Randomized Control Trials and Policy Evaluation

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

Toulouse School of Economics (TSE)

M2 PPD-EEP-EEE

Spring 2026

Overview of the Course

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, public, 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

Outline and Timeline of the Course

1. Overview and intro (week 1)
2. Econometrics of RCTs (weeks 2 to 5)
   - ⇒ Potential outcomes, SUTVA, and assignment mechanisms
   - ⇒ Randomization designs, estimation, and inference
3. Design, implementation and inference issues (weeks 5 to 7)
   - ⇒ Sample size considerations
   - ⇒ Non-compliance, spillovers, attrition, and multiple outcomes
4. Applications (weeks 8 to 10)
   - ⇒ Scaling-up RCTs
   - ⇒ Students' presentations

Course Material

1. Slides: on `matteobobba.github.io/teaching` + the course's Moodle page

   ⇒ 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"

2. TDs: software + data applications of topics covered in class

   ⇒ TA for this class is Kevin Frick (kevin-michael.frick@tse-fr.eu)

Course Requirements [relative weight]

1. Take-home exercise [25%]

   ⇒ Distributed during week 3, due by week 5

2. Paper presentation with slides [25%]

   ⇒ In-class presentations during weeks 9-10

3. Final written exam [50%]

   ⇒ March 31, 2026 at 9.30am (90 minutes)

Part 1: Introduction

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 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 set of restrictions on the distribution of S

$\Rightarrow$ Example. The linear (univariate) 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 causality between x and y
  - $\Rightarrow$ 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
  - $\Rightarrow$ The *ceteris paribus* condition: "other (relevant) factors being equal" Pictures
- This implies positing more assumptions on S
  - $\Rightarrow$ Example: if $\mathbb{E}(\epsilon_i|x_i) = 0$ then β is the causal effect of x_i on y_i

Characterization of Endogeneity Biases

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

Example of Simultaneity Bias

- Demand and supply model

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

$\Rightarrow$ Data on exchanged quantities y_i $(D_i = S_i)$

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

$\Rightarrow$ 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:

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

$\Rightarrow$ First-order condition for optimal schooling is:

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

$\Rightarrow$ Assume both MR and MC of schooling are linear

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

Example of Unobserved Heterogeneity Bias (Cont'd)

- Hence, optimal schooling in the linear case is:

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

- Minncerian regression

$$\ln(y)_i = \alpha + \beta S_i^\star + \epsilon_i$$

$\Rightarrow$ OLS estimate of β is a weighted average of $\bar{b}$ and $\bar{r}$

$\Rightarrow$ OLS $>$ average MR: people with higher b_i/lower r_i choose higher levels of $S_i^\star$

Example of Sample Selection Bias

- Individuals determine labor supply by trading off consumption for leisure

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

  - $\Rightarrow$ Interior solution: $v'(h^\star) = w$
  - $\Rightarrow$ Corner solution $v'(0) > w$
  - $\Rightarrow$ Reservation wage is $w^\star = v'(0)$. Work if $w \geq w^\star$

- Empirical specifications for reservation and offered wages

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

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

Example of Sample Selection Bias (cont'd)

- Individual i works $(D_i = 1)$ when

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

- We observe wages for individuals who work. Hence:

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

$\Rightarrow$ If $\epsilon_i \perp \nu_i$, OLS of wages on X_i recovers β on truncated sample (Tobit)

$\Rightarrow$ If $\epsilon_i \not\perp \nu_i$, $\mathbb{E}(\epsilon_i|X_i, D_i = 1) \neq 0$ and OLS is inconsistent (Heckit)

The Promise of RCTs

- Field experiments induce controlled variations in a policy variable: $D_i \perp \epsilon_i$
  - $\Rightarrow$ Sources of exogenous variation in real world economic environments
  - $\Rightarrow$ Transparent and easy-to-replicate across different contexts
- RCTs provide causal estimates with minimal statistical assumptions
  - $\Rightarrow$ Valid in the sample under study + when correctly designed and implemented

