initial soil conservation technical package for the sake of an impact analysis.

The solution: Get a clear recognition of successive technological vintages, and measure the impact across sufficiently distinct vintages of the same technology.

Measurement problem #2: What outcomes to observe?

The problem: To satisfy donors, most impact analysis proposals aim at measuring "welfare" effects such as changes in income, expenditure, or poverty. There are two problems with this. The first is that adoption of any single even "big" technological innovation is only one element in an overall income strategy. The second is that it will take time for the household to respond, and amplify the benefits of the initial adoption to see important effects. It is consequently not a surprise that a well-implemented impact analysis of adoption of even a major technological innovation tends to observe small welfare effects in the short / medium term. This is a common disappointment for donors with the contributions of rigorous analysis as they were expecting that good analysis would show large impacts.

The solution: The alternative is to focus on household response indicators as proximate causes for potential subsequent welfare effects. This includes decisions to invest more, manage better, and increase household time allocation to the innovation. These effects can be observed shortly after adoption. This is by analogy with educational programs where we look at enrollment, achievements, or learning effects without waiting for welfare effects. That ultimate welfare effects may be modest is something that should be anticipated in back-to-the envelope simulations of expected impacts given sources of income for the household, not to induce false expectations with donors. We rarely do this, in part for our grant proposals to be competitive with other submissions.

Identification problem #1: Adoption diffuses naturally in a population. As a consequence, impacts are heterogeneous and correlated with the timing of adoption

The problem: We would like to measure the average impact on adopters based on comparison with counterfactuals. Difficulty is that this is a moving target. The impact on early adopters is easier to measure because there are still lots of potential counterfactuals (though the choice of counterfactuals needs to be done carefully), but this does not tell us the impact on adopters in general, once the technology has diffused. There are no obvious low or high boundaries. Early adopters may be the most Schumpeterian that would have done better anyway, or they may be the less constrained among those with a positive correlation with potential benefits, or both.

In addition, not all impacts of adoption can be measured. This is only feasible if general equilibrium effects are not important. What can be measured are (1) impacts on relatively early adopters, (2) substitution for an earlier variety without much aggregate supply effect, (3) technology that applies to a small group or region, or even to the whole country if it is a tradable good (but labor and land are non-tradable). Once diffusion is complete, there is no counterfactual left to measure impact.

The solution: Recognize when impact can be measured and admit when it cannot. In addition, carefully document at what stage of the diffusion process the impact has been measured and on whom as it is a changing magnitude with a changing incidence.

Identification problem #2: Adoption is a choice, invalidating selection of counterfactuals based on observables

The problem: Adoption is a choice. Some farmers may adopt despite not having been "treated", and some may not adopt despite being "treated". Households select into adoption on the basis of both observable and unobservable characteristics. If we have two observationally identical farmers and one adopts while the other does not, there is a good chance that they differ importantly on unobservable characteristics. Examples of unobservables that affect adoption are the farmer's ability and entrepreneurship, differential soil quality, and weather shocks that occur early enough to affect technology adoption. The fact that we often observe low technology uptake, even after promotion efforts have been extended to targeted farmers, suggests that there are factors we do not observe that affect farmer decisions. Furthermore, even adopters choose when to adopt, and the fact that they do not all adopt immediately suggest that adoption is related to time varying idiosyncratic characteristics or context.

As a consequence, it is unlikely that the very popular Propensity Score Matching (PSM) method will be valid, even when combined with a double difference calculation. The combined method fundamentally relies on the assumption that no time varying unobservables determining adoption are correlated with the unobservables that affect outcomes (omitted variable bias). This is of course in contrast with many social programs in which there is little "non-compliance" and we observe a high uptake.

