Poverty reduction programs

Lecture 3-3

EC461 – 2/2021

Chayanee Chawanote

Slides adapted from Essential Development Economics (Ch.2)

Poverty reduction programs : Outline

- Poverty reduction programs
  - Impact evaluation: Motivation
  - “Causal effect” concept
  - Selection problem
  - Randomized controlled trials (RCTs)
  - Other method for Impact evaluations
    - Difference-in-differences
    - Example on school construction impacts in Indonesia

Impact Evaluation: Motivation

- Reducing poverty (and enhancing human flourishing) is hard!
- Many potential policies, programs and interventions exist
- Governments, donors, companies have limited budgets
- Impact evaluation typically aims to discover

  Q1: What works? [*Average Impact*]

  Q2: Why does it work? [*Impact Mechanisms or Channels*]

  Q3: Does it work better for some than for others?
  [*Impact Heterogeneity*]

- These are difficult questions…we'll focus on how to establish "What works?" (Q1)
- Very, very important! If you want to convince somebody that your idea/program/product works, how will you prove it?

“What Works?” == **Causal Impact**

- As economists, we mean: “What is the causal impact of the program on the outcome of interest?”
  - Example: “What was the average increase in maize yield (kg/acre) for farmers in Mozambique that is attributable to (caused by) gaining access to drought tolerant maize?”
- **Impact Evaluation** is a type of research that generates quantitative estimates of the causal impact of a program or intervention
  - The more cleanly a given *impact evaluation* can *identify* causality the more *rigorous* it is

"What Works?" == **Causal Impact**

- Establishing causality in economics is hard!
  - When human behavior is not "in play" (i.e., non-social sciences) it is much easier to establish causal relationships
- Main challenge: Correlation does not imply causality
- Teasing causality out of mere correlation is the major challenge to any "Impact Evaluation"

The Impact of Hospitalization on Health Status?

- The National Health Interview Survey collects data on a random sample of households in the US each year

- The survey includes the following 2 questions

  "During the past 12 months, were you hospitalized overnight?"

  "Would you say your health is excellent (5), very good (4), good (3), fair (2), or poor (1)?"

- Results

  Mean health status of hospitalized group = 3.2

  Mean health status of non-hospitalized group = 3.9

- So…hospitalization makes people sicker, right?

The Selection Problem

- Correlation does not imply causality!
  - There is a negative correlation between hospitalization and health
  - That does not imply that hospitalization ***causes*** health to get worse!

- The root of this problem is a mismatch between (i) the question we want to answer and (ii) the data we have to answer it
  - We'd like to use the health status of the un-hospitalized people to answer the question: "How healthy would the hospitalized group be *if they had not gone to the hospital*?"
  - But the data don't come in this form!

- There are systematic differences between the hospitalized and un-hospitalized group
  - The un-hospitalized group is healthier (that's why they didn't go to the hospital!)
  - So even without hospitalization the hospitalized group would have been sicker than the un-hospitalized group
  - This type of systematic difference between "Treatment" and "Control" groups underlies the "Selection Problem" and "Selection Bias"

The Selection Problem in Development: The Impact of Credit on Farm Income?

- Each year the Colombian government conducts the National Rural Survey with a random sample of 10,000 farm households
- The survey includes the following 2 questions:

  "During the past 12 months, did you receive a loan from a bank?"

  "What was your income from farming in the last 12 months?"

The Selection Problem in Development: The Impact of Credit on Farm Income?

- And the survey says....

Group	Sample Size	Mean Farm Income	Standard Error
Borrowers	1,550	$4,200	48.5
Non-borrowers	8,450	$2,500	28.9

- Our research question is: What is the ***causal*** impact of credit on farm income?
- Consider the following ***naïve*** impact estimate:

  Impact = (Mean income of Borrowers) – (Mean income of Non-borrowers)

