Randomized Control Trials and Policy Evaluation

Matteo Bobba
matteo.bobba@tse-fr.eu
Office: T.353

Toulouse School of Economics (TSE)

M2 PPD/ERNA/EEE, Winter 2025

What is this Course About?

- Randomization is a prominent tool for economists and other social scientists
  - Explosion of randomized control trials (RCTs) in development economics
  - Increasingly used in other applied-micro fields (labor, health, environ, ...)
  - Very popular among donors, international institutions, policy makers, etc.
- This course presents a broad overview of RCTs methods and applications
  - Research-oriented approach
  - Practical and hands-on approach

Course Material

1. Slides – updated frequently on the course's moodle page
2. Reading material (selected chapters/sections):
   - Imbens and Rubin (IR), "Causal Inference for Statistics, Social, and Biomedical Sciences"
   - Athey and Imbens (AI), "The econometrics of Randomized Experiments"
   - Duflo, Glennerster, Kremer (DGK), "Using Randomization in Development Economics Research: A Toolkit"
   - Other academic articles discussed in class (e.g. today's class)
3. Weekly TDs with software+data applications of topics covered in class

Course requirements [relative weight]

1. Final written exam [50%]
   - March 31, 2025 at 2pm (90 minutes)
   - See previous years' exams in moodle
2. Take-home exercises [25%]
   - Check with Kevin Frick (TA)
3. Paper presentation with slides [25%]
   - Presentations during the last two/three weeks of class
   - See previous years' presentation slides in moodle

Outline and Timeline of the Course

1. Intro and overview (Week 1)
2. Econometrics of RCTs (Weeks 2 to 5)
   - The basic framework
   - Statistical analysis of experimental data
3. Design and implementation issues (Weeks 6 to 8)
   - Sample size considerations
   - Non-compliance, spillovers, attrition, and multiple outcomes
4. RCTs applications (Weeks 8 to 10)
   - Students' presentations

Part 1: Intro and Overview

1. Endogeneity and causality in economics
   - Correlation is not causality
   - Exemplary cases of endogeneity
2. The case for and against RCTs
   - A brief history of RCTs in economics
   - The triumph of the experimental approach in development economics
   - Randomization and its discontents

Endogeneity and causality in economics

Descriptive models

- Consider an iid sample of N observations of a couple of random variables $S = \{(y_i; x_i); i = 1, ..., N\}$
- A descriptive model is a statistical restriction (a list of assumptions) on the distribution of S
- Example. The linear model: $y_i = \alpha + \beta x_i + \epsilon_i$

$$\hat{\beta} = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N} (x_i - \bar{x})^2}; \quad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}.$$

Econometric models

- Econometric models aim at isolating the direction of causality among variables of interest
  - E.g. the effect of a change in price of a good on its quantity demanded
- Causality is based on the notion of **controlled variation**
  - The *ceteris paribus* condition: "other (relevant) factors being equal"
  - This implies positing more assumptions on your model
- Example (cont'd): if $\mathbb{E}(\epsilon_i|x_i) = 0$ then β can be interpreted as the causal effect of x_i on y_i
  - The pool game analogy Pictures

Characterization of Endogeneity Biases

- Recall our sample $S = \{(y_i; x_i); i = 1, ..., N\}$
  - What is the effect of x on y?
- There are two main challenges in uncovering causal relationships from observational data
  1. x may be endogenous, either because of **simultaneity** or because of **unobserved heterogeneity**
  2. S may be endogenous, because of **selectivity** of the units that are observed

Example of Simultaneity Bias

- Demand and supply model

$$D_i = \alpha_0 - \alpha_1 p_i + u_i$$
$$S_i = \beta_0 + \beta_1 p_i + v_i$$

- Data on exchanged quantities y_i ($D_i = S_i$)

$$y_i = \frac{\beta_0 \alpha_1 + \beta_1 \alpha_0 + \beta_1 u_i + \alpha_1 v_i}{\beta_1 + \alpha_1}$$
$$p_i = \frac{\alpha_0 - \beta_0 + u_i - v_i}{\beta_1 + \alpha_1}$$

