Lecture 1: Public policy evaluation, causal inference and randomization

P.J. Messe[1]

[1]Le Mans Université GAINS-TEPP, CEET, LEMNA

Master in Applied Econometrics

The goal of this course

- To give a toolkit for public policy evaluation
  - A list of the main microeconometric tools to evaluate the impact of public policies
  - A user's guide for each tool

The goal of this course

- To give a toolkit for public policy evaluation
  - A list of the main microeconometric tools to evaluate the impact of public policies
  - A user's guide for each tool
- A course divided into two parts:
  - A theoretical one: to show the terminology, the equations and the required diagnosis test/robustness checks related to each method: 12H Lectures
  - An applied one: to manipulate data sets and apply each microeconometric approach to evaluate the causal effect of different public policies: 12H Tutorials

Resources/documentation

- Mainly based on the textbook of Angrist and Pischke (2008)
- Completed by influential articles related to the different methodologies (available in MADOC)
- Own professional experience
  - Evaluation of a new job training program targeted to job-seekers implemented by the Pays de la Loire Regional Council in 2020
  - Data sets and evaluation issues already studied in previous research work

Note: microeconometrics has not the monopoly of policy evaluation

- This course is only about microeconometric tools of policy evaluation
- But other approaches exist in the field of policy evaluation
  - Theoretical approach: building a model in which we explicitly describe the agents' behaviour : micro or macro to have general equilibrium effects
  - Structural approach: combines the insights from a theoretical model and the empirical support of data, used to estimate the unknown parameters of the model
  - A growing interest for this latter since it allows to address the Lucas critique (1976): agents REACT to the implementation of policies.

Evaluation

- 2 evaluations
  - One in the middle of the semester
  - Final exam

Outline

Causal inference and confounding

- The Neyman-Rubin's framework
- The selection bias

Randomized experiments: a gold standard?

- Randomized experiments and selection bias
- The other advantages of RCTs
- The concerns about RCTs

A key notion: potential outcomes

- Donald B. Rubin: a statistician that popularized the "potential outcomes" framework in the 1970's
  - Initially introduced by Neyman (1923)
  - Framework that borrows many terms in the field of medical experimentation

A key notion: potential outcomes

- Donald B. Rubin: a statistician that popularized the "potential outcomes" framework in the 1970's
  - Initially introduced by Neyman (1923)
  - Framework that borrows many terms in the field of medical experimentation

- Main question: what is the effect of a **treatment** on a **unit**
  - Treatment=public policy: labour market policies (training program, change in eligibility conditions to UI, rise in retirement age, ...) ; educational policies (class size reduction, ...), health policies (evolution in the health care system generosity, ...)
  - Unit=individual, firm, locality (country, region, ...)

A key notion: potential outcomes

- We associate each treatment/unit pair with a potential outcome
  - Potential: only one can be realized and possibly observed: the potential outcome corresponding to the treatment actually taken at that time.
  - Assessing the causal effect of treatment involves comparison of these potential outcomes: some realized and some others no realized.

A key notion: potential outcomes

- Some notations:
  - Outcome for an individual i: Y_i
  - The treatment: a binary random variable D_i
  - Individuals can be treated (treatment group) or non-treated (control group)

A key notion: potential outcomes

- Some notations:
  - Outcome for an individual i: Y_i
  - The treatment: a binary random variable D_i
  - Individuals can be treated (treatment group) or non-treated (control group)
- For any individual there are two potential outcomes:

$$potential \quad outcome = \begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases}$$

- The observed outcome writes as:

$$Y_i = (1 - D_i)Y_{0i} + D_i Y_{1i}$$

$$= Y_{0i} + D_i \quad x \qquad \underbrace{(Y_{1i} - Y_{0i})}_{}$$

causal effect of treatment for an individual

A common assumption in the potential outcomes approach: the no-interference assumption

- The notations used rely on a strong assumption: The Stable Unit Treatment Value Assumption (SUTVA)
  - The treatment D_i only affects the individual i : no spillover effects of the treatment on non-treated individuals

A common assumption in the potential outcomes approach: the no-interference assumption

