Chapter 1. Potential Outcomes and Causality: Treatment Effects

Joan Llull

Quantitative Statistical Methods II
Barcelona School of Economics

Introduction

Causal inference

Econometrics: The branch of economics that aims to give empirical content to economic relations.

Three ways of using econometrics:

- **Inference of association**: quantitatively describe or summarize the co-movement of different economic variables (often used to establish "facts", and more popular with the advent of big data).
- **Forecasting**: predict one variable based on the movements of the other (umbrellas and rain).
- **Causal inference**: the process of drawing conclusions about a cause-effect connection between economic variables ("what if").

We focus on the latter, which is powerful for (*ex-post*) **evaluation of public policies**.

Evaluation of Public Policies

Evaluation of public (and private) policies is very important for efficiency, and ultimately to improve welfare.

Vast **literature** in economics, mostly in public economics, but also in development economics and labor economics, devoted to the evaluation of different programs:

- training programs
- welfare programs
- wage subsidies
- minimum wage laws
- taxation
- Medicaid and other health policies
- school policies, feeding programs
- microcredit and a variety of other forms of development assistance
- ...

These analyses aim at **quantifying** the effects of these policies on different outcomes, and ultimately on welfare.

Structural Approach

Classic approach to quantitative policy eval. **structural approach**:

- specifies a class of theory-based models of individual choice
- chooses the one within the class that best fits the data
- and uses it to evaluate policies through simulation.

Main advantages:

- it allows both *ex-ante* and *ex-post* policy evaluation
- it permits evaluating different variations of a similar policy without need to change the structure of the model or reestimate it (out of sample simulation)

Main critique:

- host of untestable functional form assumptions which have unknown implications for the results (too much discretion)
- too much emphasis on external validity at the expense of a more basic internal validity
- complexity (transparency) & computational cost (difficult to replicate)

Treatment Effects Approach

Last two decades: **treatment effect approach**.

It has **changed** language, priorities, techniques, and practices, and the perception of evidence-based economics among economists, public opinion, and policy makers.

Main goal: **evaluate ex-post** the impact of an existing policy.

Compare distribution of a chosen outcome variable for individuals affected by the policy (the **treatment group**), with the distribution of unaffected individuals (**control group**).

Main challenge: comparison so that the distribution of outcome for the control group serves as a good **counterfactual** for the distribution of the outcome for the treated group in the absence of treatment.

Main focus: understanding of the sources of variation in data with the objective of identifying the policy parameters.

Pros and Cons of Treatment Effects

Pros: given its focus on internal validity, the exercise gives transparent and credible identification.

Cons: estimated parameters are not useful for welfare analysis because they are not deep parameters (they are reduced-forms instead), and as a result, they are not policy-invariant.

In that respect, a treatment effect exercise is less ambitious.

Crucial Questions for Causal Inference

In order to set up a **causal inference** analysis, we have to formulate **four crucial questions**:

- What is the **causal relation** of interest?
- What **experiment could ideally** be used to capture the causal effect of interest?
- What is your "**identification strategy**"?
- What is your **method of inference**?

Potential Outcomes, Selection Bias, and Treatment Effects

Potential Outcomes

Consider the **population** of individuals susceptible of a treatment:

- Y_{1i}: outcome for individual i if exposed to the treatment $(D_i = 1)$
- Y_{0i} be the outcome for the same individual if not exposed $(D_i = 0)$
- Treatment indicator: D_i

Note that Y_{1i} and Y_{0i} are **potential outcomes** in the sense that we only observe:

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

Main challenge of this approach: the treatment effect can not be computed for a given individual.

Our interest is not in treatment effects for **specific individuals** *per se*, but, instead, in some characteristics of their distribution.

