Economics Training Series

Introductory Course

Project Impact Evaluation

ADB

Outline

1. What is project impact evaluation?
2. Methods of project impact evaluation?
3. Operational implications.

Project Monitoring and Evaluation

Adapted from presentation by Bill Savedoff, Center for Global Development.

ADB

Terms

- Impact evaluation looks at the impact of an intervention on final welfare outcomes, rather than at project outputs or at the project implementation process (World Bank). The latter cases usually are called performance evaluation.
- Counterfactual: the hypothetical outcome that would have prevailed had there been no intervention.
- "Assessment" sometimes is used but mostly in ex-ante appraisals (e.g., poverty / environmental impact assessment) and "evaluation" usually implies an ex-post study.

ADB

Qualitative Approach

- Desk studies, reviews, interviews, secondary data, etc.
- Establish causal inferences on a basis of processes (A → B → C), behaviors (incomes, expenditures, visits to hospital), perceived changes (better schools, roads), and conditions (upgraded irrigation canals, more crops). For example, Participatory Rural Appraisal.
- Drawbacks: Subjectivity involved in data collection, the lack of a comparison group, and the lack of statistical robustness.

ADB

Quantitative Approach

- The analysis is based on a counterfactual.
- Quantitative evaluations are generally regarded as more authoritative and usually referred to as rigorous. The WB now requires a counterfactual analysis to qualify as impact evaluation.
- But good impact evaluations combine quantitative analysis and qualitative information to have both rigor and supportive contextual insights.
- This presentation discusses quantitative methods only.

