Evaluating public policies

Master Economics & Public Policies,
School of Public affairs

Denis Fougère (Sciences Po) & **Arthur Heim** (PSE & Cnaf)

January 30, 2023

Forwords

Who am I ?

- PhD candidate (5^{th} year) at Paris School of Economics (EHESS) with Marc Gurgand (PSE-CNRS)
  - CIFRE Funding from ANRT and the national family allowance fund (CNAF)
  - Now Research and evaluation officer at the Directorate for statistics, research and evaluation National familiy allowance fund (Cnaf)
- Main work: public policy evaluations (early childhood, education and labor)
- Other research interests for market design
- Former scientific advisor at France Stratégie
- Teaching causal inference and evaluation of public policies at Sciences po (since 2018), previously at Dauphine too (2019-2020).
- Former positions at national council for school system evaluation (Cnesco), France Stratégie and Cnaf

Forwords

What's the plan

- Main class: 4 first sessions with me, 8 nexts with Pr. Fougère
  - Main tools for causal inference including randomized control trials, instrumental variable, difference in difference, synthetic controls, regression discontinuity and matching.
  - Discussions of important articles as illustrations
- TA lectures: 8 sessions with me: How to do public policy evaluation **in practice**

† More advanced than the main class
  - Overview the most recent trends in the literature
  - Use simulations or open-access data to practice
  - Learn how to make reproducible research and dynamic documents using R, Rmarkdown and LaTeX
- The contents from the TA sessions are NOT in the exams, but they help a lot.

Forwords

Schedules and programs

Main class

1. 01/30/2023 (10h15-12h15): Arthur Heim: Introduction to evaluation and causal inference
2. 02/06/2023 (10h15-12h15): Arthur Heim: The conditional independence assumption and randomization
3. 02/13/2023 (10h15-12h15): Arthur Heim: Advanced experimental design
4. 02/20/2023 (10h15-12h15): Arthur Heim: Difference-in-differences, part I
5. 03/06/2023 (10h15-12h15): Denis Fougère: Difference-in-differences, part II
6. 03/13/2023 (10h15-12h15): Denis Fougère: Instrumental variables, part I
7. 03/20/2023 (10h15-12h15): Denis Fougère: Mid-Terms
8. 03/27/2023 (10h15-12h15): Denis Fougère: Instrumental variables, part II
9. 04/03/2023 (10h15-12h15): Denis Fougère: More about LATE
10. 04/17/2023 (10h15-12h15): Denis Fougère: Regression discontinuity design (Part I)
11. 04/24/2023 (10h15-12h15): Denis Fougère: Regression discontinuity design (Part II)
12. 04/28/2023 (10h15-12h15): Denis Fougère: Matching

Forwords

Schedules and programs

TA Sessions

1. 02/13/2023 (12:30-14:30): Topics on regressions: theory and practice
2. 03/06/2023 (12:30-14:30): Conducting research : power, replicability
3. 03/13/2023 (12:30-14:30): RCT in practice
4. 03/20/2023 (12:30-14:30): Applied difference in differences
5. 03/27/2023 (12:30-14:30): Synthetic controls
6. 04/03/2023 (12:30-14:30): Instrumental variables
7. 04/17/2023 (12:30-14:30): Regression discontinuity design
8. 04/24/2023 (12:30-14:30): Wrap-Up

Forwords

Online texbooks/ressources you should read (they're great !)

- **Latest book on causal inference** Nick Huntington-Klein. 2021. *The Effect: An Introduction to Research Design and Causality | The Effect.* December 21, 2021
- **Online book on causal inference** Scott Cunningham. 2018. *Causal Inference: The Mixtape*
- **10 fundamental theorems for econometrics** Thomas S. Robinson. 2020. *10 Fundamental Theorems for Econometrics.* Vol. v0.1. September 30, 2020
- **Econometrics using R** Constantin Colonescu. 2016. *Principles of Econometrics with R.* September 1, 2016
- **On R** Hadley Wickham and Garrett Grolemund. 2017. "Welcome | R for Data Science," January
- **Coding** Vikram Singh Rawat. 2021. *Best Coding Practices for R.* April 27, 2021

My repositories with useful ressources for you

- **R cheatsheets and books**: Link
- **Statistical method textbooks using R**: Link
- **Econometrics Textbooks**: Link

Forwords

Textbooks this class is built on

- Huntington-Klein 2021
- Cunningham 2018
- Joshua D. Angrist and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press
- Joshua D Angrist and Jörn-Steffen Pischke. 2015. *Mastering 'Metrics: The Path from Cause to Effect*
- Jeffrey M Wooldridge. 2012. "Introductory Econometrics: A Modern Approach," 910

Forwords

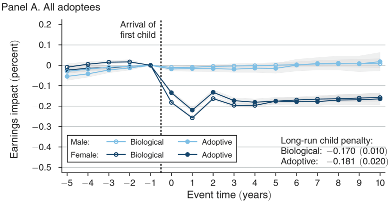
To go further

- Judea Pearl and Dana Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect.* New York: Basic Books. ISBN: 978-0-465-09761-6
- A Colin Cameron and Pravin K Trivedi. 2005. *Microeconometrics : Methods and Applications.* Cambridge University Press
- Guido W. Imbens and Donald B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* Cambridge: Cambridge University Press
- Martin Ravallion. 2007. "Chapter 59 Evaluating Anti-Poverty Programs." In *Handbook of Development Economics,* 4:3787–3846. Elsevier. ISBN: 978-0-444-53100-1
- Susan Athey and Guido W. Imbens. 2017b. "The State of Applied Econometrics: Causality and Policy Evaluation." *Journal of Economic Perspectives* 31, no. 2 (May): 3–32
- S. Athey and G. W. Imbens. 2017a. "Chapter 3 - The Econometrics of Randomized Experiments." In *Handbook of Economic Field Experiments,* edited by Abhijit Vinayak Banerjee and Esther Duflo, 1:73–140. Handbook of Field Experiments. North-Holland, January 1, 2017

