Part A: Regression and causality

A2: Potential outcomes and RCTs

Kirill Borusyak

ARE 213 Applied Econometrics

UC Berkeley, Fall 2023

Outline

1. The concept of potential outcomes

2. Identification of causal parameters

3. Limitations and alternatives

Rubin causal model

- Consider some population of units i
- Suppose each unit is observed one of several treatment conditions $D_i \in \mathfrak{D}$
  - E.g. $\mathfrak{D} \in \{0, 1\}$: control and treated
- Suppose we can imagine each unit under all possible conditions
  - Causality always requires specifying alternatives
  - Corresponding **potential outcomes** are $\{Y_i(d) : d \in \mathfrak{D}\}$
    - e.g. $(Y_i(0), Y_i(1))$ (equivalently written as (Y_{0i}, Y_{1i}))
    - e.g. demand function
  - **Causal effects** $Y_i(d') - Y_i(d)$ are defined by this abstraction
  - Writing $Y_i(d)$ encodes a possibility that D_i impacts Y_i
  - **Realized outcome**: $Y_i = Y_i(D_i)$

What can be a cause?

Imagining each unit under all possible conditions is non-trivial:

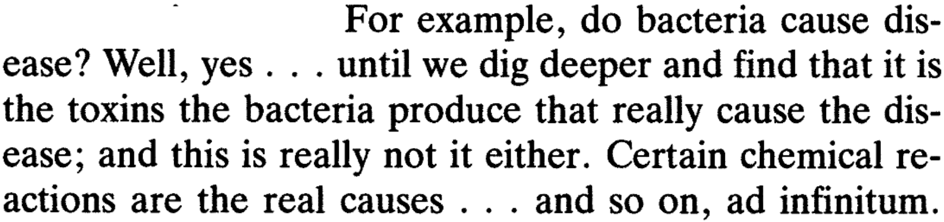
"No causation without [imagining] *manipulation"* (Holland & Rubin)

1. *"She did well on the exam because she was coached by her teacher"* (Holland 1986) ✓
2. *"She did not get this position because she is a woman"* (Imbens 2020) ✗
   - Gender is an **attribute**, not a cause; same for race
   - *"She got an orchestra job because of a gender-blind audition"* (cf. Goldin and Rouse 2000) ✓
3. *"Household's housing expenditures affect savings?"* (Wooldridge) ✗
   - Housing and savings are a joint household decision affected by rent etc.

Effects of causes vs. causes of effects

Statistical analysis focuses on effects of causes (treatments) rather than causes of effects (outcomes)

- Causes are not clearly defined



> For example, do bacteria cause disease? Well, yes . . . until we dig deeper and find that it is the toxins the bacteria produce that really cause the disease; and this is really not it either. Certain chemical reactions are the real causes . . . and so on, ad infinitum.

(Holland 1986, p.959)

- What caused increased obesity in the US?
    - Numerous factors without clear counterfactuals: e.g. rising consumption of snacks

SUTVA

In writing $Y_i(d_i)$ we implicitly imposed **SUTVA** ("stable unit value treatment assumption")

- Most common meaning: no unmodeled **interference**
  - I.e., treatment statuses of other units, d_{-i} do not affect Y_i (individually or together)
  - Frequently violated: e.g. vaccines and infectious disease; information and technology adoption
  - With interference we'd write $Y_i(d_1, \dots, d_N)$ for the population of size N
  - We may be interested in own-treatment effects $Y_i(d'_i, d_{-i}) - Y_i(d_i, d_{-i})$ and various spillover effects, e.g. $Y_i(d_i, 1, \dots, 1) - Y_i(d_i, 0, \dots, 0)$
- Original meaning: d_i summarizes everything about the intervention that is relevant for the outcome

Outline

1. The concept of potential outcomes

2. Identification of causal parameters

3. Limitations and alternatives

Common causal parameters

- We cannot learn the causal effect $Y_i(1) - Y_i(0)$ for any particular unit
    - "Fundamental problem of causal inference": multiple potential outcomes are never observed at once
    - ... but we can sometimes learn some averages
- Average treatment/causal effect: $ATE = \mathbb{E}\left[Y_i(1) - Y_i(0)\right]$
    - $ATE = \mathbb{E}\left[Y_i(1)\right] - \mathbb{E}\left[Y_i(0)\right]$
    - $Y_i(1) - Y_i(0)$ is never observed but $Y_i(1)$ and $Y_i(0)$ are: for some but not all units
- Average effect on the treated: $ATT = \mathbb{E}\left[Y_i(1) - Y_i(0) \mid D_i = 1\right]$ (a.k.a. TOT, TT)
- Average effect on the untreated: $ATU = \mathbb{E}\left[Y_i(1) - Y_i(0) \mid D_i = 0\right]$
- These are parameters of the distribution of $(Y(1), Y(0), D)$. Are they identified from observations of (Y, D)?