The case for and against RCTs

Historical Background

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

- By the early 1980s, there were more than 70 social experiments in the US
  - ⇒ Education and training
  - ⇒ Employment programs and income transfers

⇒ Small-scale, government run and individual-level randomization

An Example: The Perry Pre-school Project

- Five cohorts drawn from area surrounding the Perry elementary school

  ⇒ 123 children aged 3 in disadvantaged african-american families

  ⇒ 2.5 hours preschool plus weekly home visits by teachers for two years

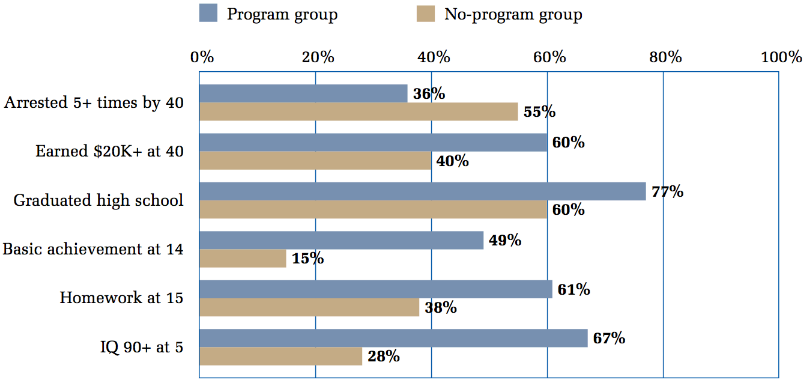
- Each cohort is followed over entire life

  ⇒ Annual surveys till age 15, follow-ups at ages 19, 27 and 40

The Perry Pre-school Project (Cont'd)

Figure 1

Major Findings: High/Scope Perry Preschool Study at 40



The Rise of RCTs in Development Economics

1. Proof-of-concept randomized trials

   ⇒ Does it work? E.g. bundled interventions

2. From proof-of-concepts to field experiments

   ⇒ Why does it work? E.g. cross cutting designs

3. From field experiments to large-scale interventions

   ⇒ Can this idea work at scale? E.g. at-scale randomization designs

The Impact of RCTs on Economics Research

- Micro-based approach
  - $\Rightarrow$ Breaking down a question in pieces, each of which could be studied via RCTs
- A greater focus on identification across the board
  - $\Rightarrow$ Large impact on other design-based methods and model-based approaches

The Impact of RCTs on Development Policy

- Evidence on mechanisms behind poverty and interventions to alleviate it
- Discover new products or policies and not just study existing ones
  - ⇒ Partnerships to facilitate, fund, and incentivize innovation
  - ⇒ Train government officials in evidence-based policy making
  - ⇒ World Bank and other multilaterals support hundreds of RCTs

Randomization and Its Discontents

⇒ 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.*

⇒ 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.*

⇒ John List's critique

> *RCTs' laudable goal has been undermined by a phenomenon known as the voltage drop, 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

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

  $\Rightarrow$ There is no effect of randomization on participation decisions

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

- How to avoid discrepancies between the distributions of $D^\star$ and D?

  $\Rightarrow$ Blind randomization

  $\Rightarrow$ Placebo group

Randomization Bias: Examples

- RCT-driven effects
  - $\Rightarrow$ Treated individual may exert more effort as they feel lucky or grateful
  - $\Rightarrow$ Control individuals may also react by exerting more or less effort
- Demand and anticipation effects
  - $\Rightarrow$ Participants may react in response of what the evaluator is trying to test
  - $\Rightarrow$ Control people alter their behaviors as they expect to receive the program
- Survey-driven effects
  - $\Rightarrow$ Surveys as nudges (e.g. they provide a reminder to use the program)

RCTs Ain't Magical

- $\mathbb{E}(Y_i^1 - Y_i^0) = \mathbb{E}(Y_i \mid D_i = 1) - \mathbb{E}(Y_i \mid D_i = 0)$

  $\Rightarrow$ This is not true for the median or any percentile of the distribution of treatment effects, or its variance

- Estimate is sample/population-specific

  $\Rightarrow$ Difficult to extrapolate to other settings

- Randomization balances treat and control groups in expectation

- Inference and implementation issues

  $\Rightarrow$ Skewness, outliers, multiple outcomes, heterogeneity, non-compliance,...