Session I
Introduction to evaluation & causal inference

Evaluating public policies

Arthur Heim (PSE & Cnaf)



Source: Kleven, Landais, and Sø gaard (2021)

Outline

1. Introduction
   - What is program evaluation ?
   - Causality
   - Public policy evaluation: a subfield of causal inference in economics
   - Intuition from an example
   - Definitions
2. Potential outcomes and causal relationships
3. The rise of design-based economics
4. Wrap-up

> *". . . we think of a cause as something imperative which exists in the exterior world. In my opinion this is fundamentally wrong. If we strip the word cause of its animistic mystery, and leave only the part that science can accept, nothing is left except a certain way of thinking. [T]he scientific . . . problem of causality is essentially a problem regarding our way of thinking, not a problem regarding the nature of the exterior world."*

Frisch (1930) quoted in James J. Heckman and Rodrigo Pinto. 2022. "The Econometric Model for Causal Policy Analysis." *Annu. Rev. Econ.* 14, no. 1 (August 12, 2022): 893–923

Introduction

What is program evaluation ?

- Program evaluation is a systematic method for:
  - **collecting, analysing, and using information**,
  - answering questions about projects, policies and programs, and
  - measurements, particularly about their **effectiveness and efficiency**
- Evaluation became particularly relevant in the U.S. in the 1960s during the period of the Great Society social programs (elimination of poverty and racial injustice)
- Program evaluations can involve **both quantitative and qualitative** methods of social research
- People who do program evaluation come from many different fields, such as sociology, psychology, economics, social work, and public policy

What is program evaluation ?

The relationship between evaluation and public policies

Two important things about this relation:

1. "The first is that information, and by extension the brand of information that is revealed by evaluation, is critical for deciding what to do next. (...) Action that does not take careful account of what is working and not working stands a chance of making social situations worse."
2. "But information is relevant to more than decisions. Evaluators have often worried about the use of their finding in terms of their effects on policy decisions. Did decision makers change the program to remedy the shortcomings that evaluation identified? Did they terminate totally unsuccessful programs and expand successful ones?"

Carol Hirschon Weiss (1927 – 2013)
Professor Emerita at Harvard University

What is program evaluation ?

Why do we evaluate ?

Non-strategic arguments:

- **Effectiveness** : ensure that the program does more good than harm
- **Efficiency** : use scarce public resources to maximize the effect of the program
- **Service orientation** meet citizens' needs/expectations
- **Accountability** transparency of what is done and why
- **Democracy** enhance the democratic process
- **Trust** help ensure/restore trust in government and public services

Introduction

Causality

- Oxford dictionary: "the relationship between cause and effect"
- Causality is a theoretical concept. It cannot be (directly) measured or tested with data
- To make a causal statement, one needs a clear theory
- The methods of causal inference are “rhetorical devices”
- they allow us to establish causality **under certain assumptions**
- Since we want to identify a causal effect, these are called **identifying assumptions**

Introduction

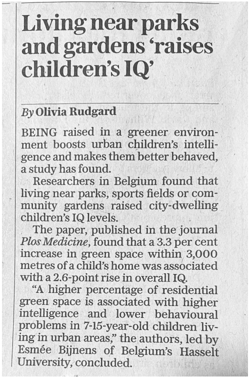
Public policy evaluation: a subfield of causal inference in economics

- Take a causal question: "*Does reducing class size improve students achievements ?*"
- Define notations to represent variables and relationships: outcomes, treatment, observable characteristics, etc.
- What are the target parameters, how are they defined ? Under which set of hypotheses can they be identified ?
- What data exists or are required to answer this question[1] ?
- Use statistics or econometrics to estimate target parameters
- Make appropriate inference to compute standard errors and test hypotheses.

1. Research questions can also come from observing patterns in data

Public policy evaluation: a subfield of causal inference in economics

When causal claims go wrong...



Living near parks and gardens 'raises children's IQ'

By **Olivia Rudgard**

BEING raised in a greener environment boosts urban children's intelligence and makes them better behaved, a study has found.

Researchers in Belgium found that living near parks, sports fields or community gardens raised city-dwelling children's IQ levels.

The paper, published in the journal *Plos Medicine*, found that a 3.3 per cent increase in green space within 3,000 metres of a child's home was associated with a 2.6-point rise in overall IQ.

"A higher percentage of residential green space is associated with higher intelligence and lower behavioural problems in 7-15-year-old children living in urban areas," the authors, led by Esmée Bijnens of Belgium's Hasselt University, concluded.

Note: Daily Telegraph August 25, 2020

Public policy evaluation: a subfield of causal inference in economics

Oh no... Did you forget about age ?

Mail Online

Home | News | U.S. | Sport | TV&Showbiz | Australia | Femail | Health | Science | Money | V

Latest Headlines | Coronavirus | Royal Family | Crime | Boris Johnson | Prince Harry | Meghan Markle | World N

Is baldness a risk factor for Covid-19? Men who have completely lost their hair are 40% more likely to end up in hospital with the virus, study claims

- **Researchers compared hair patterns of almost 2,000 British men**
- **20% of men with no hair tested positive compared to 15% of men with a full head**
- **Scientists have theorised that male hormones help the virus to enter cells**

By VANESSA CHALMERS HEALTH REPORTER FOR MAILONLINE
PUBLISHED: 14:28 BST, 22 July 2020 | **UPDATED:** 15:00 BST, 22 July 2020

Share 6.4k shares 757 View comments

Bald men are 40 per cent more likely to end up in hospital with the **coronavirus**, a study has claimed.

