Impact Assessment: Towards Sound Methodology for Informed Decision-making in Agricultural Programs

Ninth DA Lecture Forum

27 June 2012

Dr. Dennis S. Mapa
Associate Professor of Statistics and Director for Research, School of Statistics
Affiliate Associate Professor, School of Economics
University of the Philippines Diliman

Organization of the Lecture

- Motivation for Impact Assessment
- Qualitative and Quantitative Impact Assessment
- Ex ante and Ex post Impact Assessment
- Example of Ex Ante Impact Assessment
- Methods for Ex post Impact Assessment
- Some examples of Ex post impact assessment
- Steps in Impact Assessment

Motivation for Impact Assessment

- Government and donors are keen to determine the effectiveness of programs with goals such as lowering poverty, increasing employment or increasing school attendance.

- Program evaluation are often possible only through impact assessment based on hard evidence (empirical) from survey data or through related quantitative approaches.

- Methods to understand whether such programs actually work, as well as the level and nature of impacts on intended beneficiaries, are main themes of this lecture.

Motivation for Impact Assessment

- Can the conditional cash-transfer program in the country improve health and schooling outcomes for poor women and children?

- Does a new road actually raise welfare (e.g. income) in a remote area in Eastern Samar?

- Do community-based programs of the Department of Agriculture create long-lasting improvements in employment and income for the poor?

Need for Impact Assessment

- Programs might appear potentially promising before implementation yet fail to generate expected impact or benefits.

- The obvious need for impact evaluation/assessment is to help policy makers decide whether programs are:

(a) generating intended effects;
(b) to promote accountability in the allocation of resources across public programs; and
(c) to fill gaps in understanding what works, what does not, and how measured changes in well-being are **attributable to a particular project or policy intervention**.

Need for Impact Assessment

- Effective impact evaluation/assessment should be able to assess precisely the mechanisms by which beneficiaries are responding to the intervention.

- These mechanisms can include links through markets or improved social networks as well as tie-ins with other existing policies.

- The benefits of a well-designed impact evaluation are therefore long term and can have substantial spillover effects.

Need for Impact Assessment

- Broadly, the question of *causality* makes impact evaluation/assessment different from Monitoring and Evaluation (M&E) and other evaluation approaches.

- In the absence of data on counterfactual outcomes (that is, outcomes for participants had they not been exposed to the program), impact evaluations can be rigorous in identifying program effects by applying different models to survey data to construct comparison groups for participants.

- The main question of impact evaluation is one of **attribution**—isolating the effect of the program from other factors and potential selection bias.

Impact Assessment Techniques

- Impact assessment spans qualitative and quantitative methods.

- Quantitative Methods can be done through an *ex ante* and/or *ex post*.

Importance of Qualitative Impact Methods

- Qualitative analysis, as compared with the quantitative approach, seeks to gauge potential impacts that the program may generate, the mechanisms of such impacts, and the extent of benefits to recipients from in-depth and group-based interviews.

- Whereas quantitative results can be generalizable, the qualitative results may not be. Nonetheless, qualitative methods generate information that may be critical for understanding the mechanisms through which the program helps beneficiaries.

Importance of Qualitative Impact Methods

- For example, qualitative information can help identify mechanisms through which programs might be having an impact; such surveys can also identify local policy makers or individuals who would be important in determining the course of how programs are implemented.

Example: The Role of the School Principal in the Impact of Education Television (ETV) on National Achievement Test (NAT) Result

Limitation of the Qualitative Impact Methods

- A qualitative assessment on its own cannot assess outcomes against relevant alternatives or counterfactual outcomes. That is, it cannot really indicate what might happen in the absence of the program.

- A mixture of qualitative and quantitative methods (a *mixed-methods approach*) might therefore be useful in gaining a comprehensive view of the program's effectiveness.

Quantitative Impact Assessments

- Quantitative impact methods are preferred over qualitative methods in doing impact assessments.