Treatment Effects

Most of the time focus on two main parameters of interest:

The first one is the **average treatment effect** (ATE):

$$\alpha_{ATE} \equiv \mathbb{E}[Y_{1i} - Y_{0i}],$$

The second is **average treatment effect on the treated** (TT):

$$\alpha_{TT} \equiv \mathbb{E}[Y_{1i} - Y_{0i} | D_i = 1].$$

As noted, the main challenge is that we **only observe** Y_i. The standard measure of association between Y_i and D_i is:

$$\begin{aligned}\beta &\equiv \mathbb{E}[Y_i | D_i = 1] - \mathbb{E}[Y_i | D_i = 0] \\ &= \underbrace{\mathbb{E}[Y_{1i} - Y_{0i} | D_i = 1]}_{\alpha_{TT}} + \underbrace{(\mathbb{E}[Y_{0i} | D_i = 1] - \mathbb{E}[Y_{0i} | D_i = 0])}_{\text{selection bias}}.\end{aligned}$$

which differs from α_{TT} unless the second term is equal to zero.

The second term (selection bias) indicates the difference in potential outcomes when **untreated for individuals** that are actually treated and individuals that are not.

A nonzero difference may result from a situation in which treatment status is the result of individual decisions where those with low Y_0 choose treatment more frequently than those with high Y_0 (**difference in composition**).

An important assumption of the potential outcome representation is that the effect of the treatment on one individual is **independent of the treatment received by other** individuals. This excludes equilibrium or feedback effects, as well as strategic interactions among agents.

Structural vs Reduced/Form Effects

From a **structural model** of D_i and Y_i, one could obtain the implied average treatment effects.

Instead, here, they are defined with respect to the distribution of potential outcomes, so that, relative to the structure, they are **reduced-form causal effects**.

Econometrics has conventionally distinguished between **reduced form** effects, uninterpretable but useful for prediction, and **structural** effects, associated with rules of behavior.

The treatment effects provide this **intermediate category** between predictive and structural effects, in the sense that recovered parameters are causal effects, but they are uninterpretable in the same sense as reduced form effects.

Sample Average Treatment Effects

Sample analogs for α_{ATE} and α_{TT} are:

$$\alpha_{ATE}^{S} \equiv \frac{1}{N} \sum_{\mathrm{i=1}}^{\mathrm{N}} (Y_{1i} - Y_{0i})$$

$$\alpha_{TT}^{S} \equiv \frac{1}{N_1} \sum_{\mathrm{i=1}}^{\mathrm{N}} D_i (Y_{1i} - Y_{0i}),$$

where $N_1 \equiv \sum_{\mathrm{i=1}}^{\mathrm{N}} D_i$ is the number of treated individuals.

If factual and counterfactual potential outcomes were observed, these quantities could be estimated without error. The sample average version of β is given by:

$$\begin{aligned} \beta^S &\equiv \bar{Y}_T - \bar{Y}_C \\ &\equiv \frac{1}{N_1} \sum_{\mathrm{i=1}}^{\mathrm{N}} Y_i D_i - \frac{1}{N_0} \sum_{\mathrm{i=1}}^{\mathrm{N}} (1 - D_i) Y_i, \end{aligned}$$

where $N_0 \equiv N - N_1$ is the number of untreated individuals.

Identification of Treatment Effects under Different Assumptions

Independence

Identification of the treatment effects depends on the **assumptions** we make on the relation between potential outcomes and the treatment.

Simplest case is when the distribution of the potential outcomes is **independent** of the treatment (e.g. randomized experiments):

$$(Y_{1i}, Y_{0i}) \perp\!\!\!\perp D_i.$$

When this happens:

$$F(Y_{1i}|D_i = 1) = F(Y_{1i})$$
$$F(Y_{0i}|D_i = 0) = F(Y_{0i})$$

which implies that:

$$\mathbb{E}[Y_{1i}] = \mathbb{E}[Y_{1i}|D_i = 1] = \mathbb{E}[Y_i|D_i = 1]$$
$$\mathbb{E}[Y_{0i}] = \mathbb{E}[Y_{0i}|D_i = 0] = \mathbb{E}[Y_i|D_i = 0]$$

and, as a result, $\alpha_{ATE} = \alpha_{TT} = \beta \Rightarrow$ **unbiased estimate** of α_{ATE}: $\widehat{\alpha}_{ATE} = \bar{Y}_T - \bar{Y}_C = \beta^S$.