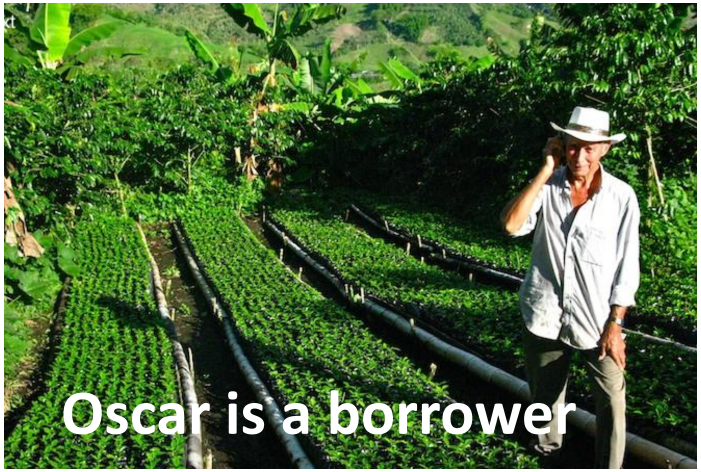
  Impact = 4,200 – 2,500 = 1,700

The Selection Problem in Development: The Impact of Credit on Farm Income?

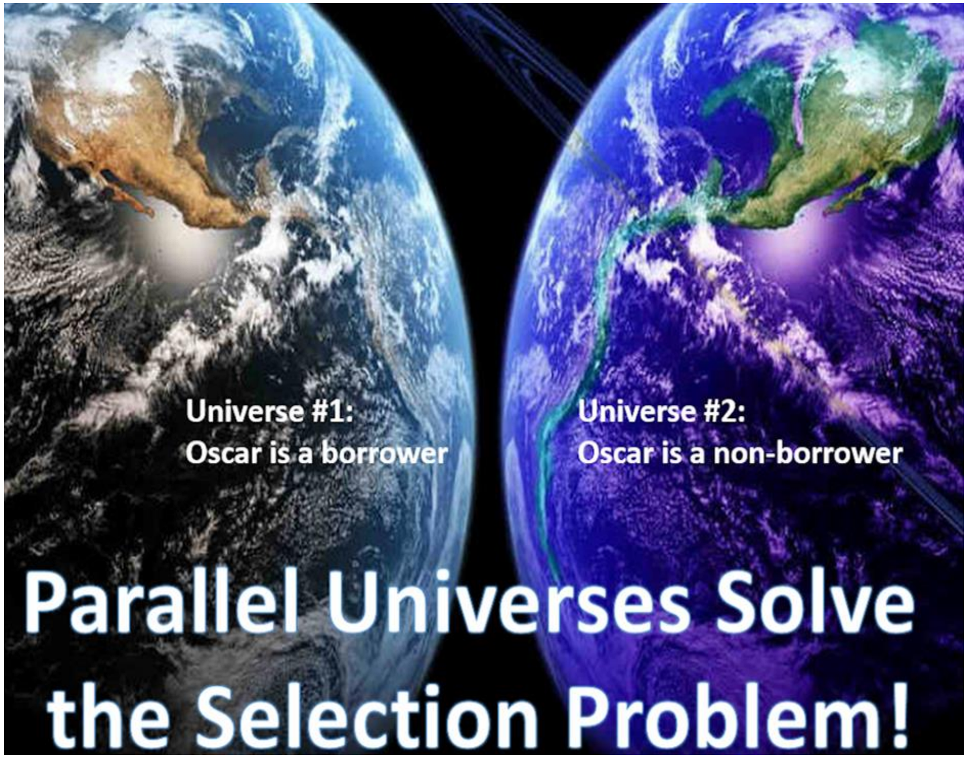
- Why is the naïve impact estimate not convincing as a *causal* impact estimate?
- In order to evaluate the *causal* impact of credit, we need to answer the following counterfactual question:

  "How much would borrowers have earned *if they had not borrowed*?"

- But we cannot directly observe the *counterfactual*
- If there is any selection process at work that makes borrowers different than non-borrowers, then we have a Selection Problem…



The Selection Problem in Development: The Impact of Credit on Farm Income?