- The notations used rely on a strong assumption: The Stable Unit Treatment Value Assumption (SUTVA)
  - The treatment D_i only affects the individual i : no spillover effects of the treatment on non-treated individuals

- If we relax this assumption, the problem becomes more complex. In a setting of N units:
  - Let $\mathbf{D}$ the N-vector assignment with typical element D_i.
  - If potential outcomes can depend on the treatments for ALL units each unit has 2^N different potential outcomes $Y_i(\mathbf{D})$

The key issue of evaluation

- One big issue: missing data
  - We cannot observe at the same time Y_{0i} AND Y_{1i} for the same individual

The key issue of evaluation

- One big issue: missing data
  - We cannot observe at the same time Y_{0i} AND Y_{1i} for the same individual

- One naive estimator: looking at the difference between the average observed outcome for the treated individuals and the non-treated ones

$$E(Y_i|D_i=1) - E(Y_i|D_i=0) = \underbrace{E(Y_{1i}|D_i=1) - E(Y_{0i}|D_i=0)}_{\text{Observed difference in average outcome}}$$

Outline

A key notion: potential outcomes

- One big issue: **selection bias**
  - Treated and non-treated individuals may be **strongly different**
  - Selection into treatment may be driven by observable or unobservable characteristics that influence the outcome.
  - A lack of independence between treatment and potential outcomes = **confounding**.

A key notion: potential outcomes

- One big issue: **selection bias**
  - Treated and non-treated individuals may be **strongly different**
  - Selection into treatment may be driven by observable or unobservable characteristics that influence the outcome.
  - A lack of independence between treatment and potential outcomes = **confounding**.

- Rewriting the naive estimator, we exhibit the selection bias

$$E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 0) = \underbrace{E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 1)}_{\text{Average effect of the treatment on the treated: ATT}}$$

$$+ \underbrace{E(Y_{0i}|D_i = 1) - E(Y_{0i}|D_i = 0)}_{\textbf{selection bias}}$$

A key notion: potential outcomes

- The Average effect of Treatment on Treated: ATT
  - Difference between observed outcome for the treated $E(Y_{1i}|D_i = 1)$ and the **counterfactual** outcome for this group $E(Y_{0i}|D_i = 1)$
  - The counterfactual is not observed: what would have happened to treated individuals had they not been treated?

A key notion: potential outcomes

- The Average effect of Treatment on Treated: ATT
  - Difference between observed outcome for the treated $E(Y_{1i}|D_i = 1)$ and the **counterfactual** outcome for this group $E(Y_{0i}|D_i = 1)$
  - The counterfactual is not observed: what would have happened to treated individuals had they not been treated?
- The selection bias
  - The difference between this counterfactual outcome and the observed outcome for the non-treated individuals $E(Y_{0i}|D_i = 0)$
  - If treated individuals are more likely to have a higher value of outcome: selection bias is positive (upward biased estimates)
  - If treated individuals are more likely to have a lower value of outcome: selection bias is negative (downward biased estimates)

A key notion: potential outcomes

- Another estimand more interesting than ATT: **Average Treatment Effect (ATE)**
  - What would happen if we generalize the treatment **to ALL individuals?**
  - TWO counterfactuals have to be estimated
    - what would have happened to treated individuals had they not been treated? (as for the ATT)
    - what would have happened to **non-treated** individuals had they been treated?

A key notion: potential outcomes

- Another estimand more interesting than ATT: **Average Treatment Effect (ATE)**
  - What would happen if we generalize the treatment **to ALL individuals?**
  - TWO counterfactuals have to be estimated
    - what would have happened to treated individuals had they not been treated? (as for the ATT)
    - what would have happened to **non-treated** individuals had they been treated?
- ATT-ATE is it the same?
  - NO in case of HETEROGENEOUS effects of the treatment
  - NO if SUTVA does not hold (cf discussion in section 2)
  - NO in case of general equilibrium effects (cf discussion in section 2)

To practice: Potential outcomes applied to the example of training program targeted to job-seekers

- For any individual there are two potential outcomes:

$$\begin{cases} Y_{1i} & \text{:earnings of one job-seeker who is assigned to this specific training program} \\ Y_{0i} & \text{:earnings of one job-seeker who is NOT assigned to this specific program} \end{cases}$$