Identifying ATT & ATE

$$\begin{aligned} ATT &= \mathbb{E}\left[Y_1 \mid D=1\right] - \mathbb{E}\left[Y_0 \mid D=1\right] \\ &= \left(\mathbb{E}\left[Y_1 \mid D=1\right] - \mathbb{E}\left[Y_0 \mid D=0\right]\right) && \text{(Difference in means)} \\ &\quad - \left(\mathbb{E}\left[Y_0 \mid D=1\right] - \mathbb{E}\left[Y_0 \mid D=0\right]\right) && \text{(Selection bias)} \end{aligned}$$

- Thus, $\beta_{\text{OLS}} = \mathbb{E}\left[Y \mid D=1\right] - \mathbb{E}\left[Y \mid D=0\right] = ATT + \text{Selection bias}$
  - Selection bias $= 0$ iff $Y_0 \perp D$
- $ATE = ATT$ iff $(Y_1 - Y_0) \perp D$
  - Simple regression identifies ATE and ATT in a randomized control trial (**RCT**) where $(Y_0, Y_1) \perp D$ by **design**
  - Regression with any (fixed set of) predetermined controls X also identifies ATE by FWL or OVB logic

Distribution of gains?

Can we identify other parameters of the distribution of $Y_1 - Y_0$, e.g. median gains?

- Generally no!
- Imagine an RCT where in both treated and control groups Y takes values 0, 1, 2 with prob. $1/3$ each
- This is consistent with no casual effect for anyone
  - Median gain $= 0$
- Or with (Y_0, Y_1) only taking values $(0,1)\,,(1,2)\,,(2,0)$ with prob. $1/3$ each
  - Median gain $= 1$

Connecting to linear models

- With a binary treatment, the potential outcomes model implies

$$Y_i = Y_{0i}(1 - D_i) + Y_{1i}D_i = \beta_0 + \beta_{1i}D_i + \varepsilon_i$$

  where $\beta_0 = \mathbb{E}\left[Y_0\right]$, $\beta_{1i} = Y_{1i} - Y_{0i}$ and $\varepsilon_i = Y_{0i} - \mathbb{E}\left[Y_0\right]$

- With homogeneous effects, $Y_i = \beta_0 + \beta_1 D_i + \varepsilon_i$ becomes a causal *model* where $Y_{1i} - Y_{0i} \equiv \beta_1$ (regardless of whether $\varepsilon_i \perp D_i$; think IV)

- With heterogeneous effects, can rederive our result about RCTs: if $(\varepsilon_i, \beta_{1i}) \perp D_i$ and denoting $\mu = \mathbb{E}\left[D_i\right]$,

$$\beta_{OLS} = \frac{\mathbb{E}\left[(D_i - \mu)\, Y_i\right]}{\mathrm{Var}\left[D_i\right]} = \frac{\mathrm{Cov}\left[D_i, \varepsilon_i\right]}{\mathrm{Var}\left[D_i\right]} + \frac{\mathbb{E}\left[D_i\left(D_i - \mu_i\right)\beta_{1i}\right]}{\mathrm{Var}\left[D_i\right]} = \mathbb{E}\left[\beta_i\right] \equiv ATE$$

Ordered and continuous treatments

Consider a RCT where D takes more than two values (e.g. different dosages)