Dermatologists uncovered the link between hair loss and coronavirus severity by analysing data from 2,000 British men in hospital.

Results showed a fifth of men with no hair tested positive for the virus in hospital — compared to 15 per cent of those who had a full head of hair.

Note: Daily Mail July, 22 2020

Public policy evaluation: a subfield of causal inference in economics

Sure, champagne isn't correlated with unobserved factor associated with health status...

Front Nutr. 2022 Jan 3;8:772700. doi: 10.3389/fnut.2021.772700. eCollection 2021.

COVID-19 Risk Appears to Vary Across Different Alcoholic Beverages

Xi-Jian Dai [1], Liang Tan [2], Lina Ren [1], Yuan Shao [1], Weiqun Tao [1], Yongjun Wang [1]

Affiliations + expand

PMID: 35047542 PMCID: PMC8761797 DOI: 10.3389/fnut.2021.772700

Free PMC article

Abstract

Objectives: To evaluate the associations of status, amount, and frequency of alcohol consumption across different alcoholic beverages with coronavirus disease 2019 (COVID-19) risk and associated mortality. **Methods:** This study included 473,957 subjects, 16,559 of whom tested positive for COVID-19. Multivariate logistic regression analyses were used to evaluate the associations of alcohol consumption with COVID-19 risk and associated mortality. The non-linearity association between the amount of alcohol consumption and COVID-19 risk was evaluated by a generalized additive model. **Results:** Subjects who consumed alcohol double above the guidelines had a higher risk of COVID-19 (1.12 [1.00, 1.25]). Consumption of red wine above or double above the guidelines played protective effects against the COVID-19. Consumption of beer and cider increased the COVID-19 risk, regardless of the frequency and amount of alcohol intake. Low-frequency of consumption of fortified wine (1-2 glasses/week) within guidelines had a protective effect against the COVID-19. High frequency of consumption of spirits (≥5 glasses/week) within guidelines increased the COVID-19 risk, whereas the high frequency of consumption of white wine and champagne above the guidelines decreased the COVID-19 risk. The generalized additive model showed an increased risk of COVID-19 with a greater number of alcohol consumption. Alcohol drinker status, frequency, amount, and subtypes of alcoholic beverages were not associated with COVID-19 associated mortality. **Conclusions:** The COVID-19 risk appears to vary across different alcoholic beverage subtypes, frequency, and amount. Red wine, white wine, and champagne have chances to reduce the risk of COVID-19. Consumption of beer and cider and spirits and heavy drinking are not recommended during the epidemics. Public health guidance should focus on reducing the risk of COVID-19 by advocating healthy lifestyle habits and preferential policies among consumers of beer and cider and spirits.

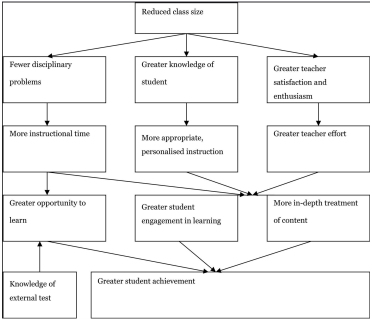
Keywords: COVID-19; SARS-CoV-2; UK Biobank; alcohol consumption; drinker; mortality; prospective cohort; risk factor.

Note: Front. Nutr., 03 January 2022 (Epidemiology)

Intuition from an example

Does reducing class size in 1^{st} grade improve student achievement ?

Figure 1: Theory on how small class size work according to Anderson (2000)



Intuition from an example

The effect or learning in a smaller class size is

- The difference between the outcome of children in small class size compared with *what would be their outcomes* if they were in normal class sizes.

Intuition from an example

The effect or learning in a smaller class size is

- The difference between the outcome of children in small class size compared with *what would be their outcomes* if they were in normal class sizes.

$\flat$ a student cannot be both in a small class size and in a large class size at the same time

Intuition from an example

The effect or learning in a smaller class size is

- The difference between the outcome of children in small class size compared with *what would be their outcomes* if they were in normal class sizes.

$\flat$ a student cannot be both in a small class size and in a large class size at the same time

$\Rightarrow$ At least one outcome is missing and other observations must be used to *impute/estimate them*

Intuition from an example

The effect or learning in a smaller class size is

- The difference between the outcome of children in small class size compared with *what would be their outcomes* if they were in normal class sizes.

$\flat$ a student cannot be both in a small class size and in a large class size at the same time

$\Rightarrow$ At least one outcome is missing and other observations must be used to *impute/estimate them*

$\Rightarrow$ We make hypotheses that are more or less plausible that allow us to use data from a comparison group to represent what would be the situation of the "small class" group of students in large classrooms.

Intuition from an example

"Nature" provides two sources of comparison

- **Before *vs* after**: We use past outcomes to represent what would be their test score in a large classroom

Intuition from an example

"Nature" provides two sources of comparison

- **Before *vs* after**: We use past outcomes to represent what would be their test score in a large classroom

♭ Problem : ability evolves with time and past test score do not represent well current academic performance

- **Treated *vs.* untreated**: We use observations of students in larger classroom to represent **counterfactual** outcomes for the small classroom students.