ADB

Why we need a counterfactual

Changes in Outcome = due to project

+

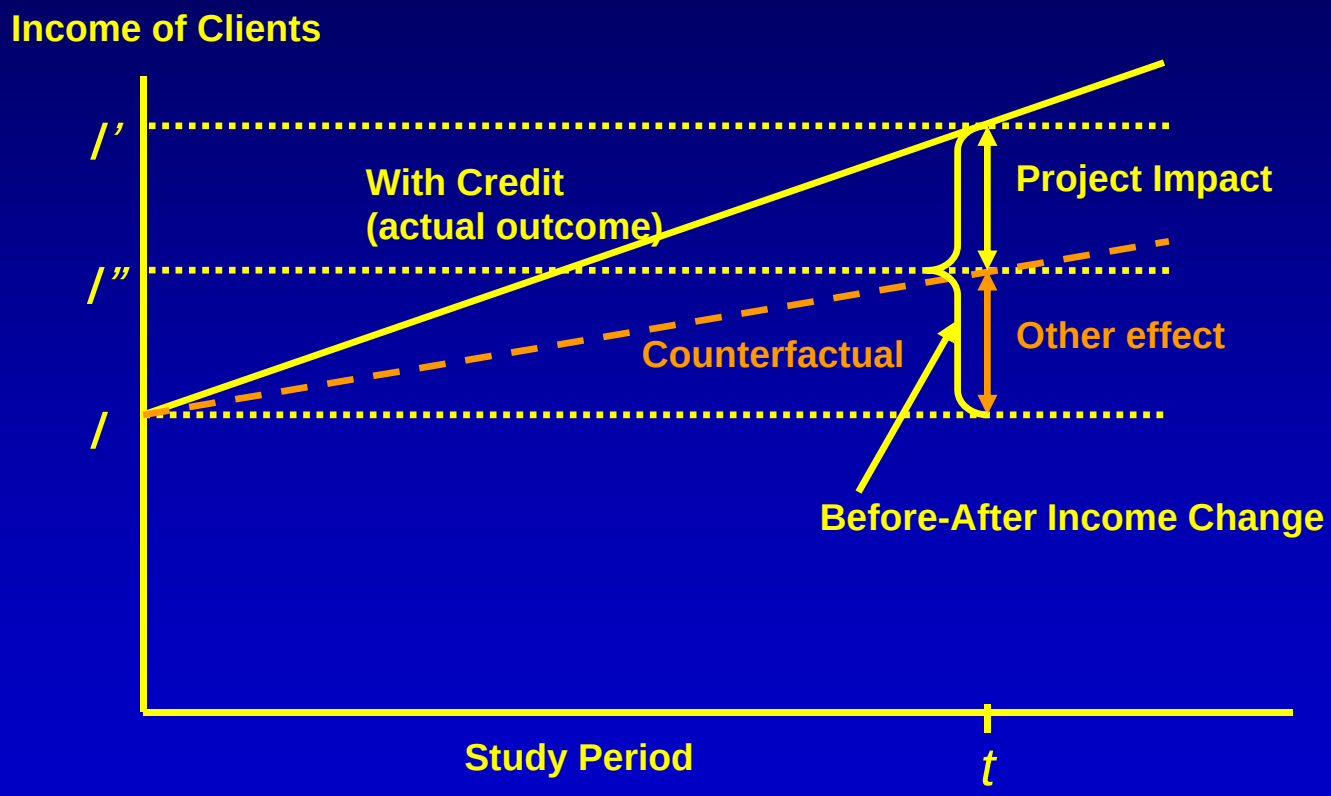
due to other factors

(environmental, personal)

- "Before" and "After": compares the same individuals before and after the project → not controlling for environmental factors.
- "With" and "Without": compares only similar but not identical individuals → not controlling for personal differences.
- Counterfactual: controls for both environmental and personal factors.

ADB

Example: Income Impact of Credit



The Problem

- The problem is that one cannot be both in and out of the project at the same time. Therefore, we need some way to "mimic" the participants. This group of non-participants that mimic the participants is called the control group. We use this control group as the counterfactual.

- All quantitative impact evaluations boil down to constructing a credible counterfactual.

ADB

Methods

Non-Experimental Methods: derive the counterfactual by statistical techniques.

- PSM and DD (IV and RDD are not presented)

Experimental Designs: construct the counterfactual by randomly assigning a group of project participants (the treatment group) and a group of non-participants (the control group).

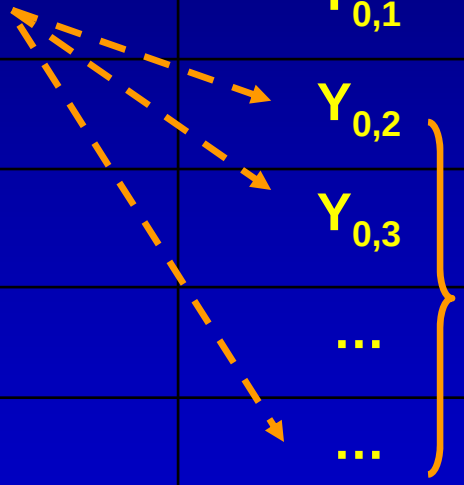
- Lottery, Phase-in, Encouragement

Non-Experimental and Experimental methods differ in the way they construct the counterfactual.

ADB

Propensity Score Matching (PSM)

Treatment Group	Comparison Group	Counterfactual
$Y_{T,1}$	$Y_{0,1}$	$Y_{C,1}$
$Y_{T,2}$	$Y_{0,2}$	$Y_{C,2}$
$Y_{T,3}$	$Y_{0,3}$	$Y_{C,3}$
…	…	…
…	…	…
$Y_{T,N}$	$Y_{0,M}$	$Y_{C,N}$



ADB

PSM Computation

- The counterfactual of each individual *i* in the treatment group is the mean of matched comparisons:

$$Y_{C,i} = \frac{1}{n}\sum_{j=1}^{n} Y_{0,j}$$

- The program impact is estimated by the mean difference between the program outcome and the counterfactual:

$$\frac{1}{N}\sum_{i=1}^{N}[Y_{T,i} - Y_{C,i}]$$

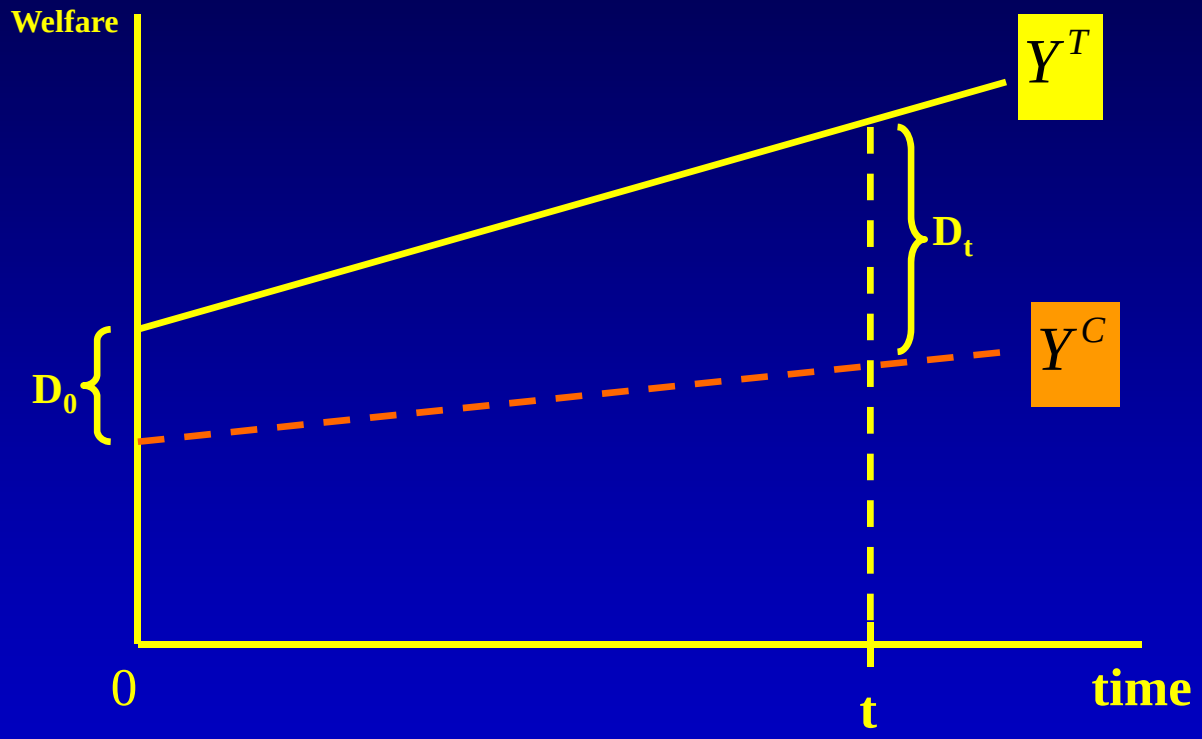
ADB

Limitations of PSM

- To get sufficient matches, large samples are required → expensive.
- Also, treatment and comparison groups must be quite similar.
- Hidden *bias may exist* because matching controls for observables only.
- If used without baseline data, *bias may occur* because matching is not controlling for pre-existing differences in characteristics.

ADB

Double Differences (DD)



Project Impact = $D_t - D_0$

ADB

Limitations of DD

- The underlying assumption is that the treatment and comparison groups would behave the same in the absence of the project. That is, the comparison is the control group. In reality, this is not true –> estimation bias.
- PSM takes care of this bias –> PSM and DD are often used together in practice.
- But then we again face the same critics as PSM above.

ADB

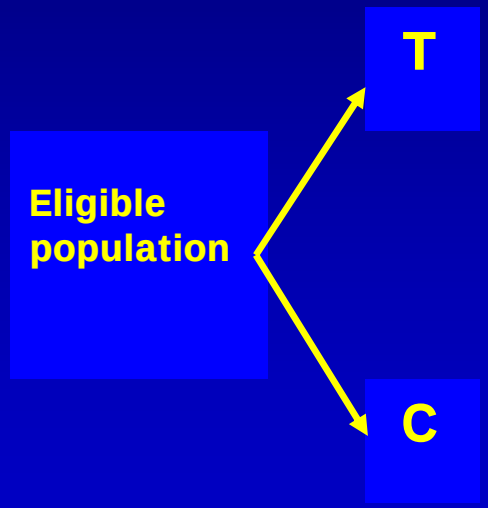
Advantages and Disadvantages of Non-Experimental Methods

- *Practical* -- applicable to almost all types of intervention and sometimes can be done retrospectively.
- Less subject to errors during project implementation.
- May yield *biased estimates* due to sample selection (not perfect control) and model specification.
- Could be data intensive (making the evaluation expensive) and computationally involved.

ADB

Randomized Evaluations

Randomized evaluation methods construct the counterfactual by randomly assigning a control group.



Project Effect:

$$\overline{Y}_T - \overline{Y}_C$$

difference between outcomes of the Treatment and Control groups

ADB

Randomized Evaluations

- Because of the random assignment of project participation, by construction, on average the Treatment group is identical to the Control group, except the participation in the project.
- This randomization effectively eliminates all pre-existing differences between the T and C groups → isolates the project effect.
- Randomized evaluations are therefore considered the "gold standard" among impact evaluation methods.

ADB

Methods of Randomization

Units of Randomization: Dependent on the nature of the intervention, can be individual (textbooks), household (micro credit), community/village (roads), school/institution (computers).

Lottery Design: Simply randomly choose participants from the target population.

- Usually used when there is no reason to discriminate among subsets of applicants and resources are limited.
- *Example*: Colombia school vouchers.