- Until we can interview both Oscars and compare their farm income to get a clean/rigorous ***causal*** impact estimate, we must seek an alternative
- E.g. Find a "control group" to provide an estimate of the counterfactual
- The "naïve" impact estimate above used non-borrowers as the control group
- Why do non-borrowers make a bad control group? Let's look at where the selection problem comes from

The Selection Problem

- Borrowers are probably richer farmers with better land, better access to input and output markets, and better farming skills than non-borrowers
- Even if they had not borrowed, the borrowing group probably would have earned more money than the group that did not borrow.
- So the average income of the non-borrowers is not a good *counterfactual* for borrowers

True Impact =

(Avg Y of borrowers) – (Avg Y of borrowers *if they had not borrowed*)

Naïve Impact Estimate =

(Avg Y of borrowers) – (Avg Y of non-borrowers)

The Selection Problem

- This is an example of the "Selection" problem, which leads to "Selection" bias.
- The "Selection" problem occurs when individual choices lead to
  1. Systematic differences between the treatment and control groups and
  2. Differences in the outcome variable of interest (income) driven by these systematic differences between the treatment and control groups
- If we don't account (control) for these differences, we get biased estimates

  (And why again do development economists worry about biased estimates?)
- Since it presents <u>the</u> major impediment in our pursuit of truth in applied economics (indeed, in applied social science more generally!), let's explore selection bias more carefully

Potential Outcome Framework

- A conceptual framework to think about impact evaluation (and econometrics in general) that…
- Will provide us notation to carefully define

1. Average Treatment Effect (ATE) on the Treated (the *Causal* Impact)
2. Selection Bias

- Will suggest ways (empirical methodologies) to avoid/eliminate selection bias

Potential Outcome Framework

- Let Y_i be the income we observe for person i:

$$Y_i = \begin{cases} Y_{1i} \;\; if\, D_i = 1 \\ Y_{0i} \;\; if\, D_i = 0 \end{cases}$$

- For borrowers:
  $D_i = 1$
  So we observe income with credit, Y_{1i}
  Y_{0i} is perfectly well defined, but not observable (in our universe)

- For non-borrowers:
  $D_i = 0$
  So we observe income without credit, Y_{0i}
  Y_{1i} is perfectly well defined, but not observable (in our universe)

Potential Outcome Framework

- We want to know the Average Treatment Effect (ATE) on the treated
- In our example, ATE on treated = Avg impact of credit on those with credit
- $ATE = E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]$ …
  $E[Y_{1i}|D_i = 1]$: Average *income with credit* for borrowers
  $E[Y_{0i}|D_i = 1]$: Average *income without credit* for borrowers
- The problem is that we don't observe that ***counterfactual*** second term
  We don't observe Oscar's doppelgänger in Universe #2
- So what can we do? Let's try our naïve estimate:
  $ATE^{Naive} = E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0]$
- Will this be a problem?

Selection Bias in Naïve Estimate

	Di = 0	Di = 1
Y0i	2,500	3,500
Y1i	3,800	4,200

ATE^{Naive} = $E[Y_{1i}|D_i = 1]$ - $E[Y_{0i}|D_i = 0]$

ATE^{Naive} =*(Avg. income with credit of borrowers) – (Avg income without credit of non-borrowers)*

Now subtract and add $E[Y_{0i}|D_i = 1]$ to RHS:

ATE^{Naive} = $\{E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]\}$ **+** $\{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]\}$

ATE^{Naive} = *{Average Treatment Effect}* + *{Selection Bias}*

ATE^{Naive} = *{4,200 – 3,500}* + *{3,500 – 2,500}*

ATE^{Naive} = *700* + *1,000*

The true causal impact of credit is $700.

The naïve approach over-estimates the impact by $1,000.

Selection Bias in Naïve Estimate

- Selection Bias = $\{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]\}$
- In our example:
  $E[Y_{0i}|D_i = 1]$ = Average of ***income without credit*** for ***borrowers = $3,500.***
  $E[Y_{0i}|D_i = 0]$= Average of ***income without credit*** for ***non-borrowers = $2,500.***

- Why is there selection bias in this example?
  $E[Y_{0i}|D_i = 1] > E[Y_{0i}|D_i = 0]$
  If they had not received credit, those who borrowed ***would have*** earned more income than those who did receive credit.
  Why? Because, in the Colombian government survey, borrowers are systematically different from non-borrowers (better farmers, better land, wealthier, etc.).
  Therefore, the average income of non-borrowers does not provide a good answer to our key counterfactual question.

$$ATE^{Naive} = \{E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]\} + \{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]\}$$

$$ATE^{Naive} = ATE + Selection\ Bias$$

Empirical strategies to estimate ***causal*** impacts must eliminate Selection Bias.