Nothing Scales?

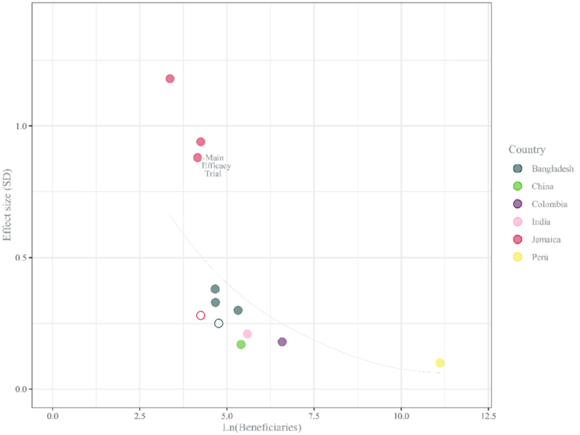
- Seemingly effective policies fail to deliver the expected benefits at larger scale
  - ⇒ Effect size of an intervention tends to shrink when it is scaled up
  - ⇒ Cost-benefit profile changes at scale
- Causes of "voltage drops"
  - ⇒ False positives
  - ⇒ Representativeness of the sampled population
  - ⇒ Representativeness of the sampled situation
  - ⇒ Spillovers/equilibrium effects
  - ⇒ Supply side considerations (diseconomies of scale)

The Jamaica Experiment

- Home-visits for improving parent-child interactions and child development
  - ⇒ Pilot/Proof-of-concept in Jamaica (N≈70)
  - ⇒ RCT designed to allow replicability at scale in Colombia (N≈700)
  - ⇒ 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***

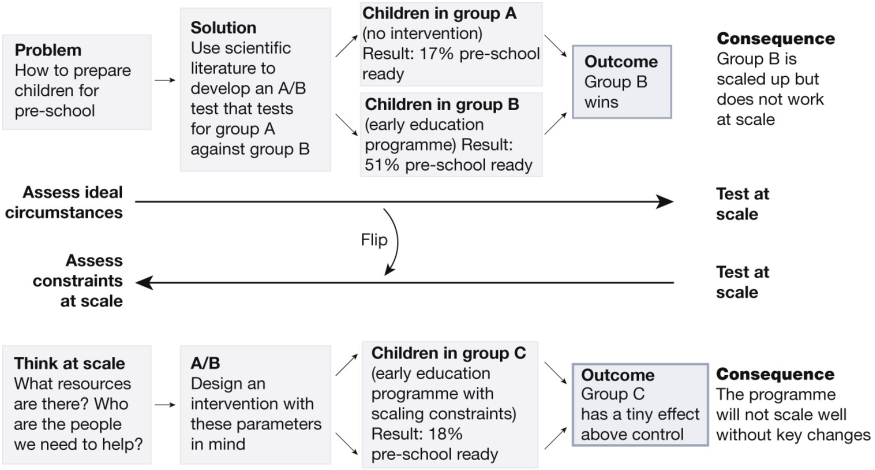


Estimated effect sizes for the Jamaican Model at different scales

RCTs at Scale

- Large and representative samples of entire populations
  - ⇒ Avoid sample/site selection bias
- Feasible implementation protocols given government constraints
  - ⇒ From evidence-based policy to policy-based evidence
- Large units of randomization
  - ⇒ Total treatment effects incorporate spillovers

Policy-based Evidence: An Example



Wrapping-up on Today's Intro Class

- RCTs have become a popular tool in empirical research
- This progress has also generated a rising tide of criticism
- RCTs are nice complements (not substitutes!) of other empirical methods
- Always be wary about spillovers, extrapolation, and/or scale-up issues

Appendix

Figure: The Pool Game Analogy: Correlation

Figure: The Pool Game Analogy: Causation