- BUT non-treated individuals may benefit from other job-search assistance or training programs

- A guess on the selection bias:
  - Only more motivated/skilled individuals receive training: upward bias
  - This specific program is targeted on individuals with poor job opportunities and high unemployment duration: downward bias

Outline

Random assignment solves the selection problem

- A key piece of information: how each individual came to receive the treatment
  - What is the assignment mechanism?

$$E(D_i | Y_0, Y_1)$$

Random assignment solves the selection problem

- A key piece of information: how each individual came to receive the treatment
  - What is the assignment mechanism?

$$E(D_i|Y_0, Y_1)$$

- Randomization is based on coin tossing, random number generation in a computer ...
  - Randomization implies that the treatment is statistically independent of potential outcomes
  - A randomized assignment mechanism is **ignorable**: the treatment assignment is **unconfounded**

$$(Y_1, Y_0) \perp D \Rightarrow E(Y_{0i}|D_i = 1) = E(Y_{0i}|D_i = 0)$$

Random assignment solves the selection problem

- In the case of a randomized assignment mechanism, the naive estimator writes as:

$$E(Y_{1i}|D_i=1)-E(Y_{0i}|D_i=0) = \underbrace{E(Y_{1i}|D_i=1)-E(Y_{0i}|D_i=1)}_{\text{Average effect of the treatment on the treated: ATT}}$$

$$+ \underbrace{E(Y_{0i}|D_i=0)-E(Y_{0i}|D_i=0)}_{=0 \quad \textbf{removal of the selection bias}}$$

Random assignment solves the selection problem

- In the case of a randomized assignment mechanism, the naive estimator writes as:

$$E(Y_{1i}|D_i=1)-E(Y_{0i}|D_i=0) = \underbrace{E(Y_{1i}|D_i=1)-E(Y_{0i}|D_i=1)}_{\text{Average effect of the treatment on the treated: ATT}}$$

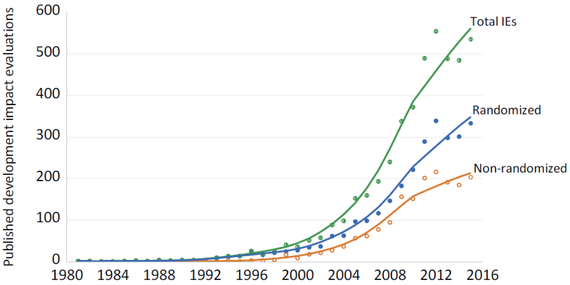
$$+ \underbrace{E(Y_{0i}|D_i=0)-E(Y_{0i}|D_i=0)}_{=0 \quad \textbf{removal of the selection bias}}$$

- Randomized experiments (or Randomized Controlled Trials: RCTs) yields **unbiased** estimate of the ATT
  - This explains the rising popularity of RCTs in public policy evaluation, in particular in development economics, since 2000.

A huge increase in RCTs in public policy evaluation in developing countries (Ravallion, 2020)

- About 60% of impact evaluations since 2000 have used RCTs (Ravallion, 2020)

Figure 1: Annual counts of published impact evaluations for developing countries



Note: Fitted lines are nearest neighbor smoothed scatter plots. See footnote 4 in the main text on likely under-counting of non-randomized evaluations in earlier years. Source of primary data: International Initiative for Impact Evaluation.

Outline

Stimulating innovative policies in development economics

- To improve educational outcomes in developing countries (reducing absenteeism, raising educational level), economic theory calls for:
  - Financial incentives (cash transfer conditional on investment in education).
  - Hiring extra teacher to reduce the teacher/student ratio (reduction in class sizes)

Stimulating innovative policies in development economics

- To improve educational outcomes in developing countries (reducing absenteeism, raising educational level), economic theory calls for:
  - Financial incentives (cash transfer conditional on investment in education).
  - Hiring extra teacher to reduce the teacher/student ratio (reduction in class sizes)

- Some programs have been implemented in this purpose but they are costly (J-PAL, 2005): For an extra child-year of education
  - Conditional cash transfers (ex : PROGRESA): 6000 $
  - Extra-teacher programs (Banerjee et al., 2005): 60 $