Randomization to the Rescue:

Randomized Controlled Trial (RCT)

"Creating a culture in which rigorous randomized evaluations are promoted, encouraged, and financed has the potential to revolutionize social policy during the 21st century, just as randomized trials revolutionized medicine during the 20th."

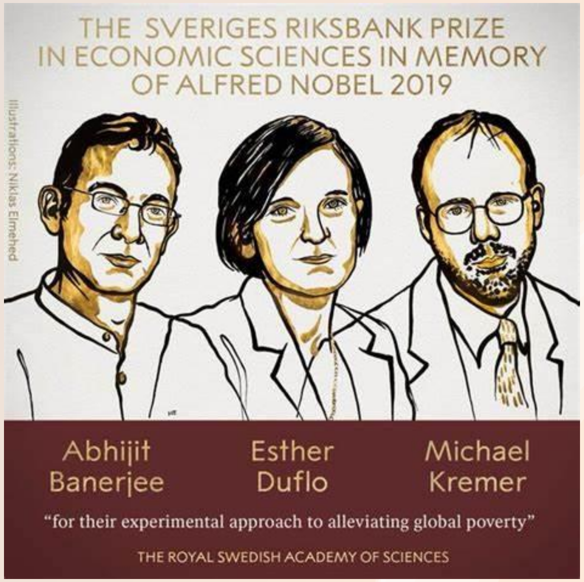
-Esther Duflo

Esther Duflo: Social experiments to fight poverty

https://www.ted.com/talks/esther_duflo_social_experiments_to_fight_poverty

https://news.mit.edu/2019/esther-duflo-abhijit-banerjee-win-2019-nobel-prize-economics-1014

https://www.nobelprize.org/prizes/economic-sciences/2019/prize-announcement/



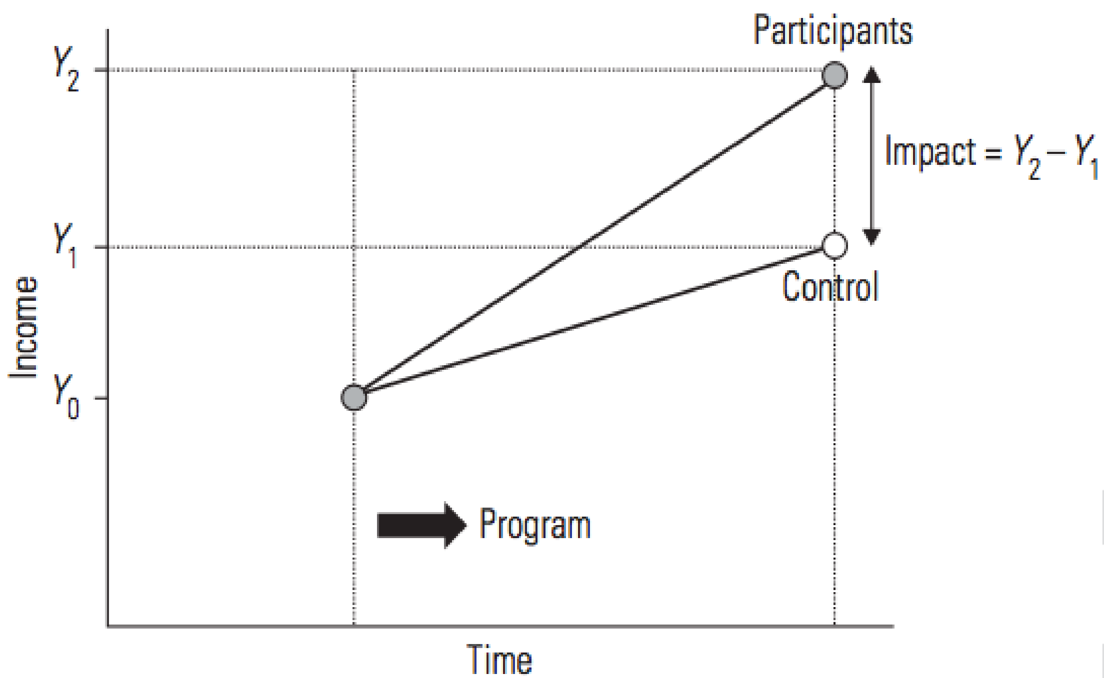
Randomized Controlled Trial (RCT)