No need to "control" for other **covariates**.

Conditional Independence

A less restrictive assumption is **conditional independence**:

$$(Y_{1i}, Y_{0i}) \perp\!\!\!\perp D_i | X_i,$$

where X is a vector of covariates.

This situation is known as **matching**: for each "type" of individual (i.e. each value of covariates) we match treated and controls.

Conditional independence implies:

$$\mathbb{E}[Y_{ji}|X] = \mathbb{E}[Y_{ji}|D_i = j, X_i] = \mathbb{E}[Y_i|D_i = j, X_i] \text{ for } j = 0, 1$$

and, as a result:

$$\alpha_{ATE} = \mathbb{E}[Y_{1i} - Y_{0i}] = \int (\mathbb{E}[Y_i|D_i = 1, X_i] - \mathbb{E}[Y_i|D_i = 0, X_i])dF(X_i),$$

For the treatment effect on the treated:

$$\begin{aligned} \alpha_{TT} &= \int \mathbb{E}[Y_{1i} - Y_{0i}|D_i = 1, X_i]dF(X_i|D_i = 1) \\ &= \int \mathbb{E}[Y_i - \mu_0(X_i)|D_i = 1, X_i]dF(X_i|D_i = 1), \end{aligned}$$

where $\mu_0(X_i) \equiv \mathbb{E}[Y_i|D_i = 0, X_i]$. The function $\mu_0(X_i)$ is used as an imputation for Y_{0i}.

Absence of Independence

Finally, sometimes we cannot assume conditional independence:

$$(Y_{1i}, Y_{0i}) \not\perp\!\!\!\perp D_i | X_i.$$

In this case, we will need some variable Z_i that constitutes an **exogenous** source of variation in D_i, in the sense that it satisfies the **independence assumption**:

$$(Y_{1i}, Y_{0i}) \perp\!\!\!\perp Z_i | X_i,$$

and the **relevance condition**:

$$Z_i \not\perp\!\!\!\perp D_i | X_i.$$

As we discuss later in the course, in this context we are only able to identify an average treatment effect for a subgroup of individuals, known as **local average treatment effect**.

Linear Regression and Treatment Effects

Expressing TE as a Linear Regression

Rearranging the relation between potential and observed outcomes yields:

$$Y_i = \mathbb{E}[Y_{0i}] + (Y_{1i} - Y_{0i})D_i + (Y_{0i} - \mathbb{E}[Y_{0i}]) \equiv \beta_0 + \beta_i D_i + U_i.$$

Homogeneous treatment effect: $\beta_i = \beta \ \forall i$, and $\beta = \alpha_{ATE} = \alpha_{TT}$.

Heterogeneous treatment effect: slope coefficient of regression $Y_i = \beta_0 + \bar{\beta} D_i + U_i$ is:

$$\bar{\beta} = \mathbb{E}[Y_{1i} - Y_{0i} | D_i = 1] + (\mathbb{E}[Y_{0i} | D_i = 1] - \mathbb{E}[Y_{0i} | D_i = 0]) = \beta.$$

Returns to Schooling

The **remaining points** to make to the link between regression and treatment effects are illustrated with the following example:

	(1)	(2)	(3)	(4)	(5)
Controls:	None	Age Dummies	Col. (2) and Additional Controls*	Col. (3) and AFQT score	Col. (2) with Occupation Dummies
	0.132 (0.007)	0.131 (0.007)	0.114 (0.007)	0.087 (0.009)	0.066 (0.010)

Data are from the National Longitudinal Survey of Youth (1979 cohort, 2002 survey). The table reports the coefficient on years of schooling in a regression of log wages on years of schooling and the indicated controls. Standard errors are shown in parentheses. The sample is restricted to men and weighted by NLSY sampling weights. The sample size is 2,434.

* Additional controls are mother's and father's education and dummies for race and census region.

In particular:

- Treatment variables that take more than two values
- Conditional independence
- Endogenous controls