RCTs at the origin of more cost-effective policies

- A treatment for intestinal worms in school in Kenya (Miguel and Kremer, 2004): **3.50$** per extra child-year of education
- Conditioning teachers' salary to pictures took twice a day to confirm the teacher/student' attendance in India (Duflo et al., 2007)

RCTs allows to decompose the effect of each element of a program

- Researchers are encouraged to implement multiple treatment experiments (Banerjee and Duflo, 2009)
  - Testing first the overall program
  - And then delving into its individual components to understand what works by varying components of treatment

RCTs allows to decompose the effect of each element of a program

- Researchers are encouraged to implement multiple treatment experiments (Banerjee and Duflo, 2009)
    - Testing first the overall program
    - And then delving into its individual components to understand what works by varying components of treatment

- Illustrations in education economics: hiring extra-teacher influences educational outcomes through three channels (Duflo et al., 2009)
    - A reduction in the teacher/student ratio
    - The use of resources by the school committee who has to hire/fire the teacher
    - The teacher type (contract teacher VS civil service one)

RCTs allows to decompose the effect of each element of a program

- A subtle design of a multiple treatment experiment allows to disentangle the effect of each channel
- For this experiment
  - An increase in resources alone has no effect on test scores (difference between comparison and treatment group non-significant)
  - A significant positive effect of hiring a contract teacher (more incentives to do a good job) than hiring a regular one
  - Additional positive effect of offering an extra training for school committee members

RCTs encourage a real involvement of researchers in public evaluation and implementation

- Economist usually deliver reports of policy recommendations... and that's it.
  - They do not engage in the details of the implementation of the policy

RCTs encourage a real involvement of researchers in public evaluation and implementation

- Economist usually deliver reports of policy recommendations... and that's it.
    - They do not engage in the details of the implementation of the policy
- RCTs change the role of economists: they become plumbers (Duflo, 2017)
    - Not only installing a "machine" (new policy/program)
    - But also looking at all the gears/joints of this machine
    - More interested in "how" to do something than "what" to do.

Economists as plumbers

- Ex : improving access to private water connections in Tangiers offering interest-free loans to poor households: OK
- BUT in reality, the take-up of this subsidized loan program was very low: 10%
- This policy becomes really effective (take-up rate=69%)...
- when a team of economists offered procedural assistance photocopying the required documents and delivering them to the municipal office (Devoto et al., 2012)

Outline

Environmental dependance

- Evaluations conducted in a few locations with specific organizations
  - While internally valid, can we generalize the results to other locations/settings?
  - Without a better understanding of the underlying mechanisms risk of misleading conclusions

Environmental dependance

- Evaluations conducted in a few locations with specific organizations
  - While internally valid, can we generalize the results to other locations/settings?
  - Without a better understanding of the underlying mechanisms risk of misleading conclusions

- Ex : Some studies show the effectiveness of offering a deworming treatment / insecticide-treated bed nets
  - But all these studies have been conducted in specific locations
  - What could be expected from implementing these policies in developed countries? Quite nothing.
  - Because the channels of these policies are specific to the countries studied (exposition to malaria, ...)

Environmental dependance OK but for all microeconometric studies

- External validity cannot be guaranteed neither in RCTs nor in non-experimental (observational) studies

- RCTs can be replicated in different settings/locations to understand the context dependence of programs (Banerjee et al., 2017)
    - For non-experimental studies, different results may also stem from the presence of biases (not the case for RCTs)
    - Some initiative to encourage replications of RCTs: systematic reviews, meta-analysis, special issues in Top 5 journals
    - The goal: not necessarily generalizing results but understanding what works what does not work.

The randomization bias

- Most of the time, ethics committees requires full disclosure : participants in a RCTs have to be informed (treated and control group).
  - Individuals can refuse to participate to the experiment

The randomization bias

- Most of the time, ethics committees requires full disclosure : participants in a RCTs have to be informed (treated and control group).
  - Individuals can refuse to participate to the experiment
- Participants may change their behaviour because they know that this program is evaluated
  - Hawthorne effect: attention to individuals may change their attitudes ("lighting" study in Hawthorne plant)
  - John Henry effect: individuals of the **control** group could do additional efforts to overcome the treated individuals' performance $\Rightarrow$ underestimate the effect of RCTs.