- Regressing y_i on p_i and assuming u_i and v_i are uncorrelated yields

$$\gamma = \frac{Cov(y_i, p_i)}{Var(p_i)} = \frac{\beta_1 \sigma_v^2 - \alpha_1 \sigma_u^2}{\sigma_u^2 + \sigma_v^2}$$

Example of Unobserved Heterogeneity Bias

- Individuals optimally decide schooling based on the following:

$$\max_{S} \log(y) - \phi(S) \text{ s.t. } y = g(S)$$

- First-order condition for optimal schooling is:

$$\frac{g'(S)}{g(S)} = \phi'(S)$$

- Optimal schooling equates the marginal rate of return to schooling with the marginal cost

Example of Unobserved Heterogeneity Bias (Cont'd)

- Assume both MR and MC of schooling are linear

$$\frac{g'(S)}{g(S)} = b_i - k_1 S$$
$$\phi'(S) = r_i + k_2 S$$

- Hence, optimal schooling is:

$$S_i^\star = \frac{b_i - r_i}{k_1 + k_2}$$

  - Variation in b_i corresponds to variation in "ability"
  - Variation in r_i corresponds to variation in "access to funds" (family wealth) or in "tastes for schooling"

Example of Unobserved Heterogeneity Bias (Cont'd)

- The OLS estimate of schooling on earnings is a weighted average of $\bar{b}$ and $\bar{r}$
  - It is larger than the average MR to schooling because people with higher marginal returns to education choose higher levels of schooling
- An IV estimate can recover the average MR if $b_i = b \, \forall i$
  - $IV > OLS$ If the instrument affects a sub-population with a sufficiently high MR

Example of Sample Selection Bias

- Individuals determine labor supply by trading off consumption for leisure

$$\max_{c,h} c - v(h) \text{ s.t. } c \leq wh + V$$

- Interior solution

$$v'(h^\star) = w$$

- Corner solution

$$v'(0) > w$$

- Reservation wage is $w^\star = v'(0)$. Work if $w \geq w^\star$

Example of Sample Selection Bias (cont'd)

- Empirical specification for reservation wage:

$$w_i^\star = X_i'\theta + \eta_i$$

- Offered wages are

$$w_i = X_i'\beta + \epsilon_i$$

- Assume that $\mathbb{E}(\epsilon_i | X_i) = 0$, so no endogeneity in the wage equation
- Individual i works $(D_i = 1)$ when

$$\begin{aligned} X_i'\beta + \epsilon_i &\geq X_i'\theta + \eta_i \\ \Rightarrow X_i'(\beta - \theta) + (\epsilon_i - \eta_i) &\geq 0 \\ \Rightarrow X_i'\psi &\geq \nu_i \end{aligned}$$

Example of Sample Selection Bias (cont'd)

- We observe wages for individuals who work ($D_i = 1$). Hence:

$$\mathbb{E}(w_i|X_i, D_i = 1) = X_i'\beta + \mathbb{E}(\epsilon_i|X_i, \nu_i < X_i'\psi)$$

- If ϵ_i and ν_i are independent, OLS of wages on X_i recovers β
  - This is equivalent of saying we have a randomly truncated sample (Tobit)
- If ϵ_i and ν_i are not independent, then $\mathbb{E}(\epsilon_i|X_i, D_i = 1) \neq 0$ and OLS on truncated sample is inconsistent (Heckit)

The impact of microcredit

- Group-based lending for the poor was the basis for the 2006 Nobel Peace Prize and embraced by policymakers, donors, and funders worldwide as an effective development tool
  - Scarce empirical evidence to support this claim
- Sources of endogeneity bias that plague ex-post evaluation studies
  - People who choose to borrow may be different than those who choose not to
  - Lenders choose which neighborhoods/markets to enter
  - Biases can go in either directions

The impact of microcredit (cont'd)

- Randomization of program placement between locations (six studies in different countries)
  - Modest take-up
  - Some substitution effects with respect to other credit sources
  - Increase in business size and profits
  - No effects on HHs income and consumption
  - No effects on child schooling and female decision power
- Microcredit does not seem to lead to any transformative effect in the lives of the poor !