- Answers to policy questions are often possible only through impact evaluations based on hard evidence (empirical) from survey data or through related quantitative approaches.

Quantitative Impact Assessment: Ex Post v. Ex Ante

- There are two types of quantitative impact evaluations: ex post and ex ante.

- An ex ante impact evaluation attempts to measure the intended impacts of future programs and policies, given a potentially targeted area's current situation, and may involve simulations based on assumptions about how the economy works.

- Many times, ex ante evaluations are based on structural models of the economic environment facing potential participants.

- These models predict program impacts.

Example - Ex Ante Impact Evaluation

The Role of Agriculture in Poverty Reduction:
Uncovering the Channels Linking
Agricultural Growth and Poverty
(Balisacan, Mapa, and Fuwa (2011))

Labor Market Channel

- The Anríquez and López's (Chile's case) approach to identify the effects of agricultural growth on labor income (via employment and wage effects):

Growth in agriculture → increase in the demand for unskilled labor → increase in wage and/or in employment of unskilled labor → poverty reduction.

Food Price Channel

- Expansion in agriculture can affect poverty by reducing the prices of food.

  - first, a fall in food prices increases real income;
  - second, the cost of the food basket used in the computation of the poverty line decreases (lowers the poverty line).

Direct Income on Poor Farmers (Spillover Effects)

Expansion in Agriculture can increase the income of poor (subsistence) farmers.

Econometric Model

- A time-series econometric model using quarterly data, from 1994 to 2010, was used to determine the marginal effect of agricultural growth on food prices.

- The model specification is given by,

$$\log(y_t) = c + \beta_1 \log(RER_t) + \beta_2 \log(PINF_t) + \beta_3 \log(Q_{at}) + \beta_4 \log(Q_{nt}) + \varepsilon_t$$

$$t = 1, 2, \dots,$$

Econometric Model

where, $y_{t\text{-}}$ is the real food price index (food component of the CPI):

RER is the real exchange rate:

PINF is the real non-food price index (non-food component of the CPI);

Q_a is the agricultural output (seasonally adjusted); and Q_n is the non-agricultural output (seasonally adjusted)
c is the constant term and ε is the error term.

The real food price index was generated from the Philippines' Consumer Price Index (CPI).

Econometric Result – Long Run Relationship

Dependent Variable: LOG(FOOD)				
Included observations: 68				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.382693	1.017943	-2.340694	0.0224
LOG(NON_FOOD)	0.504096	0.098218	5.132432	0.0000
LOG(REAL)	0.178373	0.046972	3.797415	0.0003
LOG(AGRI_SA)	-0.442682	0.120364	-3.677843	0.0005
LOG(NON_AGRI_SA)	0.713421	0.148799	4.794516	0.0000
R-squared	0.985094	Mean dependent var		4.697389

long-run elasticity of food prices to agricultural output = - 0.44

compensated elasticity of food prices to agricultural output

$$\left.\frac{d\ln y_t}{d\ln Q_{at}}\right|_{dQ=0} = \frac{\partial \ln y_t}{\partial \ln Q_{at}} + \frac{\partial \ln y_t}{\partial \ln Q_{nt}} * \left.\frac{d\ln Q_{nt}}{d\ln Q_{at}}\right|_{dQ=0}$$

= - 0.44 + 0.71 (-0.20/0.80)
= - 0.6175

Simulating the Impact of Food Prices on Poverty Incidence

- Scenario 1 assumes a 3% growth Agricultural Output

- Scenario 2 assumes a 5% growth in Agricultural Output

- Uncompensated and Compensated increased in Agricultural Output were used in the simulations

- Benchmark (2009) Poverty Incidence is 26.5% (NSCB) and Number of Poor Individuals (in 2009) is about 23.14 million (NSCB).