The randomization bias

- One other specific issue of RCTs: the site/organization-selection bias (Heckman, 1992; Banerjee et al., 2017)
  - RCTs are conducted in collaboration with SELECTED NGOs...
  - That choose specific locations/sites to conduct the experiment

The randomization bias

- One other specific issue of RCTs: the site/organization-selection bias (Heckman, 1992; Banerjee et al., 2017)
    - RCTs are conducted in collaboration with SELECTED NGOs...
    - That choose specific locations/sites to conduct the experiment
- This selection of site is not arbitrary: organizations work where they think the impact is the greatest
    - Because of limited resources
    - And because they are subject to an evaluation

The randomization bias

- One other specific issue of RCTs: the site/organization-selection bias (Heckman, 1992; Banerjee et al., 2017)
  - RCTs are conducted in collaboration with SELECTED NGOs...
  - That choose specific locations/sites to conduct the experiment
- This selection of site is not arbitrary: organizations work where they think the impact is the greatest
  - Because of limited resources
  - And because they are subject to an evaluation
- This site/organization selection has a significant impact of RCTS' results (Vivalt, 2016)
  - Comparing results of 400 RCTs, those run by NGO (or small-scale studies) tend to have higher effects than those run with governments

Market equilibrium effects

- Results obtained from small scale studies are difficult to turn into recommendations for large-scale policies
  - A nationwide policy is likely to have general equilibrium effects that are ignored in small-scale RCTs

Market equilibrium effects

- Results obtained from small scale studies are difficult to turn into recommendations for large-scale policies
  - A nationwide policy is likely to have general equilibrium effects that are ignored in small-scale RCTs

- Ignoring equilibrium effects may lead to overestimate the overall benefits of a program
  - Small-scale RCTs show that a scholarship program aiming at improving education levels has strong positive effects on earnings (Duflo et al., 2017)
  - But the generalization of this program would increase the supply of skilled workforce $\Rightarrow$ Lower returns to education (Heckman et al., 1999)

Market equilibrium effects/spillover effects

- Potential outcomes approach (not only RCTs) usually limit the possibility of interactions between units (SUTVA)
  - But in some settings, this assumption is not credible
  - And challenges the effectiveness of some programs found in partial equilibrium evaluations

Market equilibrium effects/spillover effects

- Potential outcomes approach (not only RCTs) usually limit the possibility of interactions between units (SUTVA)
  - But in some settings, this assumption is not credible
  - And challenges the effectiveness of some programs found in partial equilibrium evaluations
- A positive effect of active labor market programs on employment (Card et al., 2018)?
  - BUT if the total number of vacancies is not affected by such program they turn out to be zero-sum games
  - This stimulates employment among the treated group BUT reduces employment among the control group
  - To assess the impact of this policy at the COUNTRY-level we must account for these general equilibrium effects.

Market equilibrium effects/spillover effects

- RCTs may be used to exhibit these spillover effects with a two-stage randomization procedure (Crepon et al., 2013)
  - The proportion of job-seekers (0%, 25%, 50%, ...) to be assigned to a job search assistance program randomly assigned for each local labor market participating to the RCT
  - Randomized assignment into treatment within each local labor market

Market equilibrium effects/spillover effects

- RCTs may be used to exhibit these spillover effects with a two-stage randomization procedure (Crepon et al., 2013)
  - The proportion of job-seekers (0%, 25%, 50%, ...) to be assigned to a job search assistance program randomly assigned for each local labor market participating to the RCT
  - Randomized assignment into treatment within each local labor market

- Empirical evidence of strong displacement effects
  - The program improves the search ability of treated workers
  - BUT reduced the relative job-search success of those who do not benefit from the program
  - The overall effect of the policy is almost nil

Market equilibrium effects/spillover effects

- In some cases, ignoring spillover effects may UNDERestimate the overall effect of a program derived from small-scale RCTs
  - RCT find low impact of micro-credit on beneficiaries' consumption (Banerjee et al., 2015)
  - BUT this ignores potential multiplier effects: some part of the consumption benefits to suppliers of non-tradable goods who consume more ...