ADB

Methods of Randomization

Phase-In Design: Randomly choose some to begin first.

- Usually used when projects will be scaled up over time.
- *Example*: PROGRESA in Mexico.

Encouragement Design: Everybody is eligible to receive the project, but not all will do so.

- Can pick some people at random and encourage them to use the project, use non-encouraged as control.
- *Example*: job training programs.

ADB

Advantages

- *High internal validity* (considered to be "gold standard") because of the high quality of the counterfactual.
- Relatively easy to understand the method and to present results.
- Relatively less costly than non-experimental methods because of smaller sample sizes required.
- Integrates implementation with evaluation: focus on inputs and outcomes at the same time, allowing for possibility of improvements as program is being implemented.

ADB

Disadvantages

- Not applicable to all types of intervention: e.g., very difficult to do randomized evaluations of large infrastructure projects or projects designed to benefit a large part of or the entire population.
- Have problems with generalizability: e.g., a successful intervention in country A may not have the same impact in country B.
- *Internal validity is subject to appropriate design and implementation*: e.g., problems of attrition, spillover, contamination, randomization bias. Treatments of these may encounter the same statistical critics as non-experimental evaluations.

ADB

Operational Implications

- Financial requirement: Generally, could range from a few thousands to a few millions. But typically involve a few hundred thousands. Definitely less expensive than doing a wrong project.

- Data requirement: Baseline and follow-up surveys of with- and without-project households of, typically, a few thousands.

- Time requirement: Several years, depending on how long it takes for the project to show impacts.

ADB

Operational Concerns

- Face concerns: Expensive, technically difficult, unethical, Governments will not agree, cannot be used in many sectors, etc.

- Real problems: Lack of incentives to design and carry out rigorous impact evaluations because of the public good nature of evaluation. The result is there are fewer evaluations than needed and the quality is less than sufficient –> Evaluation Gap.

- Institutional problems: No clear mandate or institutional support.

ADB

International Experiences

- Fact: More than $55 billion (ADB, appr. $6 billion) spent on development every year, but scant hard evidence of whether they make a real difference.
- **Center for Global Development:** Reports an Evaluation Gap (in quantity and quality) and calls for international cooperation.
- **World Bank**: Conducts several evaluations. For example, Bangladesh food for education, Bolivia social investment fund, Czech active labor program, Kenya textbooks, Mexico retraining program for employment, Nicaragua school reforms, Vietnam rural roads, etc. See website: http://www1.worldbank.org/prem/poverty/ie/evaluationdb.htm
- **ADB**: Is getting onboard.

ADB

Thank you.

More Details on Non-Experimental Methods

The following slides are for additional information and will not be presented.

Propensity Score Matching (PSM)

- First proposed by Rosenbaum and Rubin (1983).
- The idea is to select from the comparison group a sample of individuals that are similar to the sample of projgram participants.
- To do this, PSM uses a predicted probability of participating in the program. This predicted probability is estimated using a logit or probit model based on observed characteristics.

Steps in PSM

- Conduct sample surveys of eligible non-participants and participants
- Pool the two samples and estimate the probability of participation using a logit / probit model with observable individual characteristics that are likely to determine participation (age, gender, income, education, etc.).
- Get predicted probability of participation (propensity score) for every sampled participant and non-participant.

Steps in PSM (cont.)

- For each individual in the sample of participants, find a small *n* (e.g., 5) observations in the non-participant sample with the closest propensity scores.
- Calculate the mean of the outcome for the chosen *n* observations. The difference between that mean and the actual outcome of the participant is an estimate of the program impact for that participant.
- Calculate the mean of these individual program impacts to obtain the average overall program impact.

ADB

PSM Computation

- The counterfactual of each individual *i* in the treatment group is the mean of matched comparisons:

$$Y_{C,i} = \frac{1}{n}\sum_{j=1}^{n} Y_{0,j}$$

- The program impact is estimated by the mean difference between the program outcome and the counterfactual:

$$\frac{1}{N}\sum_{i=1}^{N}[Y_{T,i} - Y_{C,i}]$$

ADB

Double Differences (DD)

- Developed by Heckman in the late 1970s.
- Compare outcome changes over time between the treatment group: $D^T = Y_t^T - Y_0^T$ and the comparison group: $D^C = Y_t^C - Y_0^C$.
- The underlying assumption is that the treatments would behave the same as the comparisons if the program had not happened. This makes the PSM a natural choice for determining the comparison group. And that is why PSM and DD are often used together in practice.