Effects of agricultural expansion on poverty: Food price channel and direct income channel

Effect on Income and Poverty	Scenario 1 (3% Growth)		Scenario 2 (5% Growth)	
	(a)	(b)	(a)	(b)
Increase in Average Income (%)	0.60	0.83	1.00	1.39
Percentage Point Reduction in Poverty Incidence	0.29	0.38	0.45	0.61
Poverty Incidence (%)	26.01	25.91	25.85	25.68
Number of Poor Individuals (millions)	22.72	22.64	22.58	22.44
Poverty Depth (2009 Baseline: 7.2)	7.06	7.02	6.99	6.93
Poverty Severity (2009 Baseline: 2.8)	2.77	2.73	2.71	2.69
Reduction in the Number of Poor Individuals (estimate)	252,000	332,000	392,000	534,000

Scenario 1 assumes a 3% growth Agricultural Output while Scenario 2 assumes a 5% growth in Agricultural Output; (a) Uncompensated Simulations; (b) Compensated Simulations

Ex Post Impact Assessment

Ex Post Impact Evaluation

- Ex post evaluations, in contrast, measure actual impacts accrued by the beneficiaries that are attributable to program intervention. One form of this type of evaluation is the treatment effects model.

- Ex post evaluations can be more costly than ex ante evaluations because they require collecting data on actual outcomes for participant and nonparticipant (control) groups, as well as on other accompanying social and economic factors that may have determined the course of the intervention.

- An added cost in the ex post setting is the failure of the intervention, which might have been predicted through ex ante analysis.

Basic Theory of Impact Evaluation: The Problem of Selection Bias

- Successful impact evaluations hinge on finding a good comparison group.

- There are two broad approaches that researchers resort to in order to mimic the counterfactual of a treated group:

  (a) create a control group through a statistical design;

  (b) modify the targeting strategy of the program itself to wipe out differences that would have existed between the treated and non-treated groups before comparing outcomes across the two groups.

Different Quantitative Approaches to Ex Post Impact Evaluation

Randomized Trial

Randomization

- Considered as the “gold standard” of impact evaluation.

- Allocating a program or intervention randomly across a sample of observations is one solution to avoiding selection bias, provided that program impacts are examined at the level of randomization.

- Careful selection of control areas (or the counterfactual) is also important in ensuring comparability with participant areas and ultimately calculating the treatment effect (or difference in outcomes) between the two groups.

Example of Randomization – The Tennessee STAR Experiment

- Research Problem: Impact of class size on students' achievement

- The STAR (Student-Teacher Achievement Ratio) experiment was implemented for a cohort of kindergartners in 1985-85. The study ran for four years, until the original cohort was in the third grade and involved about 11,600 children.

- The average class size in a regular Tennessee classes is about 22.3.

- The project cost is about US$ 12 million.

The Tennessee STAR Experiment

- The experiment assigned students to one of three treatments: small (13-17 children), regular class (22-25) with part-time teacher’s aide, regular class with full-time teacher’s aide.

- Randomization eliminated the selection bias, the difference in outcomes across treatment groups captures the average causal effect of class size.

The Tennessee STAR Experiment

Variable	Class Size			P-value for equality across groups
	Small	Regular/Part-Time Aide	Regular/Full-Time Aide	
Class size in kindergarten	15.10	22.40	22.80	0.00
Percentile Score in Test	54.70	48.90	50.00	0.00

The STAR study, an exemplary randomized trial in the annals of social science also highlights the logistical difficulty, long duration and high cost of randomized trial.

Double Difference

Double Difference (DD)

- Also known as Difference-in-Difference (DID) method.

- The DD estimator relies on a comparison of participants and nonparticipants before and after the intervention. For example, after an initial baseline survey of both non-participants and (subsequent) participants, a follow-up survey can be conducted of both groups after the intervention.

- From this information, the difference is calculated between the observed mean outcomes for the treatment and control groups before and after program intervention.