The notion of counterfactual

- How would individuals who (did not) participated in a program have fared in the absence (presence) of the program?
  - At any given point in time, an individual is either exposed to the program or not
- We need a comparison group
  - Compare the same individual over time (pre-post)?
  - Compare the average impact of the program between those who participate and who do not (with-without)?

Potential outcome framework

- Think about participation in microcredit as described by a binary random variable $T_i = \{0, 1\}$
- Let Y_i^1 be the potential outcome for individual i if she receives microcredit loans and Y_i^0 be the potential outcome for the same individual if she does not receive loans
- The causal effect of microcredit for individual i is

$$\Delta_i = Y_i^1 - Y_i^0$$

Selection bias

- The average causal effect of microcredit is

$$\begin{aligned} ATE &= \mathbb{E}(Y_i^1 - Y_i^0) \\ &= \mathbb{E}(Y_i^1 | T_i = 1) - \mathbb{E}(Y_i^0 | T_i = 0) \\ &= \underbrace{\mathbb{E}(Y_i^1 - Y_i^0 | T_i = 1)}_{\text{ATT}} + \underbrace{\mathbb{E}(Y_i^0 | T_i = 1) - \mathbb{E}(Y_i^0 | T_i = 0)}_{\text{Selection bias}} \end{aligned}$$

- Those who borrow from microcredit may have had different outcomes on average even if they had not borrowed
    - $\mathbb{E}(Y_i^0 | T_i = 1) > \mathbb{E}(Y_i^0 | T_i = 0)$
    - $\mathbb{E}(Y_i^0 | T_i = 1) < \mathbb{E}(Y_i^0 | T_i = 0)$

The Promise of RCTs

- Experiments induce **controlled** variations in a policy variable
  - Generate sources of exogenous variation in real world economic environments
  - Transparent and easy-to-replicate across different contexts
  - Facilitate a process of dynamic learning among researchers, implementing partners and policy makers
- RCTs provide causal estimates with minimal statistical assumptions
  - Inference is valid **in the sample under study** and **when they are correctly designed and implemented**

The case for and against RCTs

Part 1: Reading material

- Banerjee and Duflo: *The Experimental Approach to Development Economics*, Annual Reviews of Economics. 2009
- Deaton: *Randomization in the Tropics Revisited: a Theme and Eleven Variations*, NBER Working Paper 2020.
- Heckman: *Randomization and Social Policy Evaluation Revisited*, NBER Working Paper 2020.
- List: *The Voltage Effect: How to Make Good Ideas Great and Great Ideas Scale*, Penguin Books, 2022.

Historical Background

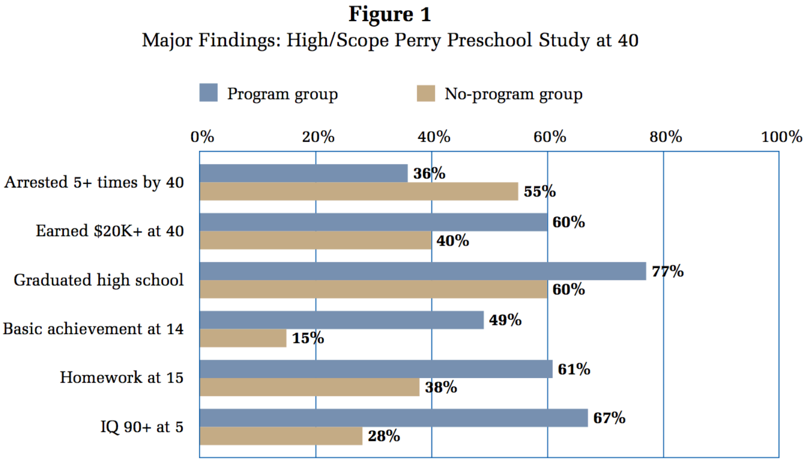
"Experimentation for policy purposes is needed to attack questions of interest to policy makers" [Orcutt and Orcutt, AER 1968]