- $D \perp \{Y(d)\}_{d \in \mathfrak{D}} \implies \mathbb{E}[Y \mid D = d] = \mathbb{E}[Y(d) \mid D = d] = \mathbb{E}[Y(d)]$
- A saturated regression of Y on dummies for all values of D (or a nonparametric regression with continuous D) traces the average structural function $\mathbb{E}[Y(d)]$
- A simple regression of Y on D identifies a convexly-weighted average of $\partial \mathbb{E}[Y(d)] / \partial d$ (or its discrete version):

$$\beta_{OLS} = \int_{-\infty}^{\infty} \omega(\tilde{d}) \frac{\partial \mathbb{E}\left[Y(\tilde{d})\right]}{\partial \tilde{d}} d\tilde{d}, \qquad \omega(\tilde{d}) = \frac{\mathrm{Cov}\left[\mathbf{1}\left[D \geq \tilde{d}\right], D\right]}{\mathrm{Var}[D]}$$

$$\text{or } \beta_{OLS} = \sum_{k=1}^{K} \omega_k \frac{\mathbb{E}[Y(d_k) - Y(d_{k-1})]}{d_k - d_{k-1}}, \qquad \omega_k = \frac{(d_k - d_{k-1})\,\mathrm{Cov}\left[\mathbf{1}[D \geq d_k], D\right]}{\mathrm{Var}[D]}$$

Outline

1. The concept of potential outcomes

2. Identification of causal parameters

3. Limitations and alternatives

Limitations and criticisms (see Heckman and Vytlacil 2007)

1. Estimated effects cannot be transferred to new environments (limited external validity) and new programs never previously implemented
    - Interventions are black boxes, with little attempt to unbundle their components
    - Mechanisms are not possible to pin down
    - Knowledge does not cumulate across studies (compare with estimates of a labor supply elasticity!)

Limitations and criticisms (2)

2. Estimands need not be relevant even to analyze the observed policy

    - Informative on whether to throw out the program entirely (ATT) and whether to extend it forcing it on everyone (ATU)
    - But not whether to extend/shrink it on the margin
    - Or a policy change that affects the assignment mechanism, e.g. available options
    - No analysis from the social planner's point of view, e.g. accounting for externalities
    - No analysis of causal parameters other than means, e.g. median gains

Exercise: which criticisms in Heckman-Vytlacil's Section 4.4 do you agree with?

Roy model

Alternative "structural" approach: to model self-selection explicitly

- Original Roy (1951) model: self-selection based on outcome comparison
  - $D =$ choice of occupation (e.g. agriculture vs not) or education level
  - $Y(d) =$ earnings for a given occupation/education
  - People vary by occupational productivities/returns to education, known to them
  - They choose based on them: $D = \arg\max_{d \in \mathfrak{D}} Y_i(d)$, perhaps with homogeneous costs
- Extended Roy model: costs are heterogeneous but fully determined by observables
  - which may or may not affect the outcome at the same time
- Generalized Roy model: self-selection based on unobserved preferences
  - $D = \arg\max_{d \in \mathfrak{D}} R_i(d)$ where e.g. $R_i(d) = Y_i(d) - C_i(d)$ for costs $C_i(d)$

Roy model: Identification

What does this structure buy us?

- No free lunch: *"for general skill distributions, the model is not identified* [from a single cross-section] *and has no empirical content"* (Heckman and Honore 1990)
- But with more data and restrictions can identify the ATE and even the distribution of $(Y_0, Y_1, R_1 - R_0) \implies$ distribution of gains
- Identification conditions depend on:
  - original vs. extended vs. generalized Roy
  - whether Y is observed for all choices (e.g. payoffs unobserved for the unemployed)
  - availability of instruments in the selection and outcome equations
- Often parametric: we'll see "Heckit" later
  - Not living up to the goal of using economic theory for identification?
- Can do better with cost shifters that shift selection but not outcomes
  - Value over traditional IV methods is not so clear?

Application: Adão 2016

“Worker Heterogeneity, Wage Inequality, and International Trade: Theory and Evidence from Brazil”

- How do commodity price shocks affect wages and wage inequality?
- Shocks affect wages in the commodity sector but also induce sectoral reallocation
- Adão imposes a simple Roy model and focuses on the role of workers’ comparative and absolute advantage
- Workers i draw $\left(L_i^C, L_i^N\right)$: efficiency units in Commodity/Non-commodity sectors
- $s_i = \log\left(L_i^C / L_i^N\right)$ denotes i’s comparative advantage in C
  - $q_i = F_s(s_i)$ sorts workers by comparative advantage in C
- $a_i = \log\left(L_i^N\right)$ denotes i’s absolute advantage in N
- Let ω^N and ω^C be market wages per efficiency unit (affected by price shocks)
- Potential log-wages are $Y_i^N = \omega^N + a_i$ and $Y_i^C = \omega^C + a_i + s_i$