Intuition from an example

"Nature" provides two sources of comparison

- **Before *vs* after**: We use past outcomes to represent what would be their test score in a large classroom
- ♭ Problem : ability evolves with time and past test score do not represent well current academic performance
- **Treated *vs.* untreated**: We use observations of students in larger classroom to represent **counterfactual** outcomes for the small classroom students.
- ♭ Problem: Being in a small classroom may be associated with results in other ways (e.g. only REP+ in France)

Intuition from an example

Other factors may affect both class size and student performances

- School headmasters may select low-achieving students to be in smaller classroom in order to improve their performance
- Demographics and background origin of students may be associated with the school they go to, their test scores,

Estimating the effects requires understanding how "treatment" is allocated

"The higher the number of students, the better they perform"

Figure 2: Naïve estimates of class size effects in France by Piketty and Valdenaire (2006)

Tableau 4: L'impact de la taille des classes sur la réussite scolaire dans les écoles primaires: estimations "naïves" (OLS)

Partie A: Impact sur les scores obtenus aux évalutions de maths de CE2 (rentrée 1999)

	Tous les élèves			Elèves avec score CP inférieur à la médiane	Elèves avec score CP supérieur à la médiane
Taille de la classe de CE1	**0,169 ****	**-0,205 *****	**-0,312 *****	**-0,449 *****	**-0,183 ****
(s.e.)	(0,066)	(0,077)	(0,065)	(0,109)	(0,082)
Contrôles socio-démographiques	Non	Oui	Oui	Oui	Oui
Contrôle pour le score global CP	Non	Non	Oui	Oui	Oui
[N.obs.]	[4 718]	[3 320]	[3 300]	[1 652]	[1 648]

Partie B: Impact sur les scores obtenus aux évalutions de français de CE2 (rentrée 1999)

	Tous les élèves			Elèves avec score CP inférieur à la médiane	Elèves avec score CP supérieur à la médiane
Taille de la classe de CE1	**0,255 *****	**-0,145 ***	**-0,254 *****	**-0.447 *****	-0,103
(s.e.)	(0,068)	(0,076)	(0,065)	(0,105)	(0,084)
Contrôles socio-démographiques	Non	Oui	Oui	Oui	Oui
Contrôle pour le score global CP	Non	Non	Oui	Oui	Oui
[N.obs.]	[4 718]	[3 320]	[3 300]	[1 652]	[1 648]

Source : Calculs des auteurs à partir du panel primaire 1997 (MEN-DEP)

Lecture : Quand la taille de classe de CE1 augmente d'un élève, le score moyen obtenu aux évaluations de maths de début de CE2 augmente de 0,169 point. Mais dès lors que l'on raisonne à caractéristiques socio-démographiques observables données ("toutes choses égales par ailleurs"), le score moyen diminue de 0,205 point quand la taille de classe augmente d'un élève. Si l'on raisonne également à score obtenu en CP donné, le coefficient passe de 0,205 à 0,312. Les variables de contrôles incluent les caractéristiques des parents de l'élève (profession, niveau d'études, nationalité, âge du père et de la mère), de l'élève (sexe, mois de naissance, nombre de frères et sœurs, rang dans la fratrie) et de l'établissement (académie, tranche d'unité urbaine, école publique/privée, école en Zep, réseau d'aide spécialisée, regroupement d'adaptation ou non). Les coefficients ont été obtenus par régression linéaire MCO des scores sur la taille de classe et les variables de contrôle. Les étoiles indiquent la significativité des coefficients (***: significatif au seuil de 1% ; ** : 5% ; * : 10%).

Note: Ces régressions portent sur les élèves de CE1 scolariés en cours unique (c'est-à-dire dans une classe contenant uniquement des élèves de CE1).

"The higher the number of students, the better they perform"

How to interpret this table ?

- Simple correlation coefficient between class size in 1^{st} grade and test score in 3^{rd} grade is positive and significant: the higher the number of students, the better the results.
- controlling for students and school socio-demographics characteristics and baseline test results, the coefficient sign switches
- For student of a given initial level and socio-economic background, being in a large class size reduces their expected results.

Multivariate regressions does not mean causal estimation.

Definitions

What is a causal impact ?

- Difference between the situation when the policy is implemented and how it would have been had it not taken place.
- Not only for policy evaluation. Most causal question can be analyzed through this framework (Does having a baby reduce labour market outcomes ? Does baldness increases risks of sever COVID-19 affection ? Does eating cheese at night increase sleeping troubles ???
- **Treatment effect is always unobserved**; it's not a sample size issue.
- We need hypotheses, theoretical constructs, estimation methods and data to estimate causal effects of interest.
- The unobserved quantity is called "counterfactual"

Definitions

Three main challenges to address

1. **The identification challenge**
   - Treatment effect is never observed, so we need hypotheses to define parameter of interests (e.g. average treatment effect on the treated)
   - The **identification strategy** refers to the set of hypotheses and methods that allow to identify a parameter of interest when these hypotheses hold true.

Definitions

Three main challenges to address

1. **The identification challenge**
   - Treatment effect is never observed, so we need hypotheses to define parameter of interests (e.g. average treatment effect on the treated)
   - The **identification strategy** refers to the set of hypotheses and methods that allow to identify a parameter of interest when these hypotheses hold true.
2. **The estimation challenge**
   - Once we have an identification strategy that defines target parameters, we need to estimate them using finite samples.
   - From a data set of N observations, what statistical or econometric methods can we use to estimate the target parameters ? (are we sure it does what we think it does ?)

Definitions

Three main challenges to address

3. **The inference challenge**
    - Relationships between variables are estimated and depend on the sample used, to a certain extend.
    - Ideally, we would like our results to be as general as possible and thus, not to depend on the specific individuals we observe.
    - We need ways to characterize our precision, certainty and if (and how much) we can tell about general relationships
    - Standard errors are the most common measures, and... they are about as hard to estimate correctly as the target parameters sometimes.