- By the early 80s, there were more than 70 social experiments in the US
  - Education and training
  - Employment programs and income transfers
- Small-scale/pilot programs, government run and individual-level randomization

An Example: The Perry pre-school project

- Five entry cohorts drawn from the population surrounding the Perry elementary school
  - Targeting: kids aged 3 in disadvantaged african-american families
  - Treatment: 2.5 hours educational preschool plus weekly home visits by teachers for two years
- Each cohort is followed over entire life and measured a wide array of cognitive, non-cognitive and socio-economic outcomes
  - Sample size: 123 children allocated over five entry cohorts
  - Measurement: annual surveys till age 15, plus follow-ups at ages 19, 27 and 40 (over 91% of original subjects interviewed)

The Perry pre-school project



The Rise of RCTs in Economics: Three Phases

1. Proof-of-concept using RCTs
   - Test bundled interventions (e.g. CCTs)
2. From proof-of-concepts to field experiments
   1. Delve into individual program components (e.g. cross cutting designs)
3. From field experiments to scalable policies
   - At-scale randomization designs

The Impact of RCTs on Economics Research

1. Shift focus from evaluating to identifying workable policies, for which one could make causal claims of impact
   - We now have a large number of concrete results on specific mechanisms behind poverty and specific interventions to alleviate it
2. Microeconomic approach
   - Breaking down a research question into smaller, more manageable topics, each of which could be rigorously studied via specifically designed randomized controlled trials
3. A greater focus on identification across the board
   - Large impact on observational methods, and model-based approaches

The Impact of RCTs on Economics Research

4. From proof-of-concept to scalable policies
   - Progress in understanding the (long) chain from the first experiments to the final adoption of policy
   - Improving programs that run at scale ?
5. Experiments as innovations, useful in developing new products or policies and not just studying existing ones
   - Institutions to facilitate, fund, and incentivize innovation
   - RCTs are collaborative and iterative

The Impact of RCTs on Development Policy

- J-PAL/IPA directly affected policy in numerous ways, and almost all continents
- Many governments have launched either long run partnership with academic researchers or their own learning units' (e.g. Minedulab in Peru, Tamil Nadu research partnership)
- World Bank and Regional Development Banks support hundreds of RCTs and training with various governments

Randomization and Its Discontents

1. Jim Heckman's critique

   [..]Proponents of randomized social experiments implicitly make an important assumption: that randomization does not alter the program being studied. Bias induced by randomization is a real possibility.

2. Angus Deaton's critique

   [..]RCTs have no special status, they have no exemption from the problems of inference that econometricians have always wrestled with, and there is nothing that they, and only they, can accomplish.

3. John List's critique

   [..]RCTs' laudable goal has been undermined by a phenomenon known as the "scale-up problem" , which is defined here as the propensity for the absolute size of an intervention's treatment effect to systematically shrink, if not vanish, when that intervention is scaled up (or, more generally, for the benefit-cost profile to change at scale).

Randomization Bias: A Framework

- Denote with $A \in \{0,1\}$ the actual participation in given program, with $p = Pr(A=1)$
- Denote with $D \in \{0,1\}$ the counterfactual participation in a non-experimental regime
- Denote with $D^{\star} \in \{0,1\}$ the random selection indicator

AS-1 There is no effect of randomization on participation decisions

$$Pr(D=1) = Pr(D^{\star}=1|p)$$

AS-2 If AS-1 is violated, either:

(a) The effect of the treatment is the same for all participants, or

(b) if agents differ in their response to treatments, their idiosyncratic responses do not influence their participation decisions

Randomization Bias: A Framework

- Failure of AS-1 can be seen as an application of the Lucas (1981) critique in the context of social experimentation
  - Changing the program enrollment process by randomly denying access to individuals who apply and are deemed suitable for a program may make the distribution of $D^\star$ different from D
  - Such randomization alters the information set of potential applicants and program administrators
- In practice, there are many reasons to suspect the validity of this assumption
  - If individuals who might have enrolled in a nonrandomized regime make plans anticipating enrollment in training, adding uncertainty at the acceptance stage may alter their decision to apply or to undertake activities complementary to training. Risk-averse persons will tend to be eliminated from the program
  - If training centers must randomize after a screening process, it might be necessary for them to screen more persons in order to reach their performance goals, and this may result in lowered trainee quality