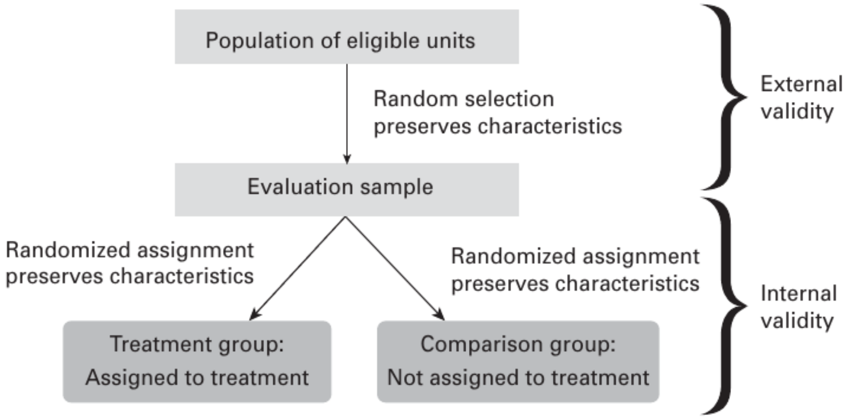
- Selection Bias again: $E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]$
- If we randomly put people into treatment and comparison (control) groups:
  - Should get no systematic differences across groups.
  - If $E[Y_{0i}|D_i = 1] = E[Y_{0i}|D_i = 0]$, then selection bias is zero
  - The observed difference in means of treatment and control group is the causal impact of the program on the treated group.
  - Randomization → the mean of our control group gives us a very good answer to our counterfactual question (what would borrowers have earned if they had not borrowed).

- This is what we do in a ***Randomized Controlled Trial*** (RCT)
  - We randomly assign individuals into treatment and control groups
  - Treatment group gets the treatment; Control does not get the treatment
  - Compare average income (or other outcome variable of interest) of treatment group to control group
  - If randomization worked, we have no selection bias
  - Thus this simple difference in means gives us the causal impact of the program (treatment)

Randomized Controlled Trial (RCT)

The Ideal Experiment with an Equivalent Control Group





Mexico’s PROGRESA Program (OPORTUNIDADES)

- In 1997, poor countries didn’t have major protection programs for the rural poor
- Most formal-sector workers in primarily urban areas had social security and medical insurance
- What about the rural poor?
- Radical new idea: Give money to the poor
  - Give mostly to women
  - Make the ‘cash transfer’ conditional on getting kids to school and clinics
- Santiago Levy (PhD Econ, Boston University) designed the rollout of PROGRESA to facilitate impact evaluation

- Impossible to launch full program at once due to capacity constraints
- Instead, rolled out to randomly chosen villages with new randomly chosen villages added each subsequent year
- Baseline surveys of everyone determined household-level eligibility
- After a year, conduct follow-up surveys to estimate ATE
- This created a near-perfect randomized experiment!
- Looked a bit like a drug trial – but without a placebo

Mexico's PROGRESA Program (OPORTUNIDADES)

Enrollment in school of kid i at time t (S_{it}=1 if enrolled, 0 otherwise; P_i=1 if a PROGRESA village; E_i=1 if household is eligible for PROGRESA):

$$S_{it} = \sum_{t=1}^{5} \alpha_{0t} + \alpha_1 P_i + \alpha_2 E_i + \alpha_3 P_i E_i ...$$

Really Shouldn't Matter

Estimated derivatives at sample means	Panel matched sample			
	Primary		Secondary	
	Female	Male	Female	Male
Progresa locality a*$_1$ (t ratio)	0.0005	−0.0088	−0.0232	0.0048
	(0.12)	(1.62)	(0.81)	(0.18)
Poor eligible household in Progresa locality a*$_3$ (t ratio)	0.0087	0.0168	0.1155	0.0572
	(2.17)	(3.97)	(4.32)	(2.51)
Net Progresa impact [significance non-zero based on joint χ^2 test]	0.0092	0.0080	0.0923	0.0620
	[0.0030]	[0.0015]	[0.0003]	[0.0027]
Sample size	33,795	36,390	13,872	14,523
Pseudo R^2	0.3728	0.3712	0.3116	0.2979

P_i: Kid i in randomly assigned PROGRESA village; E_i: eligibility dummy

T. Paul Schultz. 2004. School Subsidies for the Poor: Evaluating the Mexican PROGRESA Poverty Program. *Journal of Development Economics* 74(199-250).

