Causal Inference: Estimation methods

Álvaro A. Novo

ISEG - UTL

Lisbon, 2016

Outline

1 Econometric methodology

1.1 Fundamental identification problem

- How can we define causility? Let

$$D_i = \begin{cases} 1 & \text{if unit } i \text{ received treatment} \\ 0 & \text{otherwise} \end{cases}$$

$$Y_i : \text{ Outcome variable for unit } i$$

To define causality, use notion of **potential** outcomes:

Y_{0i} : Potential outcome **without treatment** for unit i

Y_{1i} : Potential outcome **with treatment** for unit i

Potential outcomes refer to possible **counterfactual** events

- Natural definition of **causal effect** of the treatment for unit i:

$$Y_{1i} - Y_{0i}$$

- **Fundamental problem of identification:** Cannot observe both Y_{0i} and Y_{1i} for the same individual

- Only observe the **realized** outcome Y_i:

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

- Homogeneity would solve this problem; but heterogeneity rules.

- Although we cannot compute individual treatment effects, we would like to compute average effects:
  - **Average treatment effect (ATE)**

$$\alpha_{\mathbf{ATE}} = E[Y_1 - Y_0]$$

  - **Average treatment effect on the treated (ATET)**

$$\alpha_{\mathbf{ATET}} = E[Y_1 - Y_0 | D = 1]$$

- However, comparisons between treated and untreated do not usually give the rigth answer:

$$\begin{aligned} E[Y|D=1] - E[Y|D=0] &= E[Y_1|D=1] - E[Y_0|D=0] \\ &= \underbrace{E[Y_1 - Y_0|D=1]}_{\text{ATET}} + \underbrace{\{E[Y_0|D=1] - E[Y_0|D=0]\}}_{\text{Bias}} \end{aligned}$$

  - Bias term is not likely zero: **selection** problems. Participation is usually associated with the potential outcome

- Causal inference appears a difficult task

  - Way around: Study the **assignment mechanism** (and restrict it appropriately)

1.2 Randomized experiments

- Suppose **treatment is independent of potential outcomes**:

$$(Y_1, Y_0) \perp D \tag{1}$$

- Treatment is *ignorable*. By independence of treatment $E[Y_0|D=1] = E[Y_0|D=0]$, and therefore:

$$\alpha_{ATET} \equiv E[Y_1 - Y_0|D=1] = E[Y|D=1] - E[Y|D=0]$$

Also because $E[Y_1|D=1] = E[Y_1|D=0]$:

$$\alpha_{ATE} \equiv E[Y_1 - Y_0] = E[Y_1 - Y_0|D=1] = E[Y|D=1] - E[Y|D=0]$$

- Randomized study forces $(Y_1, Y_0) \perp D$ (eq. (1)) to hold
  - Also, predetermined **observed** characteristics (e.g. age, sex, education) should be balanced between treated and controls. **Unobserved** characteristics (e.g. ability) are also balanced.
  - In practice, estimation is carried out by using sample analogs of the population ($\alpha_{ATE} = \alpha_{ATET}$):

$$\hat{\alpha} = \bar{Y}_1 - \bar{Y}_0$$

Randomized experiments: Validity and threats

1. **Internal validity**: Able to estimate the effect for our sample. **Threats** to internal validity:
    a) Failure of randomization
    b) Non-compliance with experimental protocol
    c) Attrition

2. **External validity**: Can extrapolate estimates to other populations.
   Threats to external validity:

   a) Non-representative sample

   b) Non-representative program (treatment differs in actual implementation; scale effects; actual implementations are not randomized)

1.3 Observational studies

Identification: Selection on observables

- Absence of experimental data, $(Y_1, Y_0) \perp D$ is rarely plausible
- **Selection on observables**: There is selection on observables when there exists observed predetermined variables X, s.t. the treatment is independent of the potential outcomes conditional on the covariates:

$$(Y_1, Y_0) \perp D|X \tag{2}$$

  - Statistical jargon: covariates are called confounding variables
  - **Predetermined variable**: Variable X is predetermined w.r.t. treatment if for each i, $X_{0i} = X_{1i}$. I.e., X_i does not depend on the value of D_i.

- **Selection on observables** justifies the use of estimation methods to control for differences in predetermined characteristics:
  - **Matching**
  - **Regression**
  - **Difference-in-Differences**

1.3.1 Matching estimators: Matching on covariates

- X discrete and takes on a small number of values: $\{X^1, X^2, \ldots, X^J\}$
  - n^j: number of observations in cell j
  - n_1^j: number of treated in cell j
  - n_0^j: number of controls in cell j
  - $\bar{Y}$: average value of Y
- Then

$$\hat{\alpha}_{\mathbf{ATET}} = \sum_{j=1}^{J} (\bar{Y}_1^j - \bar{Y}_0^j) \times \frac{n_1^j}{n_1}$$

- When the dimension of X is large, this strategy breaks down
  - Empty cells; Hard to find similar covariate values.