Outline

Potential outcomes and causal relationships

Rubin (1974) causal model[2]

- Rubin's causal model is a very simple framework based on one main abstract construct: **potential outcomes**
- Consider a policy that either make people "treated" or "untreated" (e.g. being in a small or large class)
- Let i denote a principal sample unit (PSU) (individual, household, firm...) and let Y be an outcome of interest (e.g. test score at the end of 3rd grade)
- Let D_i be the observed variable indicating treatment status $D_i = \mathbb{1}(\text{Treated})$

2. Important complement for this section can be found in Abadie and Cattaneo (2018)

Potential outcomes and causal relationships

Rubin (1974) causal model[3]

- Rubin define potential values for this outcomes for every PSU depending on the policy implementation. Here, there are only 2 potential values (but it works with multivalued or continuous (dosage) treatment too).
  - Every individual can theoretically be treated or untreated and for a given individual i, there exist different potential values for their outcomes: $Y_i(1)$, their outcome when treated and $Y_i(0)$ when they aren't.
  - We can only observe one potential outcome for an individual at the same time. One potential outcome is **revealed** by treatment status
  - When individual i is treated their **observed** outcome is $Y_i = Y_i(1)$, when they are not we observe $Y_i = Y_i(0)$.
- From these notations, we define parameters of interests and hypotheses for identification.

3. Important complement for this section can be found in Abadie and Cattaneo (2018)

Potential outcomes and causal relationships

Rubin (1974) causal model

- Potential outcomes can be linked to observed outcome and treatment through a switching equation

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0) \tag{1}$$

- For every PSU, treatment effect is the unobservable difference

$$\delta_i = Y_i(1) - Y_i(0)$$

- This theoretical construct is defined under a core hypothesis: **SUTVA** for *Stable unit treatment value assumption*
- SUTVA means
  1. Treatment is the same for every treated units
  2. Assigning treatment to other units does not affect one's outcome (no spillover)
- Although individual effects cannot be observed, we can define some parameters (conditional expectation for instance) as a function of these quantities that can be identified and estimated using actual data under well defined identification strategies.

Rubin (1974) causal model

Some parameters of interests

- **Population average treatment effect (ATE)**:

$$\mathbb{E}[\delta_i] = \mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)]$$

- **Average treatment effect on the treated (ATT)**:

$$\begin{aligned}\mathbb{E}[\delta_i|D=1] &= \mathbb{E}[Y_i(1) - Y_i(0)|D=1] \\ &= \mathbb{E}[Y_i(1)|D=1] - \mathbb{E}[Y_i(0)|D=1]\end{aligned}$$

- **Average treatment effect on the untreated (ATU)**:

$$\begin{aligned}\mathbb{E}[\delta_i|D=0] &= \mathbb{E}[Y_i(1) - Y_i(0)|D=0] \\ &= \mathbb{E}[Y_i(1)|D=0] - \mathbb{E}[Y_i(0)|D=0]\end{aligned}$$

Rubin (1974) causal model

Average observed differences and selection bias

- We can decompose the simple average difference in octome (SDO) by treatment status to extract parameters of interests
- Under SUTVA, observed outcomes Y_i reveal potential outcomes $Y_i(\cdot)$ for the relevant units

$$\underbrace{\mathbb{E}\Big[Y_i|D_i=1\Big] - \mathbb{E}\Big[Y_i|D_i=0\Big]}_{\text{Simple difference (SDO)}} = \mathbb{E}\Big[Y_i(1)|D_i=1\Big] - \mathbb{E}\Big[Y_i(0)|D_i=0\Big]$$

- We add and substract counterfactual values for treated individuals:

$$= \mathbb{E}\Big[Y_i(1)|D_i=1\Big] - \mathbb{E}[Y_i(0)|D_i=1] + \mathbb{E}[Y_i(0)|D_i=1] - \mathbb{E}\Big[Y_i(0)|D_i=0\Big]$$

$$= \underbrace{\mathbb{E}\Big[Y_i(1) - Y_i(0)|D_i=1\Big]}_{\text{ATT}} + \underbrace{\mathbb{E}\Big[Y_i(0)|D_i=1\Big] - \mathbb{E}\Big[Y_i(0)|D_i=0\Big]}_{\text{Selection bias}}$$

Rubin (1974) causal model

Average observed difference and selection bias

- With similar manipulation one can easily show:

$$\underbrace{\mathbb{E}[Y_i(1)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0]}_{SDO} = \underbrace{\mathbb{E}[Y_i(1) - Y_i(0)]}_{ATE}$$
$$+ \underbrace{\mathbb{E}\Big[Y_i(0)|D_i = 1\Big] - \mathbb{E}\Big[Y_i(0)|D_i = 0\Big]}_{\text{Selection biais}}$$
$$+ \underbrace{(1-\pi)\big(ATT - ATU\big)}_{\text{Treatment effect heterogeneity}}$$

- Where π is the share of treated units

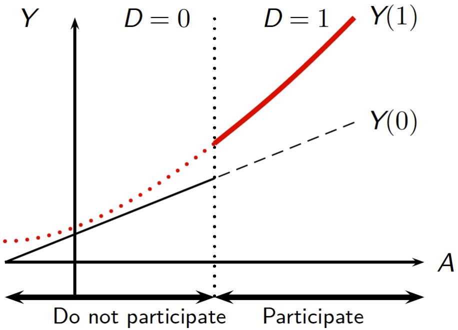
Selection and heterogeneity

Conceptual example

- Consider a policy where participants have attributes that make them more likely to participate and to benefit from the program
- Think of when you have measures of these attributes and when you don't
- Examples:
    - What's the effect of accessing Sciences Po on your career ?
    - What's the effect of childcare on mothers' labor market participation ?
    - What's the effect of organic food on risk of cardio-respiratory diseases or obesity prevalence?
- Let's plot an example and see what I mean