Adão 2016: Selection

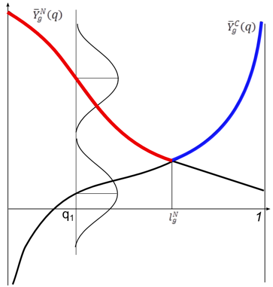
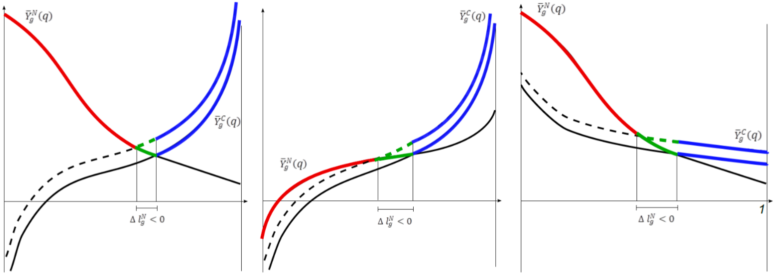


Figure 2: Sectoral Log-Wages and Employment in Equilibrium

- Selection: workers choose C iff $Y_i^C > Y_i^N \iff s_i > \omega^N - \omega^C$
- Distribution of log-wages in N is determined by the distribution of $a_i \mid s_i$
- Distribution of log-wages in C is determined by the distribution of $a_i + s_i \mid s_i$
- Slopes of $\mathbb{E}\left[a_i \mid s\right]$ and $\mathbb{E}\left[a_i + s_i \mid s_i\right]$ determine positive/negative selection

Adão 2016: Responses to a price shock



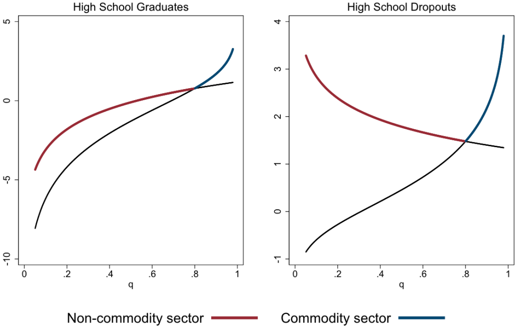
(a)
Positive selection into
both sectors

(b)
Negative selection into the
non-commodity sector

(c)
Negative selection into the
commodity sector

- Slope of F_s determines the employment response
- The gap between average and marginal productivity in each sector determines the average wage response
- Learning these responses from exogenous price shocks (e.g. across regions) yields both comparative and absolute advantage schedules

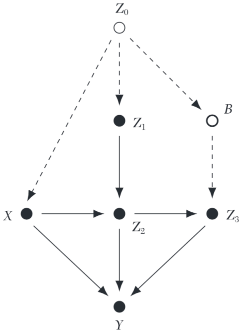
Adão 2016: Estimates



- Different selection patterns for high-school graduates and high-school dropouts
- Imply different inequality implications in response to price shocks

Another alternative: DAGs

Directed acyclic graphs of Judea Pearl represent causal relationships graphically: e.g.



$X =$ soil treatment (fumigation)
$Y =$ crop yield
$Z_1 =$ eelworm population before the treatment
$Z_2 =$ eelworm population after the treatment
$Z_3 =$ eelworm population at the end of season

$Z_0 =$ eelworm population last season (unobserved)
$B =$ bird population (unobserved)

- "Do-calculus" allows to verify whether the ATE of X on Y is identified from observing (X, Y, Z_1, Z_2, Z_3)
- Popular in epidemiology but not in economics. Why?

Some limitations of DAGs

Imbens (JEL 2020) lists some pitfalls of DAGs relative to potential outcomes:

1. Economists avoid complex models with many variables
2. Randomization and manipulability have no special value in DAGs
3. Not possible to incorporate additional assumptions, such as monotonicity (important in IV) and continuity (important in RDD)
4. Modeling simultaneity, e.g. demand and supply, is clunky
5. Interference is difficult to model
6. Too much focus on identification, relative to estimation and inference