RCT: Summary

- The Selection Problem >> Selection Bias
- Parallel Universes can't provide counterfactuals in practice
- Instead, we introduce or exploit RANDOMNESS in treatment status
    - Introduce RANDOMNESS: RCT's can provide a compelling counterfactual
    - Exploit RANDOMNESS: Natural experiments or other 'exogenous' factors can be used to "identify" causal impacts
    - E.g., How does income affect *Outcome X*? Rainfall → Exogenous income changes

- As we dial-down self-selection into treatment and dial-up random assignment into treatment, the selection problem fades and we become more confident about ***causal*** impact evaluation
    - In pure RCT, the simple difference in means of treated group versus control group is an unbiased estimate of the mean causal impact of the program/intervention on the population of interest (*ATE*)

- Note: ATE typically focus on evaluating the benefits of a program
    - While benefits are usually harder to measure, costs are also critical for impact evaluation
    - Cost-effectiveness combines ATE with costs of program to determine efficiency

- So…RCTs sound great, but they also raise ethical issues

RCT: Arguments

- **Ethical issues**: who are in the treatment group vs. who are in the control group
  - Interventions in human subjects
  - E.g., if they are not ready to take randomized loan, future debt could create harm.
- **External validity**: a pilot project may not be an accurate guide of the project's impact on a national scale
  - No one size fits all
  - Findings obtained from one population may not be generalizable to others
- **Compliance**: a fraction of individuals who are offered the treatment do not take it >> selection bias
- **Selective attrition**: people drop out of a program
- **Spillovers**: treatment helps the control group, thereby confounding the estimates of program impact

If not RCT, what can we do?

Table 11.2 Comparing Impact Evaluation Methods

Methodology	Description	Who is in the comparison group?	Key assumption	Required data
Randomized assignment	Eligible units are randomly assigned to a treatment or comparison group. Each eligible unit has an equal chance of being selected. Tends to generate internally valid impact estimates under the weakest assumptions.	Eligible units that are randomly assigned to the comparison group.	Randomization effectively produces two groups that are statistically identical with respect to observed and unobserved characteristics (at baseline and through endline).	Follow-up outcome data for treatment and comparison groups; baseline outcomes and other characteristics for treatment and comparison groups to check balance.
Instrumental variable (particularly randomized promotion)	A randomized instrument (such as a promotion campaign) induces changes in participation in the program being evaluated. The method uses the change in outcomes induced by the change in participation rates to estimate program impacts.	"Complier" units whose participation in the program is affected by the instrument (they would participate if exposed to the instrument, but would not participate if not exposed to the instrument).	The instrument affects participation in the program but does not directly affect outcomes (that is, the instrument affects outcomes only by changing the probability of participating in the program).	Follow-up outcome data for all units; data on effective participation in the program; data on baseline outcomes and other characteristics.
Regression discontinuity design	Units are ranked based on specific quantitative and continuous criteria, such as a poverty index. There is a cutoff that determines whether or not a unit is eligible to participate in a program. Outcomes for participants on one side of the cutoff are compared with outcomes for nonparticipants on the other side of the cutoff.	Units that are close to the cutoff but are ineligible to receive the program.	To identify unbiased program impacts for the population close to the cutoff, units that are immediately below and immediately above the cutoff are statistically identical. To identify unbiased program impacts for the whole population, the population close to the cutoff needs to be representative of the whole population.	Follow-up outcome data; ranking index and eligibility cutoff; data on baseline outcomes and other characteristics.

If not RCT, what can we do?

Table 11.2 *(continued)*

Methodology	Description	Who is in the comparison group?	Key assumption	Required data
Difference-in-differences	The change in outcome over time in a group of nonparticipants is used to estimate what would have been the change of outcomes for a group of participants in the absence of a program.	Units that did not participate in the program (for any reason), and for which data were collected before and after the program.	If the program did not exist, outcomes for the groups of participants and nonparticipants would have grown in parallel over time.	Baseline and follow-up data on outcomes and other characteristics for both participants and nonparticipants.
Matching (particularly propensity score matching)	For each program participant, the method looks for the "most similar" unit in the group of nonparticipant (the closest match based on observed characteristics).	For each participant, the nonparticipant unit that is predicted to have the same likelihood to have participated in the program based on observed characteristics.	There is no characteristic that affects program participation beyond the observed characteristics used for matching.	Follow-up outcome data for participants and nonparticipants; data on effective participation in the program; baseline characteristics to perform matching.