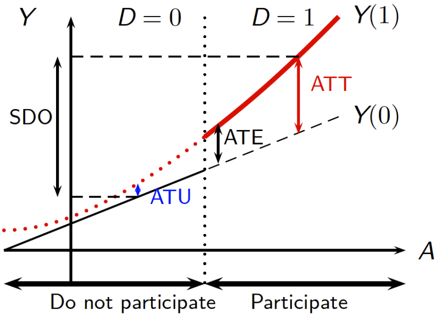
Selection and heterogeneity

Figure 3: Treatment effect with selection and heterogeneity



Selection and heterogeneity

Figure 4: Treatment effect with selection and heterogeneity



Selection and heterogeneity

Selection bias in words

In general, those who receive a treatment or benefit from a certain policy either are targeted or choose to participate. When "what" makes them more or less likely to be treated is associated with the outcomes of interest, comparing treated and untreated individuals estimate an average of treatment effect and effects of the "what" on outcome and treatment effect heterogeneity.

- You cannot test the presence or absence of selection bias when you think there are some unobservables
- Controlling for observables may worsen bias in some cases (see Cunningham (2018) chapter on "front doors" and "back doors" criterion).

Selection and heterogeneity

Selection bias in words

- In practice, how do we deal we selection bias ?
  1. When you "control" selection
     - Sometimes (not often), it's reasonable to think that when you observe certain characteristics there's no others that affect both treatment status and outcomes
     - There are statistical methods that allow to control for theses characteristics and retrieve causal estimates.
     - ♭ Which variables, how they are measured, and what type of relationship is modelled matter.
  2. When selection is based on "unobservable" characteristics
     - More often, selection arise from latent characteristics (e.g. motivation) that's usually not measured.
     - Without alternative identification strategy, making causal statement in such settings is bold.

Selection and heterogeneity

When there are spillovers

- Rubin's causal model is based on SUTVA which means no externalities, no spillover.
- In practice, externalities can come from four channels
  1. Through direct negative effect on untreated units (e.g. labor queue in intensive training for the unemployed)
  2. Through market adjustment and general equilibrium effects (e.g. increased share of graduates may reduce the wage premium associated with diploma)
  3. Behavioural adjustments of participants, non participants (moral hazard, discouragement, sharing between groups,...)
  4. Other changes from other agents (other actors endogenously react to the existence of treatment: eg provision of micro-loan may change behaviour of informal lenders)
- Sometimes you can account for and measure externalities with a clever design (e.g. testing labour queue effect as in (Crépon et al. 2013))

Selection and heterogeneity

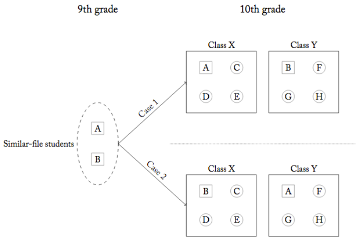
Internal validity *vs.* External validity

- Identification strategies focus on the internal validity of the estimator i.e. minimizing bias for the estimatand to retrieve target parameters.
- For the sub-population of interest, or even this particular sample, we want to be confident in the causal interpretation of our results
- *Research designs* with good internal validity may not provide good estimates of what would be the effect at a larger scale, in another context, for another population.
- External validity question our ability to interprete empirical estimates as more general evidence
- Large scale policy usually don't yield estimates as large as small-scale experiments (see Duncan and Magnuson (2013) for a discussion on early childhood policies)
- External validity is usually a "case-by-case" discussion although there's a lot of work on replicating and scaling-up good ideas. (Read the book John A. List. 2022. *The Voltage Effect: How to Make Good Ideas Great and Great Ideas Scale.* Currency. Chicago, February 1, 2022)

Internal validity *vs.* External validity

Illustration from Riegert (2016, chapter 2)

- **Research question**: Does being with a persistent classmate from 8th grade in 9th grade increase student achievement
- **Identification strategy :** Comparing "similar in files" students assigned to classroom with no or more or less former classmates.
- Starting from a database with 3 500 000 students, they identify 28 000 similar in files students among which only 8 900 are "at risk"
- Very convincing identification strategy but really hard to tell whether their results are valid for a broader population



Causal evaluation of public policies: how to ?

Solution I : Experiments

- Randomized control trials (RCT) allocate treatment randomly across a sample so assignment is independent from potential outcomes and therefore the simple difference in outcomes identify the average treatment effect.
- Randomization $\Rightarrow D_i \perp Y_i(0), Y_i(1)$

$$\Rightarrow \mathbb{E}\Big[Y_i^0|D_i = 1\Big] - \mathbb{E}\Big[Y_i(0)|D_i = 0\Big] = 0$$

- Theoretically, a well implemented RCT needs no other assumption but SUTVA to identify the ATE or other interesting constructs.
- In practice, there are many considerations that make them more complicated.

Causal evaluation of public policies: how to ?