Double Difference (DD)

- In order to control for systematic differences between the control and the treatment groups, we need two periods of data, one before the intervention and one after the intervention.

- Our sample is broken down into four groups: the control group before the intervention, the control group after the intervention, the treatment group before the intervention and the treatment group after the intervention.

- If we let A be the control group and B the treatment group, dB = 1 if the individual is in the treatment group and 0, otherwise.

- Let d2 denote a dummy variable with value 1 for the second (post intervention) period and 0, otherwise.

Double Difference (DD)

The equation of interest is,

$$y = \beta_0 + \delta_0 d2 + \beta_1 dB + \delta_1 d2 * dB + other \quad factors + \varepsilon$$

where y is the outcome variable of interest. The parameter of interest δ_1 measures the effect of the intervention.

Other controlled variables are included in the "other factors."

Without "other factors" in the regression, the estimate of δ_1 is known as the **"difference-in-differences"** estimator.

$$\hat{\delta}_1 = (\bar{y}_{2,B} - \bar{y}_{2,A}) - (\bar{y}_{1,B} - \bar{y}_{2,A})$$

Example of Double Difference (DD)

Research Problem: Impact of a feeding program on the school attendance of Grades 1 and 2 pupils in public elementary school.

Treatment Schools: 63 schools

Control Schools: 63 schools

Number of Pupil-Respondents

School Type	Pre-Test Visit	Post-Test Visit
with feeding program	2,029	2,457
without feeding program	1,605	1,600
Total	3,634	4,057

Regression Result Explaining Determinants of Absences of Pupils

Regression Model Explaining Number of Absences			
Variable	Estimated Coeff	Standard Error	P-Value
School Type	-0.541	0.152	0.000
Time Effect	4.840	0.168	0.000
Program Effect	**-1.644**	**0.214**	**0.000**
Control Variables			
Gender (Female=1)	-0.370	0.105	0.000
Age of Pupils	0.202	0.036	0.000
Mindanao Area	1.287	0.144	0.000
Visayas Area	-0.124	0.118	0.293
Constant	0.002	0.316	0.996

Regression Discontinuity Design

Regression Discontinuity (RD) Design

- In a non-experimental setting, program eligibility rules can sometimes be used as instruments for exogenously identifying participants and nonparticipants.

- To establish comparability, one can use participants and nonparticipants within a certain neighborhood of the eligibility threshold as the relevant sample for estimating the treatment impact.

- Known as regression discontinuity (RD), this method allows observed as well as unobserved heterogeneity to be accounted for. Although the cutoff or eligibility threshold can be defined non-parametrically, the cutoff has in practice traditionally been defined through an instrument.

Instrumental Variable (IV)

Instrumental Variable (IV) Regression

- Instrumental variable (IV) methods allow for endogeneity in individual participation, program placement, or both.

- The IV approach involves finding a variable (or instrument) that is highly correlated with program placement or participation but that is not correlated with unobserved characteristics affecting outcomes.

- Instruments can be constructed from program design (for example, if the program of interest was randomized or if exogenous rules were used in determining eligibility for the program).

Propensity Score Matching

Propensity Score Matching (PSM)

- Propensity score matching (PSM) constructs a statistical comparison group that is based on a model of the probability of participating in the treatment, using observed characteristics (such as income, household characteristics, etc.).