Source: Adapted from the Abdul Latif Jameel Poverty Action Lab (J-PAL) website.

School construction in Indonesia: Duflo (2001)

- The paper presents both a program evaluation of school construction and a clever way to identify returns to education
- Between 1973 and 1979, Indonesia undertook a massive primary school construction program - more than 1 school was built for every 500 children across the country
- Question : What was the effect of the program on
  - Educational attainment?
  - Wages?
- Regress outcomes (Y) on number of schools built per 1,000 students (P)?
  $Y_{ij} = \alpha_0 + \alpha_1 P_j + \epsilon_{ij}$, for individual i and region j

School construction in Indonesia
(Duflo, 2001)

Context: regions would have had differences in enrollment (and wages) in the absence of the program

- Control for 'pre-existing' differences in enrollment across regions using older individuals.
- Children who were 2-6 years old in 1974 were likely exposed to the program for a number of years.
- Children who were 12-17 years old in 1974 were unlikely to have been in primary school when the new schools were constructed.
- Thus, individuals aged 12-17 in 1974 can serve as a proxy for baseline enrollment differences across regions.

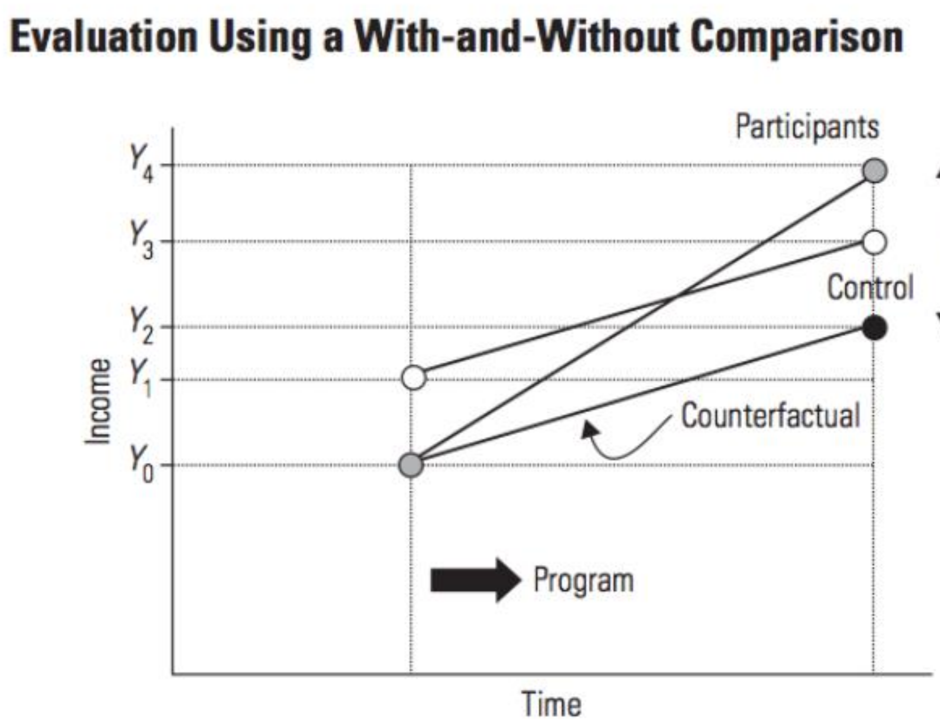
School construction in Indonesia
(Duflo, 2001)

- We want to know whether building schools affect children outcomes.
- Let there are 2 treatments categories we care about: low intensity school construction ares ($P_j = 0$), and high-intensity school construction areas ($P_j = 1$).
- We are interested in estimating the effect of high-intensity construction, relative to low-intensity construction.
- Y_{0i} = the outcome for individual i in the state where she receives low-intensity construction,
- Y_{1i} = the outcome in the state where she receives high-intensity construction
    - m = individual in the younger group (exposed to the program)
    - n = inidividual in the older group (not exposed to the program)