Solution II : Quasi-experiments and other identification strategies

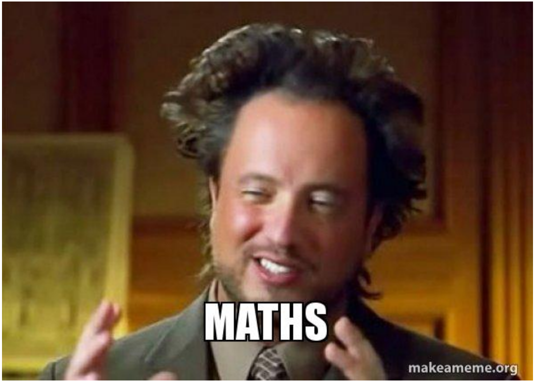
- Most non-experimental identification strategies rely on one or more of these assumptions
  1. Treatment is independent from potential outcomes **conditional on a set of observed characteristics X**. ⇒ Conditional independence assumption
  2. Treatment is independent from changes in time → Difference in difference.
  3. One other variable predicts treatment participation but is not related to outcomes but through treatment participation. ⇒ Instrumental variables

Solution III : Set identification

- Use combination of restrictions on data and hypotheses to identify a set of plausible values for the treatment effect (Not this year)

Causal evaluation of public policies: how to ?

Solution IV : Structural models



Outline

A short story of empirical economics

How econometrics was used

- Read the excellent lecture by Card (2012) available **HERE**
- pre-1970
  - Most econometric modelling focused on macro data, using systems of linear equations
  - Labour economics: Mincer equations for wage and education, role of unions on wages,...
- 1970s:
  - Micro data and computer power
  - McFadden's logit model
  - Heckman's selection model
- 1980's - 2000
  - Many critics of econometric models
  - Emergence of "design-based" approach especially with Lalonde, Angrist, Card, Kruger, Lavy, Imbens,...
  - Identification equated with research design
  - Research design defines the counterfactual
  - Randomization is 'gold standard' for asserting causality

A short story of empirical economics

Where are we now (inspired from Card (2012))

- Design-based papers use simplified working models (e.g., single equation causal model) with no formal derivation of the DGP for the observed data. Emphasis is on "sharp" predictions and falsification of theories.
- Structural papers follow McFadden program: derive DGP from assumed model (though estimation may only use moment conditions). Emphasis on sophisticated agents, equilibrium behaviour.
- Design-based approach estimates $\mathbb{E}[y|X] = \boldsymbol{X'}\beta$ with focus on "clean identification" of β
- Model-based approach poses model $g(y, x, \varepsilon; \theta) = 0$ with focus on estimating θ, sometimes using the implied linear projection $L(y|X) = \boldsymbol{X'}\beta(\theta)$
- design-based practitioners emphasize "credible" identification, with testing/evaluation of assumptions
- model-based studies are often identified by combination of stochastic assumptions, functional form and exclusion restrictions, and "outside information" (e.g. interest or discount rates are calibrated).

The credibility revolution

Design based approach consecrated

- Rubin's potential outcomes rely on minimal assumptions that derive directly from the randomisation framework and (allegedly) careful use of statistical modelling and econometrics.
- **Design-based** approach to assess causal questions.
- Opened the path through what is now know as the "credibility revolution" (Angrist and Pischke 2010) consecrated with two recent Nobel prizes
  1. 2019: Abhijit Banerjee, Esther Duflo, and Michael Kremer for the experimental approach in economics and
  2. 2021: Joshua Angrist, David Card and Guido Imbens for their work on causality with observational data
- Main critics in the field from two other Nobel prize winners: Angus Deaton (e.g. (Deaton 2020)) and James J. Heckman (e.g. (Heckman and Pinto 2022))

The credibility revolution

What are the main critics ?

Angus Deaton is the 2015 Nobel Prize winner for his work on consumer choices and happiness. His critics against design-based economics focus mostly on these arguments:

- Implementation to restricted sets of questions, overly focused on short-term, measurable outcomes at the expense of long-term and broader impacts ("under-the-spotlight critics").
- Overly prescriptive and fail to account for the complexity of real-world economic systems (Replicability, general-equilibrium effects).
- Ethical concerns with experiments
- Lack of replicability (across sites but also design issues such as power)

The credibility revolution

What are the main critics ?

James J. Heckman is the 2000 Nobel Prize winner for his work on econometrics theory and methods to correct for selection effects. Heckman is rather severe although he uses randomised experiment in his work. (He is very critical in general though...) Heckman's critics are more about the interpretability and limits of what you can learn from design-based analysis, especially with instrumental variables

- Design-based analysis estimate participation equations and get a "how much ?" pretty-much "assumption-free".
- (Heckman 2005) proposes the "scientific model of causality" that emphasizes the use of multiple methods and sources of evidence to infer causality, rather than relying solely on experimental design.

The credibility revolution

Toward reconciliation ?

- Academic debate: better LATE than nothing ? (Imbens 2010)
- More and more convergence in the two approaches, even from Heckman himself (Heckman and Pinto 2022).
- Model-based analysis get the same "how much" with structure on the "why" (Kline and Walters 2019)
- Ability to extrapolate more general parameters from experimental manipulations (Mogstad and Torgovitsky 2018)
- Design-based estimations can be used to calibrate structural models (e.g. (Card and Hyslop 2005))

Outline

Wrap-up

From Rubin's causal model to public policy evaluation

- Rubin's causal model is a powerful yet simple framework to think of causality
- Emphasize hypothesis about assignment, participation,...
- Lead to design-based economics and the credibility revolution

Next week: Randomised control trials

- **To read: mandatory**:Esther Duflo, Rachel Glennerster, and Michael Kremer. 2008. "Using Randomization in Development Economics Research: A Toolkit." In *Handbook of Development Economics,* 4:3895–3962. Elsevier
- **To watch**: YouTube video by J.D. Angrist on how to read economics papers on RCT https://youtu.be/s-_3s3OMeqs