ADB

Steps in DD

- Conduct a pre-intervention baseline survey of both participants and non-participants.
- Conduct follow-up surveys, ideally of the same sampled observations as the baseline survey. If this is not possible, surveys should be in the same geographical areas.
- Construct the comparison group for participants in the baseline and follow-up surveys, using PSM.

ADB

Steps in DD (cont.)

- Calculate before-after difference for each participant.
- Calculate before-after differences for non-participants in the comparison group.
- Evaluate difference of those differences.

$$DD = D^T - D^C$$

ADB

$y = a + bT + cX + e$

Instrumental Variables (IV)

- Consider a simple model:

$$y = a + bT + cX + e$$

- If the program participation T is not exogenous, then the estimate of program effect b is biased. We then need to find a replacement for T that correlates with T but not directly with outcome y.

- *Example*: distribution of flyers in a micro credit program. Reading flyers may induce program participation but will not directly affect income. Whether flyers were distributed or not can be used as an *instrument* for participation.

- Problem: Easier said than done.

ADB

Regression Discontinuity Design (RDD)

- Applied when program participation is determined by an exogenous rule. For example, students receive free textbooks if income below a certain level.
- This method is based on the assumption that people around the cut-off point have similar characteristics. Then persons near the other side of the cut-off point can be used as the counterfactual.

ADB

RDD Illustration

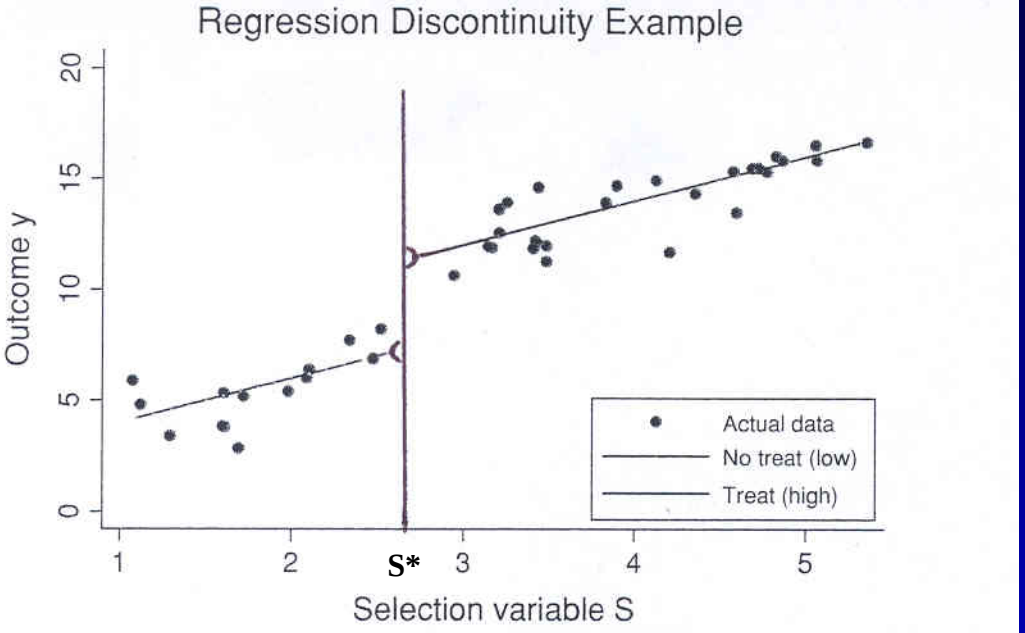


Figure 25.1: Regression-discontinuity design: example.

- The project impact is estimated by the mean difference in outcomes of persons just above and below the cut-off point S*.
- Problem: cannot say any thing about people far away from the cut-off point.

ADB

Recommended References

- **Judy Baker**. *Evaluating the impact of development projects on poverty – A handbook for practitioners.* World Bank, 2000. (Very practical introduction to non-experimental methods with examples)
- **Francois Bourguignon and Luiz A. Pereira da Silva**, editors. *The impact of economic policies on poverty and income distribution – Evaluation techniques and tools.* World Bank, 2003. (Excellent review of most robust micro and macro methods)
- **Peter H. Rossi, Mark W. Lipsey, and Howard E. Freeman**. *Evaluation: A systematic approach. Seventh Edition*. Sage Publications, 2004. (Most complete overview of methods)
- **Carol Weiss**. *Evaluation. Second Edition.* Simon and Schuster, 1998. (Basic introduction to evaluation)

ADB