School construction in Indonesia: **Causal effect**

- Define the causal effect of high intensity construction on those who actually received high-intensity construction as $E(Y_{1i}|P=1) - E(Y_{0i}|P=1)$ where the conditioning in the expectation indicates the category that the individuals actually falls into.
  - The latter term is not observed, called the counterfactual
- What we observe is $E(Y_{1i}|P=1) - E(Y_{0i}|P=0)$
- Therefore, the assumption that you have to make for this estimator to be valid is $E(Y_{0i}|P=0) = E(Y_{0i}|P=1)$
- Think about this as the young-old diffrence in the high-intensive areas relative to the young-old difference in the low-intensive program areas: $DD = [E(Y_{m1}|P=1) - E(Y_{n1}|P=1)] - [E(Y_{m0}|P=0)$ $-E(Y_{n0}|P=0)]$

Causal effect: Difference-in-differences method



	Before	After	Difference
Treatment	Y_0	Y_4	$Y_4 - Y_0$
Control	Y_1	Y_3	$Y_3 - Y_1$
Difference	$Y_0 - Y_1$	$Y_4 - Y_3$	DiD = $Y_4 - Y_2$

- We need the assumption that the young-old difference in the low-intensity areas is equivalent to the young-old difference in the high-intensity areas, had they received the low-intensity intervention.

Causal effect: Difference-in-differences method

- See Duflo (2001) Table 3

TABLE 3—MEANS OF EDUCATION AND LOG(WAGE) BY COHORT AND LEVEL OF PROGRAM CELLS

	Years of education			Log(wages)		
	Level of program in region of birth			Level of program in region of birth		
	High (1)	Low (2)	Difference (3)	High (4)	Low (5)	Difference (6)
Panel A: Experiment of Interest						
Aged 2 to 6 in 1974	8.49	9.76	−1.27	6.61	6.73	−0.12
	(0.043)	(0.037)	(0.057)	(0.0078)	(0.0064)	(0.010)
Aged 12 to 17 in 1974	8.02	9.40	−1.39	6.87	7.02	−0.15
	(0.053)	(0.042)	(0.067)	(0.0085)	(0.0069)	(0.011)
Difference	0.47	0.36	0.12	−0.26	−0.29	0.026
	(0.070)	(0.038)	(0.089)	(0.011)	(0.0096)	(0.015)
Panel B: Control Experiment						
Aged 12 to 17 in 1974	8.02	9.40	−1.39	6.87	7.02	−0.15
	(0.053)	(0.042)	(0.067)	(0.0085)	(0.0069)	(0.011)
Aged 18 to 24 in 1974	7.70	9.12	−1.42	6.92	7.08	−0.16
	(0.059)	(0.044)	(0.072)	(0.0097)	(0.0076)	(0.012)
Difference	0.32	0.28	0.034	0.056	0.063	0.0070
	(0.080)	(0.061)	(0.098)	(0.013)	(0.010)	(0.016)

Notes: The sample is made of the individuals who earn a wage. Standard errors are in parentheses.

Causal effect: Difference-in-differences method

	Before	After	Difference
Treatment			
Control			
Difference			

DD in the regression model:

$$Y_{ijk} = \alpha_0 + \alpha_1 P_j + \alpha_2 T_k + \alpha_3 (P_j T_k) + \epsilon_{ijk}$$

- Y_{ijk} = the outcome for student in i in region j in birth cohort k
- P_j = a dummy variable for high-intensity areas
- T_k = a dummy variable representing the younger cohort
- α_3 is the DD estimate for the effect of the program

Impacts of school construction in Indonesia (Duflo, 2001)

- In both cohorts, the average educational attainment and wages in regions that received fewer schools are higher than in regions that received more school. More schools were to be built in regions where enrollment rates were low.
- In both types of regions, average educational attainment increased more in regions that received more schools.
- Schooling program raised educational attainment by 0.15 years on average for each school built per 1000 students in the region, and raised wages by about 2 percent.

- The simplified of the regression in the paper: $Y_{ir} = \gamma_0 + \gamma_1 P_{ir} + \gamma_2 C_{ir} + \epsilon_{ir}$
  - Y_{ir} = outcome of interest for individual i in region r
  - P_{ir} = the number of schools constructed per 1000 students in region r for the younger cohort
  - C_{ir}= controls for differences across regions among the older cohort
- See Duflo (2001) Table 4