Matching estimators: Propensity score methods

- **Propensity score**: Under selection on observables, we call propensity score to the selection probability conditional on the confounding variables, $P(D = 1|X)$.
- **Conditioning on the propensity score is enough** to have independence between treatment and potential outcomes

$$\text{If } (Y_1, Y_0) \perp D|X \quad \text{then} \quad (Y_1, Y_0) \perp D|P(D = 1|X).$$

- **Two step estimation** procedure:
  1. Estimate the propensity score, $P(D = 1|X)$
  2. Do matching on the propensity score

1. $P(D = 1|X)$ can be estimated with a **probit** or **logit** model.
2. Matching (selecting controls based) on the propensity score:
   - **Nearest neighbor**: Control with $P(X)$ closest to i's $P(X)$
   - **Caliper**: Control(s) with $P(X)$ within a range of i's $P(X)$
   - **Kernel**: Weight each control unit by the kernel (density); control units closer to i's $P(X)$ have larger weights

- **Matching average treatment effect on the treated**:

$$\alpha_{ATET} = \frac{1}{n_1} \sum_{i=1}^{n_1} \{Y_{1i} - \sum_{j=1}^{n_0} \omega_{(i,j)} Y_{0j}\}$$

where $\omega_{(i,j)}$ is the weight of j as a control for unit i

1.3.2 Regression

How to interpret linear regression in this context?

- Under $Y_1, Y_0 \perp D|X$, $E[Y|D,X]$ is a conditional causal response:

$$E[Y|D=1,X] = E[Y_1|D=1,X] = E[Y_1|X]$$
$$E[Y|D=0,X] = E[Y_0|D=0,X] = E[Y_0|D=1,X] = E[Y_0|X]$$

  So $E[Y|D,X]$ provides the average potential responses:

$$E[Y|D=1,X] - E[Y|D=0,X] = E[Y_1 - Y_0|X]$$

- Functional form of $E[Y|D,X]$ is typically unknown. If we assume linear, then OLS estimate of α gives impact of treatment on treated:

$$Y = \alpha D + X\beta$$

1.3.3 Difference-in-Differences

- Previous methods: control for observed differences. However, treated and non-treated may **differ in unobservables**
- With pre- and post-treatment data available:
  - Compare outcomes of treated **after** the treatment with outcomes **before** the treatment
  - Before-after comparisons are likely to be contaminated: **temporal trends** and effects of **other events** on outcome
- **Use untreated comparison group** to identify the temporal variation in the outcome not due to treatment exposure.
  - **Difference-in-Differences (DinD)** is based on this idea