Market equilibrium effects/spillover effects

- In some cases, ignoring spillover effects may UNDERestimate the overall effect of a program derived from small-scale RCTs
  - RCT find low impact of micro-credit on beneficiaries' consumption (Banerjee et al., 2015)
  - BUT this ignores potential multiplier effects: some part of the consumption benefits to suppliers of non-tradable goods who consume more ...

- In the case of program involving positive externalities, results from RCTs underestimate overall effect of the program
  - Introduction of insecticide-treated bed nets may have small effects on malaria in small-scale studies (Tarozzi et al., 2014)
  - BUT scaling up this experiment should lead to positive externalities (ex: fewer mosquitos due to contact with insecticide)
  - making the policy more effective (Cohen and Dupas, 2010)

Market equilibrium effects/spillover effects

- Education or training programs are also expected to have strong positive spillover effects
  - Peer effects stemming from knowledge spillovers or peer pressure (Cornelissen et al., 2017)

Market equilibrium effects/spillover effects

- Education or training programs are also expected to have strong positive spillover effects
  - Peer effects stemming from knowledge spillovers or peer pressure (Cornelissen et al., 2017)
- Spillover effects could also be informational: more users of a technology/program $\Rightarrow$ better information $\Rightarrow$ higher take-up rate (Dupas, 2014)
  - Free distribution of insecticide-treated bed-nets in Kenya increase the willingness to pay of the treated individuals' neighbors one year later

Ethical issues and feasibility of RCTs

- Main critics against RCT: some treated individuals would not need the treatment and some non-treated would need this
  - One point commonly used by policy-makers to refuse the RCTs

Ethical issues and feasibility of RCTs

- Main critics against RCT: some treated individuals would not need the treatment and some non-treated would need this
  - One point commonly used by policy-makers to refuse the RCTs
- One solution that raises new issues: phased-in design of RCTs
  - To improve acceptance of control units: planning an expansion of the program in the future for these units
  - Potential biases if the anticipation of receiving the treatment later leads non-treated individuals to change their behaviors.

Ethical issues and feasibility of RCTs

- Other solution to tackle policy-makers' reluctance: blocked/stratified RCT
  - First identifying a subset of individuals who really need the program
  - Then randomly assigning the treatment among this subset of individuals

Ethical issues and feasibility of RCTs

- Other solution to tackle policy-makers' reluctance: blocked/stratified RCT
  - First identifying a subset of individuals who really need the program
  - Then randomly assigning the treatment among this subset of individuals

- This design can also disallow undesirable assignments
  - Random assignment does not ensure that characteristics are balanced across both groups (treatment and control)
  - If randomization leads to a treatment group entirely composed of specific individuals, resulting estimates are likely to be uninformative.

Ethical issues and feasibility of RCTs

- Ethics in medical research considers the equipoise principle
  - OK for randomizing if we are ignorant about whether it is better to be in the treatment or control group

Ethical issues and feasibility of RCTs

- Ethics in medical research considers the equipoise principle
  - OK for randomizing if we are ignorant about whether it is better to be in the treatment or control group

- But in social sciences RCTs, equipoise rarely applies (Ravallion, 2020) and goes in the opposite direction
  - Funders of RCTs gave preference to RCT proposal likely to have sizeable impacts on outcomes
  - In that case one could be worried about withholding a treatment from those who need it
  - Especially when it is about health mortality (ex : Cohen and Dupas (2010) : households who do not benefit from free distribution of bed nets experienced a higher child mortality rate)

Ethical issues and feasibility of RCTs

- In order to improve knowledge RCTs may have negative consequences

Ethical issues and feasibility of RCTs

- In order to improve knowledge RCTs may have negative consequences
- A financial reward offered to Indian individuals to help them having quickly a drivers' licence (Bertrand et al., 2007)
  - The study aims at verifying the effect of corruption: driving licence are obtained bribing officials
  - This has been verified: so one part of the RCT's cost has enriched some corrupted officials
  - And some individuals have had licence without really knowing how to drive

Ethical issues and feasibility of RCTs

- Such ethically-contestable evaluations may be justified by expected benefits from new knowledge
  - but the choice between RCT or non-expermental study should better weigh these benefits with regards to the generated costs