Bibliography I

- Abadie, Alberto, and Matias D. Cattaneo. 2018. "Econometric Methods for Program Evaluation." *Annual Review of Economics* 10, no. 1 (August 2, 2018): 465–503.
- Anderson, L.W. 2000. "Why Should Reduced Class Size Lead to Increased Student Achievement?" In *How Small Classes Help Children Do Their Best,* Temple University Center for Research in Human Development and Education., edited by M.C WANG and J.D. FINN, 3–24. Philadelphia.
- Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.
- Angrist, Joshua D, and Jörn-Steffen Pischke. 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics." *Journal of Economic Perspectives* 24, no. 2 (May): 3–30.
- ———. 2015. *Mastering 'Metrics: The Path from Cause to Effect.*
- Athey, S., and G. W. Imbens. 2017a. "Chapter 3 - The Econometrics of Randomized Experiments." In *Handbook of Economic Field Experiments,* edited by Abhijit Vinayak Banerjee and Esther Duflo, 1:73–140. Handbook of Field Experiments. North-Holland, January 1, 2017.
- Athey, Susan, and Guido W. Imbens. 2017b. "The State of Applied Econometrics: Causality and Policy Evaluation." *Journal of Economic Perspectives* 31, no. 2 (May): 3–32.
- Cameron, A Colin, and Pravin K Trivedi. 2005. *Microeconometrics : Methods and Applications.* Cambridge University Press.

Bibliography II

- Card, David. 2012. "Model-Based or Design-Based? Competing Approaches in "Empirical Micro"." Woytinsky lecture.
- Card, David, and Dean R. Hyslop. 2005. "Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare-Leavers." *Econometrica* 73 (6): 1723–1770.
- Colonescu, Constantin. 2016. *Principles of Econometrics with R.* September 1, 2016.
- Crépon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot, and Philippe Zamora. 2013. "Do Labor Market Policies Have Displacement Effects? Evidence from a Clustered Randomized Experiment." *The Quarterly Journal of Economics* 128 (2): 531–580.
- Cunningham, Scott. 2018. *Causal Inference: The Mixtape.*
- Deaton, Angus. 2020. *Randomization in the Tropics Revisited: A Theme and Eleven Variations.* w27600. Cambridge, MA: National Bureau of Economic Research, July.
- Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2008. "Using Randomization in Development Economics Research: A Toolkit." In *Handbook of Development Economics,* 4:3895–3962. Elsevier.
- Duncan, Greg J, and Katherine Magnuson. 2013. "Investing in Preschool Programs." *Journal of Economic Perspectives* 27, no. 2 (February): 109–132.
- Frisch, Ragnar. 1930. "A Dynamic Approach to Economic Theory," 246. Lectures by Ragnar Fritch. Yale University.
- Heckman, James J. 2005. "The Scientific Model of Causality." *Sociological Methodology* 35:1–97.

Bibliography III

- Heckman, James J., and Rodrigo Pinto. 2022. "The Econometric Model for Causal Policy Analysis." *Annual Review of Economics* 14, no. 1 (August 12, 2022): 893–923.
- Huntington-Klein, Nick. 2021. *The Effect: An Introduction to Research Design and Causality | The Effect.* December 21, 2021.
- Imbens, Guido W. 2010. "Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)." *Journal of Economic Literature* 48, no. 2 (June): 399–423.
- Imbens, Guido W., and Donald B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* Cambridge: Cambridge University Press.
- Kleven, Henrik, Camille Landais, and Jakob Egholt Sø gaard. 2021. "Does Biology Drive Child Penalties? Evidence from Biological and Adoptive Families." *American Economic Review: Insights* 3, no. 2 (June): 183–198.
- Kline, Patrick, and Christopher R. Walters. 2019. "On Heckits, LATE, and Numerical Equivalence." *Econometrica* 87 (2): 677–696.
- List, John A. 2022. *The Voltage Effect: How to Make Good Ideas Great and Great Ideas Scale.* Currency. Chicago, February 1, 2022.
- Mogstad, Magne, and Alexander Torgovitsky. 2018. "Identification and Extrapolation of Causal Effects with Instrumental Variables." *Annual Review of Economics* 10, no. 1 (August 2, 2018): 577–613.
- Pearl, Judea, and Dana Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect.* New York: Basic Books. ISBN: 978-0-465-09761-6.

Bibliography IV

- Piketty, Thomas, and Mathieu Valdenaire. 2006. *L'impact de la taille des classes sur la réussite scolaire dans les écoles, collèges et lycées français: estimations à partir du panel primaire 1997 et du panel secondaire 1995,* Les Dossiers enseignement scolaire 173. Paris: Ministère de l'éducation nationale, de l'enseignement supérieur et de la recherche, Direction de l'évaluation et de la prospective.
- Ravallion, Martin. 2007. "Chapter 59 Evaluating Anti-Poverty Programs." In *Handbook of Development Economics,* 4:3787–3846. Elsevier. ISBN: 978-0-444-53100-1.
- Riegert, Arnaud. 2016. "Inégalités Scolaires, Ségrégation et Effets de Pairs," EHESS.
- Robinson, Thomas S. 2020. *10 Fundamental Theorems for Econometrics.* Vol. v0.1. September 30, 2020.
- Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66(5):688–701.
- Singh Rawat, Vikram. 2021. *Best Coding Practices for R.* April 27, 2021.
- Wickham, Hadley, and Garrett Grolemund. 2017. "Welcome | R for Data Science," January.
- Wooldridge, Jeffrey M. 2012. "Introductory Econometrics: A Modern Approach," 910.