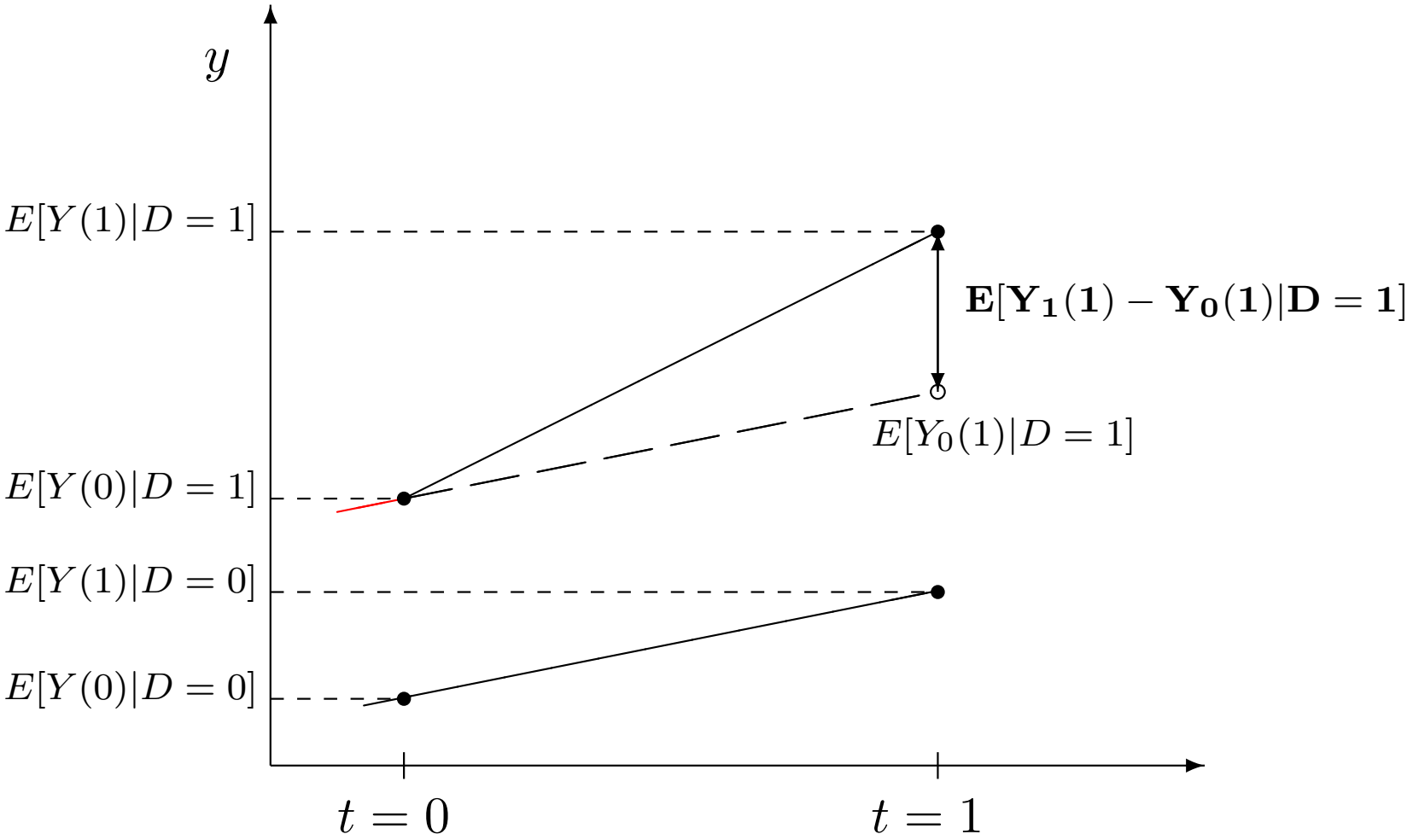
- **Assumption: Common trend**

$$E[Y_0(t=1) - Y_0(t=0)|D=1] = E[Y_0(t=1) - Y_0(t=0)|D=0]$$

- If the assumption holds, then

$$\begin{aligned} E[Y_1(1) - Y_0(1)|D=1] &= \{E[Y(1)|D=1] - E[Y(1)|D=0]\} \\ &\quad - \{E[Y(0)|D=1] - E[Y(0)|D=0]\} \end{aligned} \tag{3}$$

- Graphical interpretation of the DinD


y
$E[Y(1)|D=1]$
$\mathbf{E[Y_1(1) - Y_0(1)|D=1]}$
$E[Y_0(1)|D=1]$
$E[Y(0)|D=1]$
$E[Y(1)|D=0]$
$E[Y(0)|D=0]$
$t=0$
$t=1$

- Estimator for **longitudinal data**:

$$\left\{\frac{1}{n_1}\sum_{D_i=1} Y_i(1) - \frac{1}{n_0}\sum_{D_i=0} Y_i(1)\right\} - \left\{\frac{1}{n_1}\sum_{D_i=1} Y_i(0) - \frac{1}{n_0}\sum_{D_i=0} Y_i(0)\right\} =$$

$$= \frac{1}{n_1}\sum_{D_i=1}\{Y_i(1) - Y_i(0)\} - \frac{1}{n_0}\sum_{D_i=0}\{Y_i(1) - Y_i(0)\}$$

- Estimator for **repeated cross-sections data**:

$$\left\{ \frac{1}{n_1^1} \sum_{D_i(1)=1} Y_i(1) - \frac{1}{n_0^1} \sum_{D_i(1)=0} Y_i(1) \right\}$$
$$- \left\{ \frac{1}{n_1^0} \sum_{D_i(0)=1} Y_i(0) - \frac{1}{n_0^0} \sum_{D_i(0)=0} Y_i(0) \right\}$$

- The ATET estimate (α) can be obtained using regression techniques
  - **Longitudinal**

$$\Delta Y = \delta + \alpha D + X\beta + u$$

  - **Repeated cross-sections**

$$Y = \mu + \gamma D + \delta T + \alpha(D \cdot T) + X\beta + \epsilon$$

- Previous estimators did not include covariates, $\mathbf{X}$; they can be included to **control for confounders**.

- Two threats to the validity of DinD estimators:

    1. **Compositional differences**: The use of repeated cross-sections is only valid when the composition of the target population does not change between the two periods. Can be tested: Distribution of (D, X) must be the same for pre- and post-treatment periods.

    2. **Non-parallel dynamics**: Sometimes non-parallel dynamics can be controlled for with covariates. However, if dynamic of outcome variable depends on non-observables, identification breaks down. Test parallel dynamics: If there are data for 2 or more periods before treatment, run regression and test $\alpha = 0$.

Questions?

Thank you.

ISEG - UTL

Lisbon, 2016