The solution: "Selection on observables" is unlikely to hold when selection into adoption has occurred. It is in my opinion time to put to rest using PSM in adoption-impact studies. We need instead research designs that allow for unobservables to differ between adopters and non-adopters. Because adoption is endogeneous, it cannot be directly randomized. This requires some intervention that will induce adoption. This is often done on the demand side such as through randomized encouragement designs. Note however that demand side intervention will measure the impact on those that have been induced to adopt by the intervention and would have normally not adopted. It however does not measure the impact on adopters that would adopt anyway, or on adopters that adopted very early before our intervention. Hence, only when we are truly interested in the impact on this population that could be induced to adopt, or when we can assume that impact is pretty homogeneous across these different types of adopters, can we use a demand-side encouragement design. Otherwise, intervention ought to be done on the supply side, respecting "normal' conditions, such as the rollout of availability of the technology. This is the preferred method if we have either a natural experiment where rollout is not correlated with outcome, or if we can design an RCT where rollout is randomized across potential users.

Identification problem #3: Addressing spillovers from adoption

The problem: Adoption by one farmer might affect the outcomes for other farmers, both adopters and non-adopters, even without General Equilibrium changes in the economy. Spillovers occur through such effects as:

– Local employment and wage effects
– Local effects on input and output prices
– Learning-from-other effects: one's own adoption raises other farmers' returns to adoption
– Environmental externalities

These spillover effects make the design of counterfactuals more difficult. However, they are part of the benefits from adoption. The average impact of the technology is not just the average treatment effect on the treated. It is the average treatment effect on the treated plus the average spillovers from adoption. Ultimately, spillovers may be very large contributors to average impact.

The solution: Counterfactuals need to be designed not to be affected by spillover effects, while treatments must be designed to account for spillovers. This requires choosing the unit of analysis to internalize spillovers. Intensity of treatment within this unit of analysis can be varied to allow quantification of direct and spillover effects.

Identification problem #4: The difficulty of addressing the long-term aggregate impact of a technology or string of technologies

The problem: After spillover and general equilibrium effects have taken place (leaving, at the limit, no available counterfactual), impacts will be very different in both magnitude and incidence. To measure them, we need very different methods and units of analysis.

Econometric estimations require having observations on the past, before technology adoption has occurred. This is consequently data demanding and requires that systematic efforts have been made in due time to collect these data.

The most notable achievement following this approach is the work of Foster and Rosenzweig for the Green Revolution in India[1]. They conduct a panel analysis of small "economies" (villages) that have differentially benefited from technological change over time. Village and time fixed effects are used to control for much of the potential confounding factors. This allows estimating the effect of yield changes on welfare, poverty, etc. The approach is very demanding in terms of data and econometric skills. It goes a long way toward estimating the desired welfare effects, but still cannot identify the aggregate effect that applies to the whole of India as it is absorbed in the time fixed effects.

The solution: Measuring welfare effects including spillover and general equilibrium effects can be done, but needs anticipation and resources. It has been done ex-post for India given the unique richness of data for that country. With the Green Revolution in progress in Africa, with the concerted effort of AGRA, we strongly recommend that a similar concerted effort be made in data collection. It would add only a small cost to the resources already committed. It would help achieve the objectives of using impact analysis to define priorities for agricultural investment, identify its effectiveness, and assist in improving implementation.

Conclusion

- We need methodological pluralism, including use of Randomized Control Trials as one option in a portfolio of methods and not always as the superior approach.
- Each case is different. Hence, each case needs to be carefully diagnosed (through case studies, descriptive statistics) to understand in particular who has adopted and why we see non-adoption.
- Impact analysis is all in design and robustness checks.

1. 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.

- Not everything can be measured for impact. It is consequently better not do it when it cannot be done well than to do it wrong.
- Impact analysis is an art, not a mechanical application of methods.

Hence, high quality impact evaluation can be very useful to help define priorities for agricultural investment, identify its effectiveness, and assist in improving implementation, … if rigorously done when it can be done.

Créée en 2003, la **Fondation pour les études et recherches sur le développement international** vise à favoriser la compréhension du développement économique international et des politiques qui l'influencent.

Contact
www.ferdi.fr
contact@ferdi.fr
+33 (0)4 73 17 75 30