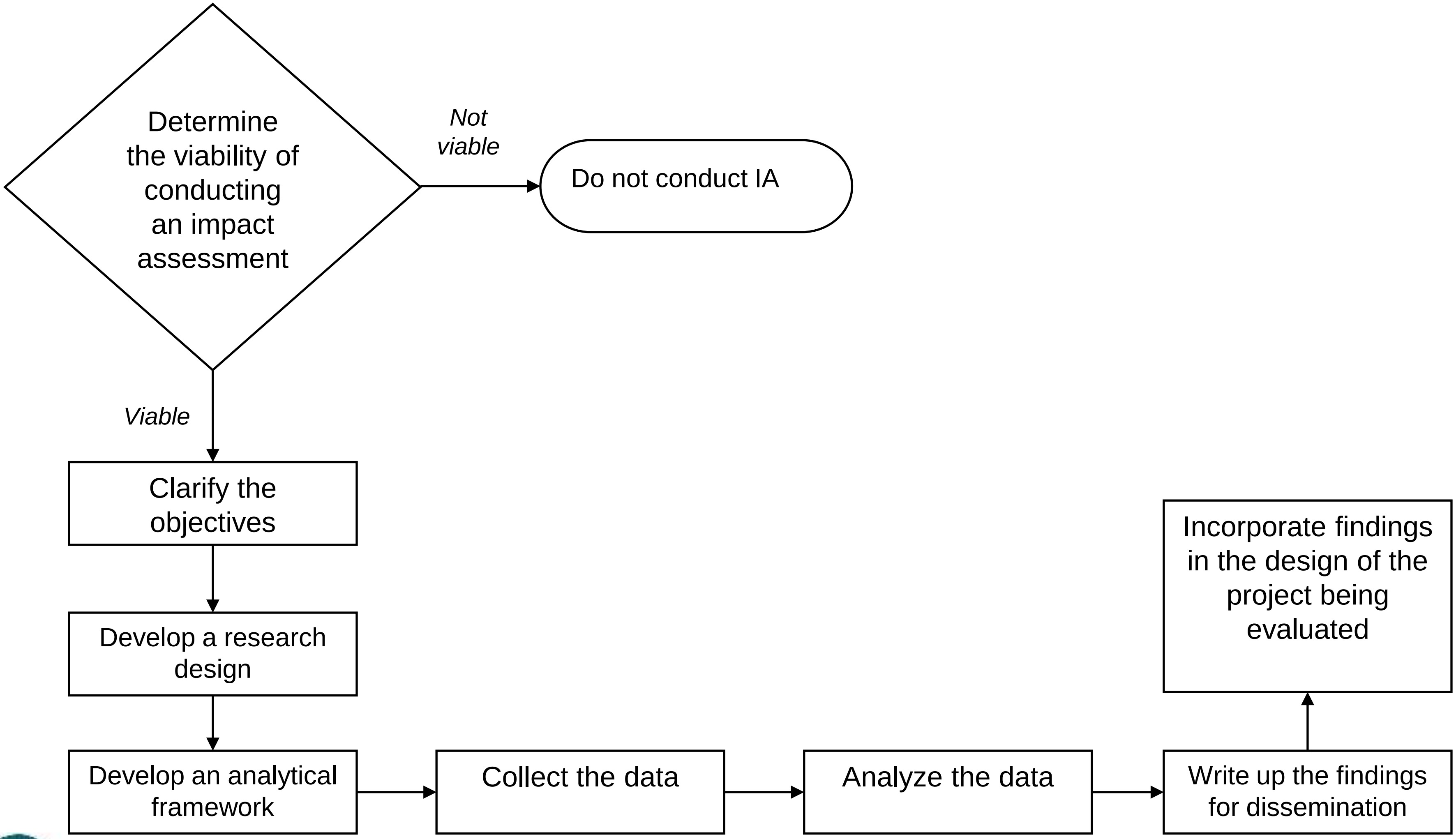
- Participants are then matched on the basis of this probability, or propensity score, to nonparticipants.

- The average treatment effect of the program is then calculated as the mean difference in outcomes across these two groups.

- The validity of PSM depends on two conditions: (a) conditional independence (namely, that unobserved factors do not affect participation) and (b) sizable common support or overlap in propensity scores across the participant and nonparticipant samples.

Conduct of Impact Assessment

The Impact Assessment Framework



Steps: Preparation Stage

- Determine viability to carry out IA
- Clarifying objectives of IA
- Exploring data availability
- Designing the IA
- Forming the IA team

Thank you.