Other Instances of Randomization Bias

- Evaluation-driven effects
  - Treated individual may feel lucky or grateful or simply by being conscious of being observed exert more effort
  - Control individuals may also react by exerting more or less effort
- Demand and anticipation effects
  - Participants may react in response to their perception of what the evaluator is trying to test
  - Control people alter their behaviors as they expect to receive the program in the future
  - Just the fact of being surveyed may alter behavior (e.g. it provides a reminder to use the program)
- How to minimize those effects?
  - Blind randomization
  - Placebo group

Nothing Scales?

- One of the key drivers behind the increasing adoption of randomized evaluations has been a genuine ambition to directly inform policy making
  - Evidence from field experiments ensures that governments implement programs that actually work, and avoid those that do not work
- This laudable goal has been undermined by a "voltage effect"
  - Effect size of an intervention tends to shrink when it is scaled up
  - Cost-benefit profile changes at scale
- Vast waste of resources on seemingly effective policies that ultimately fail to deliver the expected benefits when implemented at larger scale

Causes of Voltage Drops: Five Vital Signs

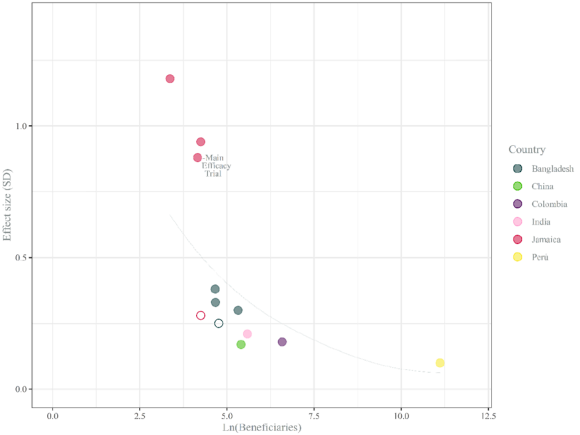
1. False positives
2. Representativeness of the sampled population
3. Spillovers
4. Supply side (diseconomies of scale)
5. Representativeness of the sampled situation

Example: the Jamaica Experiment

- A home-visit intervention aimed at improving parent-child interactions and child development
  1. Efficacy trial in Jamaica (N≈70)
  2. Pilot designed to allow replicability at scale in Colombia (N≈700)
  3. At-scale government program in Peru (N≈70,000)
- Plus a few other replications in other countries
- Some differences in implementation and targeted population

Effect Sizes of Different Replications of Jamaica

Figure 2. *Effect Sizes of Different Replications of the Jamaican Model*



Estimated effect sizes for the Jamaican Model at different scales

Experimentation at Scale

1. Representative samples of large populations
   - Randomization/Site selection bias (e.g. Allcott. 2015)
2. Feasible implementation protocols given government constraints
   - From evidence-based policy to policy-based evidence (List, 2022)
3. Large units of randomization
   - Randomizing at a large unit is essential to get total treatment effects incorporating spillovers (e.g. Miguel and Kremer, 2004)

RCTs and other methods

- RCTs as benchmark to evaluate accuracy of other ex-post studies
  - But this reflects a narrow view of empirical work in economics
- Deductive method does not always involve causal effects
  - Graphs and cross-tabulations can be so powerful when they arrange data in a way that contradicts a mass of prior understanding about how the world works
- An array of econometric approaches can be seen as complementary, not substitute, to RCTs

Wrapping-up on RCTs and their Critics

- RCTs have become a popular tool in development economics research
  - Owing to the close collaboration between researchers and implementers, RCTs allow the estimation of parameters that would not otherwise be possible to evaluate
  - RCTs allow for replication of research design across different contexts
- Perhaps inevitably, this progress has also generated a rising tide of criticism
  - Do not discard other methods!
  - Always be wary about scale-up issues

Figure: The Pool Game Analogy: Correlation

Figure: The Pool Game Analogy: Causation

back to main slide